

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

**A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313:761-4700 800:521-0600**

ACQUIRING SYNTACTIC GENERALIZATIONS FROM POSITIVE EVIDENCE:
AN HPSG MODEL

by
NINA WACHOLDER

A dissertation submitted
to the Graduate Faculty in Linguistics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy, The City University of New York.

1995

UMI Number: 9605677

**Copyright 1995 by
Wacholder, Nina
All rights reserved.**

**UMI Microform 9605677
Copyright 1995, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI
**300 North Zeeb Road
Ann Arbor, MI 48103**

© 1995

NINA WACHOLDER

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Sept. 14, 1995
Date

Janet Dean Fodor
Chair of Examining Committee

Sept 14, 1995
Date

Charles E. Cairn
Executive Officer

Janet Dean Fodor

Martin S. Chodorow

Robert Levine
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

SYNTACTIC LEARNING FROM POSITIVE EVIDENCE:
AN HPSG MODEL

by

Nina Wacholder

Adviser: Professor Janet Dean Fodor

This dissertation investigates how children can acquire, from positive evidence only, a correct grammar which does not miss significant generalizations or license ungrammatical constructions. A learning model based on the GPSG prototype of Fodor (1992) is adapted to HPSG. Fodor showed that correct, compact, language-particular syntactic generalizations can be acquired from input tree structures, if the learning mechanism strips out of syntactic trees all universally determined feature values, including, crucially, default values. However, HPSG differs from GPSG in that it assumes that all language-particular information is lexical rather than syntactic, and it renounces syntactic defaults because of their computational intractability. Two questions are therefore addressed: (i) Can apparently syntactic phenomena be acquired by lexical learning? and (ii) How can the work of syntactic defaults be done in HPSG?

These issues are illustrated using complementizers as the initial working example. A complementizer has a lexical feature in which selectional information is stored, so that properties of the clauses types with which it co-occurs can be acquired by lexical learning. A lexical entry is

acquired when the feature specifications on a lexical node in an input tree are copied into the lexicon for storage. However, unless lexical entries are compacted by eliminating feature values, they would be costly to store, redundant, and linguistically unsatisfactory. It is argued here that eliminating only fully predictable features is insufficient. Default values must also be factored out of lexical entries. What is stored is then maximally compact feature descriptions consisting only of marked feature values. Default specifications are thus needed in UG, but since in this model they are restricted to the lexicon, they do not create the computational problems that unrestricted defaults can engender.

Evidence is presented that lexical interrogatives (e.g., *who*) and phrasal interrogatives (e.g., *to which panda*) also select properties of their sister node. It is shown that even in the phrasal case, these selectional properties can be stored in the lexicon and acquired by lexical learning. The question is then raised whether other types of apparently syntactic selection (as between adjunct and matrix clauses) can similarly be re-cast as lexical.

ACKNOWLEDGEMENTS

Because of the inter-disciplinary nature of my topic, I have received advice and support from individuals who work in many different areas of linguistics. Most importantly, I owe an immense debt to Janet Dean Fodor, who offered me the opportunity to work on this topic which I have thoroughly enjoyed, and who generously shared her knowledge and experience. Her quest for rigor, her commitment to her students, and her boundless energy are inspiring. Many of Fodor's insights are now an inextricable part of my understanding of the process of language learning; I apologize if I have incorporated some of her ideas without attribution or modified them without proper acknowledgement. I am also very grateful to Martin Chodorow and Robert Levine, the other members of my committee, for their insightful comments, helpful suggestions, and steady encouragement.

In addition, I owe thanks to members of several other linguistic communities. These include the HPSG 'circle', particularly Bob Carpenter, Jo Calder, Georgia Green and Ivan Sag; my current and former colleagues at the IBM T.J. Watson Research Center, especially Roy Byrd and Yael Ravin; and my teachers and fellow students at the CUNY Graduate Center. Special thanks are also due to Judith Klavans and Mitchell Marcus for their advice and concern.

I would also like to acknowledge here those who taught me Latin and Greek, including the late Rosemary Hope of Walnut Hills High School in Cincinnati, and Nathan Greenberg, James Helm, and the late Charles T. Murphy of Oberlin College. It was their teaching of classical grammar that led me to linguistics.

It is not possible to list all of the friends and relatives who have provided support and a sympathetic ear, offered help, served as role models, refused to be insulted when I cut phone calls short, or otherwise contributed to this effort by being part of my life during the writing of this dissertation. Without them, I never could have taken on this project.

Finally, I have been privileged to have a family that encouraged me in this endeavor. My father, Ben Zion Wacholder, an exemplar of dedication to learning, provided advice and reassurance. My children, Shifra and Jacob Goldenberg, and my stepson, Alex Goldenberg, tolerated my pre-occupation and took pride as something that they recognized as a product at last emerged. Most importantly, my husband, Robert Goldenberg, has been a partner in all stages of this project; without his understanding and his willingness to give any kind of help that he could, this dissertation would never have been completed.

TABLE OF CONTENTS

CHAPTER 1	
ACQUIRING CORRECT GENERALIZATIONS FROM POSITIVE EVIDENCE	1
1.1 THE LEARNING PROBLEM	1
1.2 THE SYNTACTIC PROBLEM	10
1.3 PRINCIPLE-BASED ACQUISITION OF PHRASE STRUCTURE GRAMMAR	11
1.3.1 THE HPSG FRAMEWORK	17
1.3.2 THE NATURE OF LEARNING	22
1.4 OUTLINE OF THE DISSERTATION	26
1.5 NOTE TO READERS	28
CHAPTER 2	
HPSG AND UNIVERSAL GRAMMAR	30
2.1 THE COMPLEX FEATURE STRUCTURES OF HPSG	32
2.2 MONOTONICITY	39
2.3 SYNTACTIC CONSTRAINTS AND DECLARATIVITY	42
2.4 CONCLUSION	57
CHAPTER 3	
ACQUIRING THE SELECTIONAL PROPERTIES OF COMPLEMENTIZERS	59
3.1 THE GPSG MODEL	61
3.2 PROPERTIES OF COMPLEMENTIZERS IN HPSG	69
3.3 FEATURES THAT COMPLEMENTIZERS ARE SENSITIVE TO	76
3.4 LEXICAL ACQUISITION OF SELECTIONAL PROPERTIES	83
3.5 ACQUIRING CORRECT GENERALIZATIONS	87
3.6 GENERALIZATION BY COLLAPSING	93
3.6.1 INTERSECTIFICATION	94
3.6.2 CONSERVATIVE DISJUNCTION	96
CHAPTER 4	
DEFAULTS AND LINGUISTIC THEORY	104
4.1 THE LINGUISTIC BASIS FOR SYNTACTIC DEFAULTS	109
4.1.1 CRITERIA FOR RECOGNIZING DEFAULT VALUES	112
4.1.2 SYNTACTIC MARKEDNESS: A CASE STUDY	119
4.1.2.1 CONTEXT-FREE DEFAULT FOR ROOT	119
4.1.2.2 CONTEXT-FREE DEFAULT FOR MARKING	126
4.1.2.3 CONTEXT-SENSITIVE DEFAULT FOR ROOT AND MARKING	127
4.1.3 OTHER PLAUSIBLE DEFAULTS	137
4.2 SYNTACTIC DEFAULTS AND LEARNING	140
4.3 INDETERMINATE DEFAULTS	144
4.4 DISCARDING SYNTACTIC DEFAULTS: THE SWITCH FROM GPSG TO HPSG	149
4.5 LEXICAL DEFAULTS IN HPSG	158
4.6 CONCLUSION	165

CHAPTER 5	
LEXICAL DEFAULTS AND HPSG LEARNING	168
5.1 LEXICAL VS. SYNTACTIC: THE DISTINCTION	170
5.2 THE 'SYNTACTIC' PROPERTIES OF LEXICAL DEFAULTS	174
5.2.1 CONTEXT-SENSITIVE LEXICAL DEFAULTS	174
5.2.2 CONTEXT-FREE LEXICAL DEFAULTS	176
5.3 THE LEARNING PROCESS	179
5.3.1 ACQUISITION OF SPEC VALUES	182
5.3.2 GENERALIZATION BY DEFAULT UNIFICATION	194
5.3.3 THE LEARNING MODEL AND GRAMMAR MODEL	198
5.4 COMPUTATIONAL SAFEGUARDS	200
5.4.1 LEXICAL DEFAULTS	201
5.4.2 DETERMINATE DEFAULTS	203
5.4.3 SUMMARY	206
5.5 GENERALIZATION BY COLLAPSING OF SPEC VALUES	207
5.6 CONCLUSION	211
5.7 APPENDIX: ACQUIRING SYNTACTIC GENERALIZATIONS BEFORE	213
 CHAPTER 6	
FILLER-GAP CONSTRUCTIONS: SYNTAX AND LEARNABILITY	220
6.1 TOWARD A UNIVERSAL VERSION OF SCHEMA 6	221
6.2 SELECTIONAL PROPERTIES OF INTERROGATIVE FILLERS	228
6.3 LEXICAL FILLERS	234
6.3.1 ACQUIRING THE VALUE OF SELECT	238
6.3.2 MARKEDNESS SPECIFICATIONS	240
6.3.3 CORRECT GENERALIZATION OVER SELECT VALUES	250
6.4 PHRASAL FILLERS	252
6.4.1 LEXICAL STORAGE OF PHRASAL SELECT	253
6.4.2 ACQUIRING THE VALUE OF SELECT	258
6.4.3 COMPACT STORAGE OF PHRASAL SELECT VALUES	264
6.4.4 PHRASAL SELECT: LEXICAL OR SYNTACTIC	269
6.5 GENERALIZATION WITH POINTERS	271
6.6 CONCLUSION	281
APPENDIX: RELATIVE CLAUSES	284
 CHAPTER 7	
LEXICAL LEARNING AND LEXICAL DEFAULTS	287
7.1 THE SELECTIONAL PROPERTIES OF ADJUNCTS	289
7.2 PHRASES WITH NO OVERT LEXICAL SELECTOR	294
7.2.1 PARTICIPIAL ADJUNCTS	294
7.2.2 TOPICALIZED PHRASES	296
7.3 TOWARD A THEORY OF SELECTING CATEGORIES	303
7.4 RECONSIDERING DEFAULTS	310
 REFERENCES	 315

CHAPTER 1 ACQUIRING CORRECT GENERALIZATIONS FROM POSITIVE EVIDENCE

In this dissertation, I address the question of how (neurologically normal) children are able successfully to acquire the syntactic distinctions that they encounter in their native languages. In particular, the question to be addressed is how the learning mechanism ensures that each child will acquire a correct grammar, which does not miss significant generalizations, but also does not over-generate.

1.1 THE LEARNING PROBLEM

Syntactic generalization is the process by which a child creates a syntactic rule for a collection of syntactic phenomena that occur in its language.¹ Without generalization, each sentence type would have to be recorded individually, and so the grammar would be bulky and redundant.

However, any learning mechanism that is able to generalize runs into the risk of over-generalizing, and in the absence of negative data (see below), there is a danger that the grammar will license ungrammatical constructions. That this is a serious problem can be illustrated very clearly in a learning system that makes use of feature

¹The use of the term rule is convenient and does not necessarily imply old-style phrase structure rules. A rule can be interpreted as a condition on node admissibility, where only those nodes that are consistent with the features in the 'rule' are grammatical. This will be discussed in Section 3.1.

structure representations. A feature structure that is more general contains fewer features than a less general one, as pointed out originally by Halle (1962) and discussed for syntax by Fodor and Crain (1987). For example, (1.a) is a generalization about complementizers that are grammatical with finite verbs; (1.b) is a generalization about complementizers that are grammatical with finite verbs in embedded clauses. Because (1.b) contains an additional feature, it is more specific than (1.a). (COMPL stands for COMPLEMENTIZER.)

1. a. MORE GENERAL b. MORE SPECIFIC

COMPL+
FIN

COMPL+
FIN
ROOT-

However, natural languages do not always exhibit maximally general patterns. In fact, (1.b) is more characteristic than (1.a) is. Therefore a learner cannot acquire a correct grammar unless, in at least some circumstances, more specific (i.e. more complex) generalizations are given precedence over broader (i.e. simpler) generalizations.

For example, in English, as in many languages, complementizers like *that* and *for* are grammatical only in embedded clauses, but not in root clauses. Also, children acquiring English must learn that *that* is associated with finite verbs and *for* with infinitives. A grammar which does not distinguish between root and embedded clauses in determining the distribution of complementizers is simpler

than a grammar that does make such a distinction; a grammar that does not distinguish between *that* and *for* with respect to the clauses that they select is simpler than a grammar that does distinguish between them. Still, children do not acquire grammars that permit all complementizers in all clause types.

This example shows that syntactic learning must proceed in a way that produces generalizations, but gives priority to more specific generalizations rather than to broader, simpler generalizations. (An overly specific rule can be broadened later, in response to additional positive evidence.) Some mechanism must serve as an antidote to what would seem to be a natural preference for generalizations that mention fewer features over those that mention more features.

The question of how children acquire correct syntactic generalizations is rendered especially perplexing by the fact that learners acquire the grammar of the language of their linguistic community *solely* from exposure to a finite subset of grammatical sentences of the target language.²

An elementary theorem of mathematical linguistics states that there are an infinite number of different grammars that can generate any finite set of strings. Each grammar will make predictions about the strings not in the set. Consider the sample of the single sentence "the dog barks". It could have been taken from the language consisting of 1) all three-word strings; 2) all article-noun-verb sequences; 3) all sentences with a noun phrase; 4) that sentence alone;

²For a mathematical proof of this, see Gold (1967).

5) that sentence plus all those in the July 4, 1976 edition of the New York Times; as well as 6) all English sentences. When the sample consists of more than one sentence, the class of possible languages is reduced, but is still infinitely large, as long as the number of sentences in the sample is finite. Therefore it is impossible for any learner to observe a finite sample of sentences of a language and always produce a correct grammar for the language.

Pinker (1979:225)

Negative evidence is information about what word strings are not well-formed sentences of the language. This information could be conveyed by explicit corrections by caretakers, or in more indirect ways. It would make language-particular grammars learnable from a finite sample of sentences without innate linguistic knowledge. However, psychologists and linguists largely agree that children do not have access to systematic negative evidence that is at all commensurate with the complexity of the facts that must be acquired. (Grimshaw and Pinker 1989, Marcus 1993). Therefore children must be able to acquire the grammar of the target language from positive evidence alone.

It is unreasonable to assume that the learning mechanism requires children to engage in complex calculations, such as would be involved in comparing sentences produced by a set of candidate grammars in order to determine which one is optimal. Furthermore, since children exposed to similar input apparently acquire essentially the same grammar, the process of acquisition cannot rely on

individual creativity, insight, or random hypothesis formation (Fodor 1989b, Lightfoot 1989).³

The answer to the conundrum must be that children are genetically endowed with the ability to acquire a correct grammar.

Any kind of systematic learning -- whether in humans or in rats in cages pressing bars -- presupposes some innate similarity gradient for stimuli. Without this, as we know, language learners could generalize in infinitely many crazy ways from a single datum. They could formulate structure-independent rules for moving the seventh word of a sentence to the front, they could observe one word order and conclude that all word orders are acceptable, they could delete unrecoverable items, move non-constituents, cross tree branches with impunity. But in fact, learners don't exhibit these kinds of wild generalization patterns. Even for highly eccentric constructions like *Out popped the cuckoo*, they appear to generalize in a fairly conservative and specifically linguistic fashion -- respecting, insofar as possible, the usual syntactic categories, the usual principles of tree structure, of case assignment, binding, bounding and all. Furthermore, different learners appear to generalize in essentially the same way from the same data. It is widely agreed, for example, that *There stood the butler* is acceptable, but *Fast walked the butler* is not. It is clear, then, that the periphery is acquired not by some general purpose induction system, but by a highly constrained and specifically linguistic learning mechanism.

Fodor (1989b:133-134)

In particular, all children must construct grammars from the same basic building blocks. Chomsky (1965) noted that children need at least:

³In what follows, I sometimes describe a learner as if it has consciously observed or studied some grammatical phenomenon. This language is strictly metaphorical. The learner's perception of grammatical distinctions is typically fully unconscious, and the learner's knowledge of syntactic generalizations is abstract and not normally verbalizable.

- (i) some initial delimitation of a class of possible hypotheses about language structure; ...
- (ii) a method for selecting one of the (presumably indefinitely many) hypotheses that are allowed by (i) and are compatible with the given primary linguistic data.

Chomsky (1965:30)

In order to guarantee that all acquired grammars will be essentially the same, the "initial delimitation" provided by Universal Grammar (UG) must include a representation for mentally encoding abstract linguistic information, and a set of "construction-neutral principles" that shape the learner's expectations of the kind of linguistic patterns it will encounter. Since acquisition of a grammar is incremental, and since the learner must adopt many intermediate grammars in the process of achieving the target grammar, UG must also provide the learner with an internal yardstick for choosing better grammars over worse grammars, an 'evaluation metric'. Without such a metric, a learner might never acquire a correct grammar.⁴

⁴The goal of the Principles and Parameters approach was to alleviate the need for an evaluation metric.

In early work, economy considerations entered as part of the evaluation metric, which, it was assumed, selected a particular instantiation of the permitted format for rule systems, given PLD (primary linguistic data). Within the principles-and-parameters approach, it is generally assumed to be completely dispensable: the principles are sufficiently restrictive so that PLD suffice in the normal case to set the parameter values that determine a language.

Chomsky (1993:4-5)

However, Fodor (1989b) observed that it is doubtful that all syntactic learning can be reduced to parameter setting; if parameter setting is insufficient to account for differences among languages, then there must be a learning mechanism other than parameter setting, and it is very likely that this mechanism will require an evaluation metric.

Chomsky (1965:37) emphasized that what the evaluation metric is like is subject to empirical investigation. However, it is a standard assumption in generative grammar that the evaluation metric includes at least a simplicity measure, so that simpler generalizations like (1.a) are preferred over more complex ones like (1.b). In addition, the conflict between simplicity and generality noted above makes it apparent that the evaluation metric must include an antidote to the tendency of learners to over-generalize by simplifying. At the very least, the evaluation metric must ensure that learners postulate grammars in an order consonant with the Subset Principle.⁵

... the Subset Principle states that learning hypotheses are ordered in such a way that positive examples can disconfirm them. For many cases, the ordering will force the narrowest possible language to be hypothesized first, so that no alternative target language can be a subset of the hypothesized language. More precisely, no other target language compatible with the triggering data that led to the new hypothesis language can be a proper subset of that language.

(Berwick 1985:23; italics in original)

Among the implications of the Subset Principle for learning are the following:

○ If a learner has a choice between guessing that some collection of feature values is grammatical in some context (e.g., (1.a)) or guessing that a more highly specified set of features values is grammatical (e.g., (1.b)), the Subset Principle requires that the learner always prefer the more

⁵See also Wexler and Manzini (1987).

highly specified feature structure, since the language it licenses (other things being equal) is a subset of the language licensed by the more general feature structure.

○ Since a language with obligatory constituents is a proper subset of an otherwise comparable language with optional constituents, the Subset Principle requires that a learner always prefer obligatoriness over optionality; the learner may assume that optionality is grammatical only if there is clear (positive) evidence for it. (Berwick 1985)

○ Since a fixed word order is a subset of completely free word order, the Subset Principle dictates that learners make the initial assumption that word order is fixed. If children made random guesses about word order, or favored free word order over specific orders, they would never have a basis for correcting their grammar. Under these circumstances, fixed word order languages would not be stable; within a generation or two, they would become free word order languages. (Fodor and Crain 1990)

○ In the absence of positive evidence that a verb is subcategorized for a particular argument type (e.g. a direct object), the Subset Principle requires that the learner assume that the verb does not take that argument. Therefore the Subset Principle requires that the learner not assume that all lexical items that are subcategorized for certain realizations of arguments are also subcategorized for alternative realizations of those arguments. For example,

the sentences in (2) must not lead the learner to assume that all verbs that are subcategorized for indirect objects also occur with double object constructions (Baker 1979).

2. a. Kim gave some money to the zoo-keeper.
- b. Kim gave the zoo-keeper some money.

A learner who made this assumption would never learn that (3.b) is ungrammatical.

3. a. Kim donated some money to the zoo.
- b. *Kim donated the zoo some money.

The learning problem that I address in this dissertation is how learners can acquire correct generalizations that are economical and compact, from positive evidence. The Subset Principle is part of the solution since it guarantees conservative learning without over-generalization. But it leaves two problems. (i) What mechanism computes subset relations between candidate grammars so that the learner knows what the conservative choice is? (Fodor 1992b). (ii) Does the Subset Principle, which is essential for *learning*, necessarily entail the choice of complex grammars which are not *linguistically optimal*?

I will argue (following Fodor 1989a, 1989b, 1990, and 1992a) that the solution lies in the assumption that UG includes innate defaults, and that this assumption is independently motivated by linguistic facts. It offers an answer to the following question: why do human languages

typically display only some of the logically possible combinations of feature values?

1.2 THE SYNTACTIC PROBLEM

Human language is largely characterized by asymmetrical distributions of feature associations. For example, across languages, complementizers tend to be associated with embedded clauses but not with root clauses. So there is a special association between [COMPL+] and [ROOT-] on the one hand, and [COMPL-] and [ROOT+] on the other.

4. MORE TYPICAL FEATURE STRUCTURES

- a.

COMPL+
ROOT-

 b.

COMPL-
ROOT+

But it is not impossible for [COMPL-,ROOT-] or [COMPL+,ROOT+] to occur. [COMPL-,ROOT-] is grammatical in English in sentences like (5.c), but the contexts in which it is grammatical are a subset of the contexts in which [COMPL+,ROOT-] is grammatical.

5. a. Ruth thinks that Kim likes pandas.
 b. Ruth quipped that Kim likes pandas.
 c. Ruth thinks Kim likes pandas.
 d. *Ruth quipped Kim likes pandas.

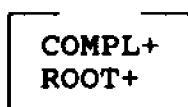
Although English does not permit (4.b), [COMPL+,ROOT+] is grammatical in some languages under some circumstances. For example, the Latin complementizer *utinam*, which introduces clauses expressing wishes, may be used in both embedded and matrix clauses; it selects subjunctive verbs.

6. Atticus 3,3 (cited in Lane (1903, par.1540)
 Utinam illum diem
 Would that that.ACC.SING.MASC day.ACC.SING.MASC
 videam
 see.PRESENT.SUBJUNCTIVE.1PERS.SING
 'I hope that I may see that day.'

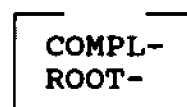
So the feature structures in (7) do occur, but are less typical of natural languages than the ones in (4).

7. LESS TYPICAL FEATURE STRUCTURES

a.



b.



The theoretical issue that I address in parallel with the learnability problem is why certain feature values tend to be associated in human language, while others associations are relatively rare. If we can discover what these asymmetrical associations are, we may be able to get insight into the principles underlying them.

1.3 PRINCIPLE-BASED ACQUISITION OF PHRASE STRUCTURE GRAMMAR

Fodor (1989a, 1990) suggests that the answer to both the learning question and the syntactic question is the same: universal defaults. The fact that languages contain 'typical' patterns that are not universal suggests that universal principles are actually of two kinds:

- Absolute universal principles that can never be violated in any language, or in any grammar that a child acquires in the process of learning the target language. They require no learning.
- Default values for features (or markedness specifications) which a language will exhibit unless its

grammar explicitly states the contrary. Default values shape the child's expectations of what grammars are like, and are reflected in recurrent asymmetrical feature associations across languages. However, the fact that they are default values means that they are not absolutely required; non-default values can be expected to be grammatical in some languages in some contexts. A child must adopt a grammar that overrides an innate default only on the basis of positive evidence.

Given this view of UG, Fodor (1990, 1992b) and Fodor and Crain (1990) proposed that the problem of accounting for correct syntactic generalizations can be fruitfully approached within a research framework where a 'learner' is conceived of as subtracting out of a 'fully specific' tree those properties which are universal. The learner, a hypothetical acquirer of syntactic generalizations, is a learning device analogous to the parser in sentence processing research. The fully specific tree is a structure in which a correct values is specified for every feature.

8. FULLY SPECIFIC TREE - UG = LANGUAGE-PARTICULAR RULE

On the assumption that the input to the process of generalization consists of grammatical sentences and (correct) fully specified tree structures, the learner's task is simply to factor novel constructions into two parts: those properties which are predictable because they follow from principles of UG, including UG defaults, and those

properties which must be recorded in the grammar because they are language-particular or item-particular, and hence are not predictable. Default values play an important role in the acquisition of compact generalizations because they are predictable values that do not need to be acquired or stored. After all universally determined feature values, including default values, are stripped away, what is left for learners to store is a purely language-particular generalization, consisting only of marked values. If only fully universal feature values could be stripped off, the residue would be too complex to be stored as part of a grammar. Stripping off default values too achieves a much more significant reduction in complexity, so that this learning method results in linguistically plausible grammars.

Fodor observed that a feature-based phrase structure grammar naturally lends itself to this "principle-based" approach.

The adoption of syntactic feature notation turns out to be the one most important factor in permitting the universal principles to be abstracted away from language-specific rules. ...every construction has some aspects that are universal and some that are language-specific; the generalizations cut across the constructions. ...In a phrase structure theory employing feature notation, the principles specify all universally predictable feature values within trees. Any feature value so specified would be redundant in a rule, so it can be omitted from the rule, thus simplifying the grammar. When all the redundant specifications are eliminated, what is left is a schematic rule whose remaining features encode just the properties of the construction in question that are peculiar to it. ... What would have been a monolithic

rule in traditional phrase structure theory can thus be factored, thanks to feature notation, into two parts: its universal aspects and its language-specific or construction-specific residue. The latter is all that needs to be recorded in the grammar for the language. The former need not be specified, since it can be supplied by universal feature instantiation principles.

Fodor (1990:242-243)

Moreover, these stripped-down phrase structure rules can be viewed as a kind of parameter, defining the range of variation among natural languages.

Note that it is not a contradiction to talk of rules as parameters. It is true that they are technically different from many typical GB parameters, in that they do not represent alternative choices with respect to *some part of a principle*...But as far as their function is concerned, they are exactly like parameters. (i) They specify all and only those aspects of (the syntax of) a language which the principles themselves do not determine. (ii) They define the units by which the theory permits one grammar to differ from another. (iii) Hence they characterize the steps that children can take from one grammar to the next in their advance towards the adult language.

Fodor (1990:237)

One of Fodor's pioneering insights was that the learnability problem could be fruitfully explored within the framework of feature-based phrase structure grammars, given the assumption of innate feature value defaults. In this dissertation, I present a kind of formal experiment designed to investigate the implications, for syntax and for learnability, of her hypothesis about UG, and its role in the acquisition of syntactic generalization. I update her model to make it consistent with more recent theoretical work on phrase structure grammars, and I confront some formal and practical objections that have been made against default

feature values. If these potential problems can be avoided, Fodor's general approach to solving learnability problems for phrase structure grammars will still be viable.

Adoption of Fodor's approach allows my research to concentrate on that part of linguistic information which evidently must be learned, the language-particular, in the hope that it will ultimately shed further insight into universal principles. In order to focus on how language-particular generalizations are acquired, I make certain other idealized assumptions about the learning process.

○ The structure available to the learner at the point where it is ready to acquire generalizations has fully specific, correct feature values on all nodes. This idealization allows me to direct my attention to the more tractable process of acquisition of generalizations, rather than the more obscure 'string-to-structure' process by which the learner assigns a syntactic analysis to a novel sentence which is not licensed by the current (incorrect) grammar. The problem of how the learner knows what structure to assign, and can correctly establish feature values for a novel sentence is as yet too hard to solve (though see Fodor 1989a and 1995 for speculations on the form of a solution to this problem, whose importance was pointed out by Virginia Valian).⁶ However, without a parsed structure, the learner

⁶Green (1994) discusses the string-to-structure problem in the context of HPSG but makes radically different assumptions about the nature of learning. First, she assumes that learners

would not have the vocabulary to form syntactic generalizations, so as a methodological strategy, I simply stipulate that the learner has (somehow) already achieved a fully specified tree which is the input to the learning process, and I consider only issues that arise in inducing the target grammar from such trees.⁷ In what follows, I occasionally refer to a 'fully specific tree' with a correct value for every feature on every node as a 'fully parsed tree' for descriptive convenience, but I do not mean by this to take a stand on how the parse is achieved, or whether it must necessarily precede rule induction operations.⁸

○ The learner is linguistically mature so that all relevant UG principles are available to it simultaneously. (See Borer and Wexler (1987), Weissenborn, Goodluck and Roeper (1992) and Weinberg (1990a and 1990b) for a discussion of maturational issues.)

have have very little or no innate linguistic knowledge. Second, she assumes that learners first acquire more general characteristics of some linguistic object and then "make incrementally finer distinctions".

⁷For successful learning, the learner must also be able to correct any errors it has made in the string-to-structure stage. See Gibson and Wexler (1994) and Fodor (1995).

⁸The hope underlying this research strategy is that progress on the generalization problem will provide insight into the string-to-structure problem. There is no doubt that discovery of a solution to the string-to-structure problem will in turn necessitate re-working of the solution to the generalization problem.

1.3.1 THE HPSG FRAMEWORK

The general approach toward the learning problem discussed in the previous section was developed by Fodor for Generalized Phrase Structure Grammar (GPSG) specifically for the syntactic theory presented by Gazdar, Klein, Pullum and Sag (1985) (hereafter, GKPS). An algorithm for stripping default values off of fully specific GPSG trees was developed by Maxfield (1990). In this dissertation, I adopt Head-Driven Phrase Structure Grammar (HPSG), the successor theory to GPSG, as the framework for my learning model.

GPSG and HPSG share many properties that make them very useful for a study of this sort.⁹

○ A central concern of both GPSG and HPSG is with the internal architecture of the entire linguistic system. They are theories both about the representation of linguistic information and about matters of linguistic substance. As such, they provide an ideal framework for development of an explicit learning model, in which all steps in the learning process are represented.¹⁰ P&S-94 state that their

⁹The general approach was originally developed by Gerald Gazdar (e.g. Gazdar, 1981).

¹⁰ Pinker noted the importance of formal models in clarifying issues of learnability.

(A) potential benefit of formal models is the explicitness that they force on the theorist, which in turn can clarify many conceptual and substantive issues that have preoccupied the field. ... Developing explicit, mechanistic theories of language learning may be the only way that these issues can be stated clearly enough to evaluate.

Pinker (1979:219-220)

commitment to precise syntactic analyses worked out in explicit detail is undertaken with the following goal:

...by taking our system of constraints -- appropriately formalized -- as a hypothesis about the structure of human linguistic knowledge, we hope eventually to exploit formal and computational properties of the theory to develop integrated models of language processing...

P&S-94:59

The learning model that I develop in this dissertation takes advantage of exactly these properties of HPSG.

○ All linguistic information (syntactic, semantic, phonological, and lexical) is represented by complex feature structures. Features are arbitrary labels that capture syntactic distinctions that people are capable of making. Feature structures thus provide a concrete and explicit representation of human linguistic knowledge. Feature notation also facilitates sorting out of universal and syntactic properties across constructions, as discussed above.

○ In feature-based systems, universal principles and language-particular syntactic generalizations (including lexical entries) are all represented by feature structure. Thus the relationship between universal principles, language-particular syntactic generalizations, and fully specific trees is transparent: the features which label nodes of trees are themselves the vocabulary in which universal principles and language-particular syntactic generalizations are expressed.

instantiated on every node that is between the filler and the gap. All syntactic constraints, even those on long-distance relationships, can thus be stated in terms of features instantiated on local trees.

○ HPSG is a declarative, unification-based theory in that the grammatical constraints are in principle order-free. This means that a local tree can be built up by applying grammatical principles in any order. As a result, the grammar itself is free of implicit psychological claims about how the grammar is used, and a single set of grammatical principles can be used for parsing and for generation.¹¹ Sag et al. (1992:302) contrast declarative models with "transformational models (including GB models where 'Move α ' is interpreted literally)".

Rules perform mappings from total representations of a sentence at a particular 'level of representation' to other total representations, and these mappings have typically involved deletions, permutations, insertions and the like. Moreover such mappings are not in general reversible; nor are their computational properties well understood.

Sag et al. (1992:302)

In what follows, I assume that the input to the generalization process takes the form of fully specific syntactic trees which accord in all respects with the theory of Pollard and Sag (1994), Chs. 1-8, hereafter P&S-94,¹²

¹¹How the grammar is used in human sentence production and perception will not be addressed here.

¹²Ch. 9 makes theoretical revisions which I will not refer to unless they are directly relevant.

and of Pollard and Sag (1987), hereafter P&S-87, to the extent that this is not superceded by P&S-94. I modify the syntactic theory of these two works only as much as necessary to account for successful language learning. For example, it will sometimes be necessary to separate an HPSG principle into those components which are plausibly universal and those which are evidently language-particular. Few, if any, substantive changes are needed, other than the adoption of default values for features in lexical entries, as explained below.

The adoption of HPSG as my working syntactic framework means that the learning model that I construct must take into account two important ways in which HPSG differs from GPSG. In P&S-94, as in much current linguistic theory, most language-particular facts are assumed to be lexical rather than syntactic. In fact, a trend towards treating the lexicon as the only source of language variation is apparent even in GB, the framework in which parameter setting was first introduced (e.g., Hermon 1994). P&S-94 posit some parameters in HSPG¹³ Since parameter setting has been

¹³For example, P&S-94 posit an "English-particular parameterization of the Trace Principle.

i. P&S-94 (Ch.4(33))

Trace Principle (parameterized for English)

Every trace must be strictly subcategorized by a substantive head.

The stipulation that the subcategorization must be 'strict' constitutes the parameterization.

They also describe the Clausal REL Prohibition (Ch.5(31)), which outlaws in situ relative phrases, as a "parameter of variation" (P&S-94:220).

studied extensively elsewhere (e.g., Roeper and Williams (1987), Clark (1990), and Fodor 1990), I do not discuss it here but concentrate instead on building a model to account for the acquisition of lexical entries. My research supports the claim that much apparently syntactic information is actually stored in the lexicon.

The second major difference between GPSG and HPSG is that HPSG has eliminated syntactic defaults, which, as noted above, were central to Fodor's learning model for GPSG. In what follows, I explore the consequences, for learnability theory and for syntactic theory, of the absence of syntactic defaults.

1.3.2 THE NATURE OF LEARNING

As discussed above, I make certain fundamental working assumptions about learning that shape the discussion that follows. These include:

- The normal process of first language acquisition by neurologically intact human infants requires no conscious study or complex calculation.
- Language learning is automatic and mechanical in that it requires no creativity or hypothesis formation and testing.
- Learners do not have access to systematic negative evidence.
- Learners obey the Subset Principle.

In order to avoid being side-tracked by other issues, and focus on the problem of how learners are able to acquire syntactic generalizations, I make several other working assumptions in this dissertation.

○ Children are 'least effort learners' (Fodor 1985) in the sense that they never adopt a grammar more complex than is necessary for comprehending novel sentences, and they never engage in activities or computations (e.g., hypothesis testing or seeking out counter-examples) which do not directly contribute to the licensing of the sentences that they encounter.

○ Children are conservative learners (Fodor 1992b) who do not change their grammar unless positive evidence indicates the need to do so. Wexler and Culicover (1980) refer to this as 'error-driven learning'.

○ Upon encountering a novel construction, the learner will revise its grammar to one that does license that construction (if it can find a new grammar that does so); however, it does not shift to a new grammar that does not license the construction. Gibson and Wexler (1994) call this the Greediness Constraint. (Of course, a new grammar thus adopted is not necessarily the target grammar.)

○ Since the local tree is the domain of all syntactic principles, it is also the domain of syntactic learning. The learner's task is to acquire language-particular

syntactic constraints that operate upon local trees (Fodor 1990).

○ The inventory of possible syntactic features is innately given. A learning model that requires the learner to invent an entire inventory of features might perhaps be developed but is not a realistic possibility at present (Fodor 1989b). The syntactic distinctions captured by features (especially those that do not carry meaning) are often abstract and are often not a priori the most obvious ones for an arbitrary communication system to draw.¹⁴

○ UG pre-specifies the organization of features within the complex feature structures, as well as all 'appropriate' features for all environments.¹⁵ Thus VERB and NUMBER are innately specified as appropriate for verbs, and NOUN, NUMBER and CASE are specified as appropriate features for nouns.¹⁶

¹⁴Pullum (1984) argues that universal grammar does not limit the range of agreement features.

¹⁵In HPSG terms, a feature is 'appropriate' if it is instantiated on a node in accordance with the requirements of the grammar. See Ch. 2 for further discussion of this notion.

¹⁶How the learner is able to recognize that particular items in the target language exemplify particular values of particular features is part of the string-to-structure problem referred to above. Even if the feature VERB is innate, the learner might encounter verbs without knowing that they are instances of VERB. Semantic bootstrapping theories such as that of Pinker (1984) are intended to solve this problem. Since the 'fully specified tree' assumption ensures that all features have already been correctly identified, I do not address this problem here.

○ Properties of phonologically null elements with language-particular properties are potentially hard to learn and therefore should be avoided if possible.

○ The learner has no way of knowing when its grammar is complete, since it cannot know when it has heard all of the relevant types of local trees that are grammatical in its language. Therefore it is impossible for the learner to adopt the strategy of waiting until its linguistic knowledge is complete to form generalizations. (This seems independently unlikely, since, if learners had to store acquired linguistic knowledge without generalization, their grammars would be much more cumbersome than that of adults.)

In order to focus on syntactic learning, I also set aside several other types of learning that interact with syntactic learning in complex and interesting ways: phonological learning, morphological learning, semantic learning, and learning of lexical rules. I do not discuss these matters unless they directly affect the issues at hand.

I also do not consider how the learner can tell the difference between relatively rare constructions and speech errors. (See Valian (1990) and Kapur (1993) for a discussion of this problem.)

In addition, I generally refrain from discussing empirical findings on the actual course of first language acquisition by children. Pinker (1979) notes that the

language learning problem is so hard that we must start by exploring particular facets of the problem, and hope that a battery of different approaches will eventually lead to a unified solution.

... a theory that is powerful enough to account for the fact of language acquisition may be a more promising first approximation of an ultimately viable theory than one that is able to describe the *course* of language acquisition, which has become the traditional focus of developmental psycholinguistics. As the reader shall see, the Learnability criterion is extraordinarily stringent, and it becomes quite obvious when a theory cannot pass it. On the other hand, theories concerning the mechanisms responsible for child language per se are notoriously underdetermined by the child's observable linguistic behavior. This is because the child's knowledge, motivation, memory, and perceptual, motor, and social skills are developing at the same time that he is learning the language of his community.
Pinker (1979:219-220)

1.4 OUTLINE OF THE DISSERTATION

In Chapter 2, I review the universal principles of HPSG.

In Chapter 3, I illustrate the difficulty of acquiring correct generalizations without defaults. My working example is acquisition of the selectional properties of complementizers. This was chosen because the facts are relatively straightforward and permit efforts to be focussed on development of a learning model. It also allows illustration of the interplay of lexical and syntactic information, which is of linguistic interest and turns out to be central to solving learnability problems.

In Chapter 4, I argue on general grounds that default (or unmarked) values are justified syntactically in that

they account for strong cross-linguistic tendencies that are not universal. Although there is not yet a well-worked-out theory of defaults, (or markedness), linguists frequently rely on defaults to account for properties that do not follow from other universal principles. Even though P&S-94 (Ch.1, fn.22) state "Except for expository convenience, HPSG does not employ any notion of defaults.", defaults are intrinsic to the lexical model presupposed in their theory. After reviewing the computational problems that led to the abandonment of defaults in the transition from GPSG to HPSG, I conclude that the evidence from both syntactic theory and learnability theory creates a strong presumption that defaults are needed, and that therefore the computational problems must somehow be solved or side-stepped, so that defaults can be taken seriously as a component of a feasible learning model.

In Chapter 5, I present a model for learning the selectional properties of complementizers in which all learning is lexical. The syntactic approach of P&S-94 lends itself to an approach in which 'horizontal' syntactic information involving selectional relations between sisters can be stored lexically. Language-particular lexical entries are acquired when information from fully specified lexical nodes of input sentences is transferred into the lexicon. There, UG-predictable and default values are factored out, so that learners store compact lexical entries that include only

marked feature values. Importantly, since defaults are restricted to the lexicon, they are not accessible to recursive syntactic operations, and thus do not create the computational problems that unrestricted defaults can engender.

In Chapter 6, I show that the learning model that I developed for complementizers can be extended to lexical interrogative fillers. I also demonstrate that the selectional properties of phrasal fillers, in which the interrogative lexical item (e.g., *which*) is a non-head daughter within the interrogative phrase filler phrase (e.g. *to which panda*) are stored in the lexicon, and thus can also be acquired by lexical learning.

In Chapter 7, I take stock of how much of the grammar can be safely acquired on these assumptions.

1.5 NOTE TO READERS

In order to create a unified account of the language learning process, it is necessary to bring together ideas from psycholinguistics, computational theory, and syntactic theory. Because of the interdisciplinary character of this project, I have tried to make my account accessible to as many readers as possible by explaining all necessary concepts. This attempt is responsible for some of the length of this dissertation. In case my effort has failed, I apologize and offer some background reading which may be helpful.

Readers who are unfamiliar with the logic of learnability theory may wish to consult Pinker (1979) for an excellent introduction. Shieber (1986a), Carlson and Linden (1987), and the early chapters of P&S-87 discuss the notion of unification in an accessible way. Sells' (1985) introduction to the concepts and motivations underlying GPSG, though outdated in technical detail, may be useful to those who have not previously been introduced to phrase structure grammars.

CHAPTER 2
HPSG AND UNIVERSAL GRAMMAR

In this chapter, I discuss the relationship between the syntactic theory of HPSG (P&S-94) and the approach to learning that I adopt in this dissertation. In order to construct a model of how language is learned, it is useful to adopt one of two basic research strategies:

- i. Assume that no language facts are innate, except where there is evidence to the contrary.
- ii. Assume that all language facts are innate, except where there is evidence to the contrary.

The research strategy that I employ in this dissertation is the second one. This move allows me to concentrate on phenomena that are language-particular and evidently must be acquired. The expectation is that this strategy will ultimately provide further insight into universal linguistic principles, even though at this early stage, it may overestimate their strength. Once we have an understanding of the process by which those linguistic phenomena which are language-particular and therefore indubitably must be learned, then it could be considered whether other phenomena are also so learned.

The particular approach that I take toward studying syntactic learning is based on the formula for syntactic learning developed by J.D. Fodor for GPSG, the predecessor theory of HPSG.

1. Fodor's formula (=Ch.1(8))
FULLY SPECIFIC TREE - UG = LANGUAGE-PARTICULAR RULE

This formula is based on the assumption that the information in UG is the complement of the information that must be acquired and stored. In order to acquire a generalization, the learner must factor a novel construction into two parts: those properties which are predictable because they follow from principles of UG, and those properties which must be learned because they are language-particular. After the predictable properties have been factored out, the unpredictable properties, which compose the acquired generalization, must be stored.

In order to proceed with this approach, it is necessary to have a concrete hypothesis about what is contained in UG. Such an explicit hypothesis is available in HPSG. P&S-94 characterize the division between universal grammar and the grammar of a particular language "roughly" as follows (examples and explanations will be given in this chapter):¹

Universal Grammar:

Linguistic ontology: the inventory of universally available sorts of linguistic entities, together with a specification of their appropriate attributes and their value sorts.

Schemata: a small, fixed inventory of universally available phrase types (schematic immediate dominance rules).

Universal constraints on well-formed phrases.

¹I use 'universal grammar' to refer to P&S-94's conception, which is agnostic on the ontology of universal grammar, and 'UG' to refer to the innate Universal Grammar that I will be assuming for purposes of the learning model. The hope is that the two will coincide.

Particular Grammar:

Lexicon: a system of lexical entries (possibly interrelated by lexical rules).

Linguistic ontology: selection from and further articulation of the universal linguistic ontology.

Schemata: selection from and further specification (e.g. for the constituent order) of the universally available schemata.
(P&S-94:Ch.1(67))

However, P&S-94 define universal grammar as belonging to linguistically mature adults.

... we take it to be the central goal of linguistic theory to characterize what it is that every linguistically mature human being knows by virtue of being a linguistic creature, namely universal grammar.
P&S-94:14

This definition includes all linguistic information in the domain of universal grammar, without distinguishing between the grammar of particular languages and universal grammar.

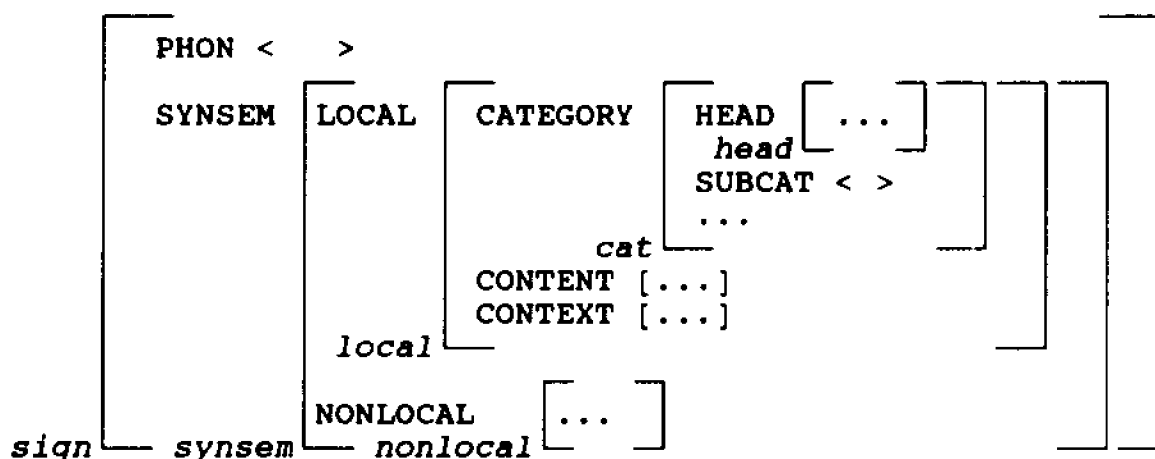
In line with strategy (ii) above, for purposes of modelling language learnability, I make the working assumption that any aspect of the theory of P&S-94 (Chs.1-8) which is plausibly universal is innate and therefore does not need to be learned. I will review these facts here, necessarily very briefly, and with attention on matters of particular relevance in later chapters. For further details, readers should consult P&S-94.

2.1 THE COMPLEX FEATURE STRUCTURES OF HPSG

All linguistic information in HPSG is represented by complex feature structures. This is in part because feature structures represent linguistic abstractions very

conveniently, but also because feature structures are computationally tractable and their formal properties are relatively well understood.² The complex feature structure in (2), which is shared by all linguistic signs, thus embodies a claim about the organization of linguistic information.

2. An HPSG sign



(2) is technically an attribute value matrix (AVM). It is hierarchically organized from the more general on the left to the more specific on the right. An attribute is a feature whose value is non-Boolean; in fact, the value of an attribute can be another feature and its value. So that each feature can be identified unambiguously, its full name includes the specification of the path of features on which it lies; a reference to SUBCAT is actually a reference to SYNSEM|LOCAL|CATEGORY|SUBCAT.

²For an introduction to feature theory as applied to linguistics, see Shieber (1986a), Kasper and Rounds (1990), and P&S-87.

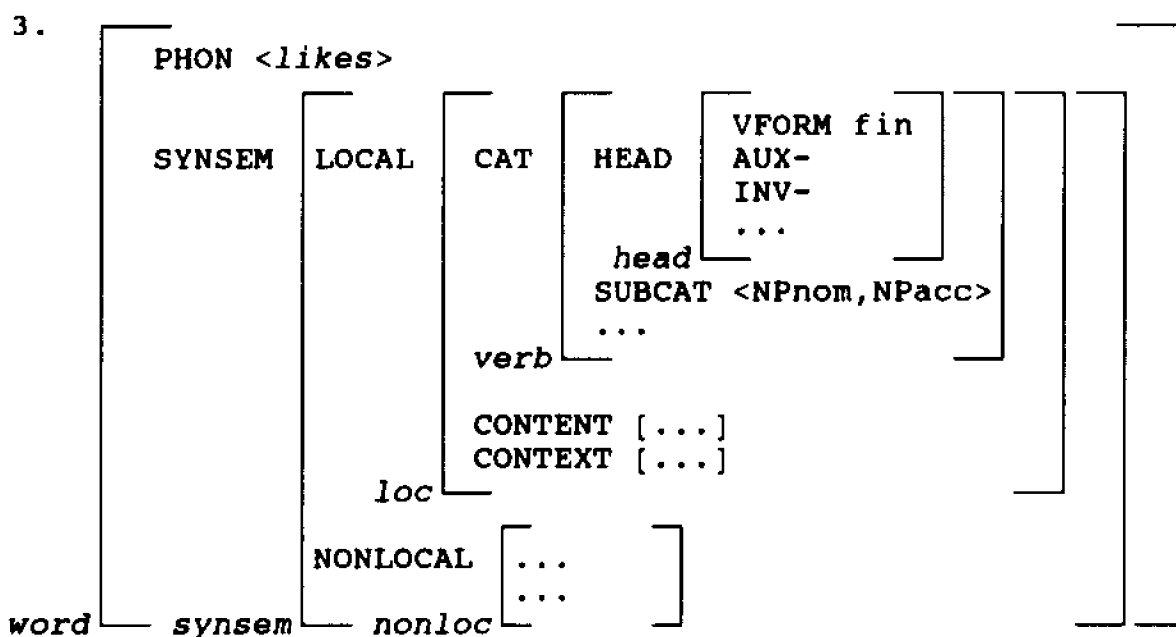
The italicized designations of subparts of the complex feature structure, called 'sorts' or 'types', are labels for the kind of object that is being modelled in that subpart.³ Sorts can be understood as "organizing features into natural classes" (Carpenter 1993:11). Sorts must satisfy a technical condition of 'appropriateness'; every sort may include only appropriate features. (2) is grammatical, for example, only if PHON and SYNSEM are appropriate attributes for the sort *sign*. In practice, the names of sorts and attributes are frequently redundant, so it is often sufficient to refer to only one of them.

With regard to linguistic substance, (2) reflects the HPSG hypothesis that all linguistic information, syntactic, semantic, and phonological, is available at a single level. According to the analysis of P&S-94, any member of the class *sign* has at least two appropriate attributes, PHON, a phonological feature whose value will be represented simply by the spelling of the sign, and SYNSEM, an attribute that includes much of the information that current transformational models distribute between D-structure and LF. It can be seen in (2) that the attribute SYNSEM is subdivided into two sorts, *local* and *nonlocal*. The sort *local* contains syntactic and semantic information and is subdivided into three sorts: *cat*, which contains information about syntactic

³Carpenter (1992) discusses the logic of sorted feature structures; his work was an outgrowth of P&S-87, and itself provides the formal underpinnings for P&S-94.

properties of the sign; *content*, which specifies the sign's interpretation, especially with regard to reference and semantic roles; and *context* which contains background information relating to presuppositions and felicity conditions. Features of sort *nonlocal*, which are associated with unbounded dependency constructions, and will be discussed below, pass *local* information between a filler and its corresponding gap.

(3) represents the feature structure of the verb *likes*, as in *Kim likes red pandas*.



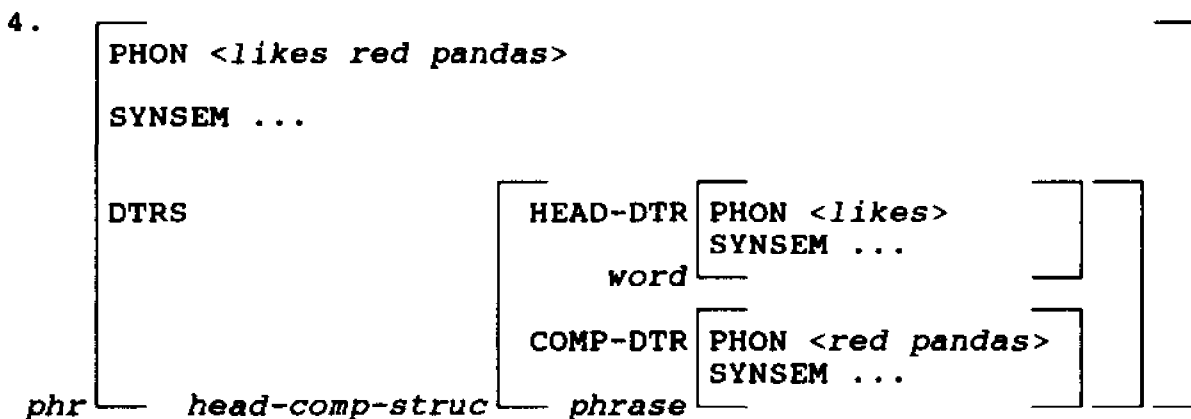
(3) is a *word*, one of the legal subsorts of *sign*. The sort *verb* indicates that the category of *like* is a *verb* rather than a *noun* or some other subsort of *cat(egory)*.

Head features (technically *SYNSEM|LOCAL|CAT|HEAD*) are subject to a universal principle, the *Head Feature Principle*, which requires that the value of *HEAD* features be

identical on a phrase and its head daughter. Head features which are appropriate for the sort *verb* include the attribute V(erb)FORM, whose values include at least *finite*, *infinitive*, and *base*, and the Boolean features AUX(iliary), and INV(ersion).⁴

The only non-head attribute of verbs which I consider here is SUBCAT(egorization), whose value represents the fact that the verb *like* is subcategorized for a nominative subject NP and an accusative direct object NP; by convention, the subject is specified first and then other arguments are listed in descending order of obliqueness.

A *sign of sort phrase* differs from a *sign of sort word*, in that it has the additional attribute DTRS (daughters).

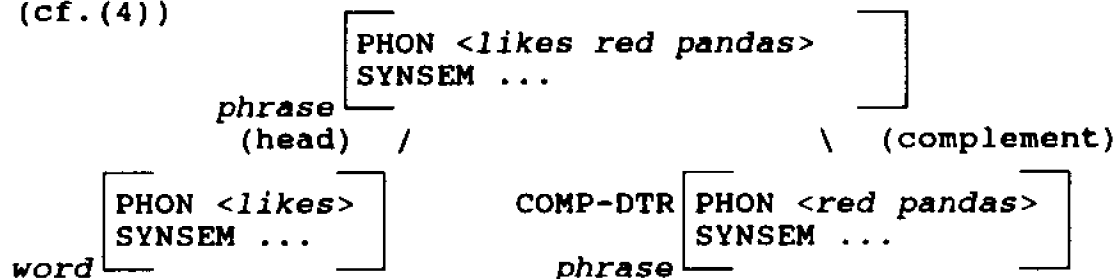


In the example here, the value of DTRS is of sort *head-comp(lement)-struc(ture)*, which has a lexical HEAD-DTR, and zero or more COMP(lement)-DTRs. Embedded within a phrase is thus the entire feature structure of all of the constituents

⁴I follow P&S-94 in assuming that INV and AUX are head features, as proposed in GKPS, but other possibilities could be considered.

of the phrase, so that the structure of any phrase, no matter how complex, can be represented in a single AVM. But for convenience of exposition, a tree-like format is frequently used, as in (5), and predictable attributes and sorts are often omitted.

5. (cf.(4))



Structural and linguistic well-formedness constraints are formally imposed on local trees by the requirement that they be 'totally well-typed' (Carpenter 1992). A sort is 'well-typed' if only features that are appropriate for a particular sort are present, so that, for example, the feature NFORM (nounform) is not used with the sort *verb*. A sort is totally well-typed and therefore complete if every feature appropriate for the sort is present.⁵

In Carpenter's system (1992:84), the user must specify the types in advance. In the case of HPSG, P&S-94 (Appendix A.1) specify a sort hierarchy and a declaration of feature appropriateness which reflects their hypothesis about the organization of linguistic content. A feature structure is legal and complete only if all appropriate features, as

⁵These sorted feature structures replace the Feature Co-occurrence Restrictions of GPSG.

defined by P&S-94, are present and properly organized according to sorts.

One of the underlying assumptions of this dissertation is that features, which are simply arbitrary labels for syntactic distinctions that human beings are capable of making, provide a useful representation of human linguistic knowledge. However, the requirement that the grammar be totally well-typed places a strong burden on a learner: if the learner had to invent a system of sorts, what would ensure that it could invent a system that is well-typed? Furthermore, what would ensure that the learner would recognize when the sort system that it has hypothesized is totally well-typed? Given my working assumptions, I simply assume that these matters are pre-specified by UG in the Universal Feature Appropriateness Principle.

6. UNIVERSAL FEATURE APPROPRIATENESS PRINCIPLE (UFAP)
UG pre-specifies a sort hierarchy and a declaration of appropriate features.

The UFAP entails that UG imposes a system for representing and organizing linguistic information on the learner, thus absolving the learner of the perhaps impossible task of inventing its own representational system.⁶

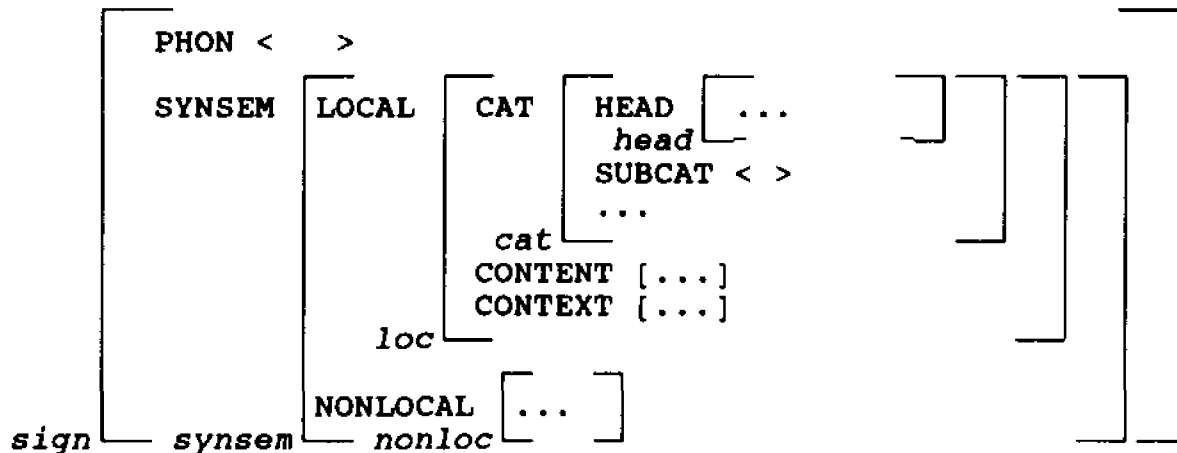
⁶The UFAP entails that all features in all languages be universal. However, note that neither AUX nor INV is universal. INV is a relational feature whose effects vary with the base order: inversion in a VSO language is different than inversion in an SVO language. Auxiliaries are not universal because they are associated with different syntactic behavior in different languages. Therefore, neither AUX nor INV are learnable under the working assumptions of this dissertation. However, it would take me to far afield to consider what universal features might

The UFAP entails one important revision of P&S-94's distinction between universal grammar and language-particular grammar. I noted above (p.32) that P&S-94 label as language-particular, "selection from and further articulation of the universal linguistic ontology", by which they presumably mean language-particular features. However, the Universal Feature Appropriateness Principle, in combination with the requirement that feature structures be totally well-typed, means that all appropriate features must be present on all nodes in all languages, even if some language appears specifically not to make all the feature distinctions. The consequences of this assumption for the process of acquiring linguistic generalizations will be explored in what follows.

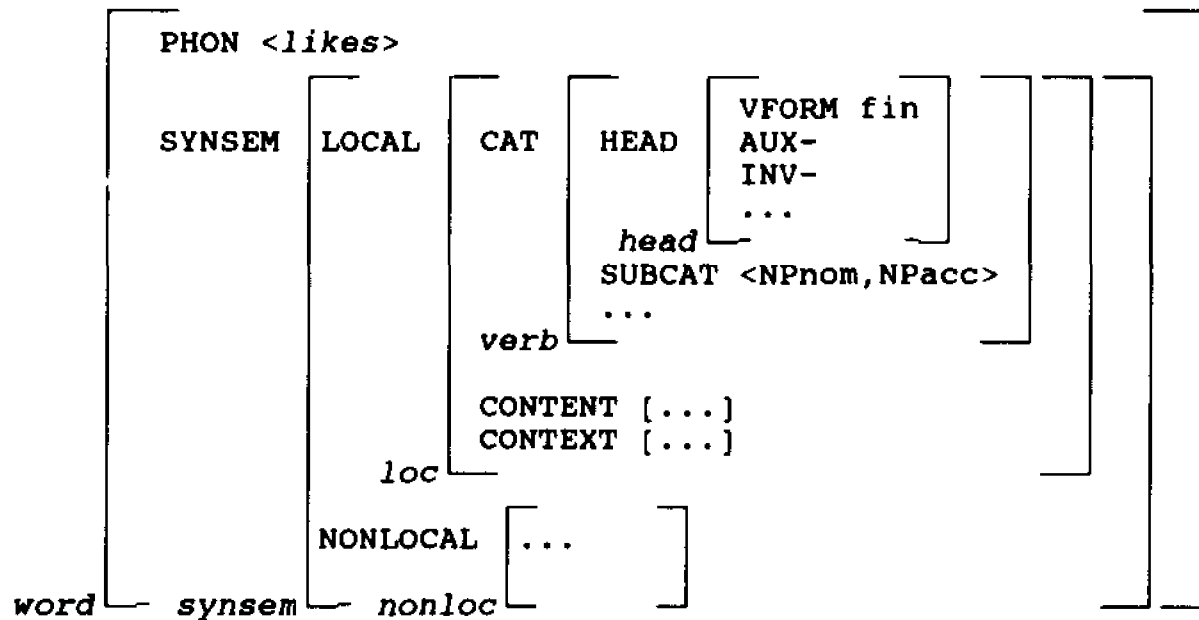
2.2 MONOTONICITY

(8) is a 'further specification' of (7) in that information is added to (7) but no information is removed or over-ridden.

7. An HPSG sign (= (2))



8. (= (3))



A basic assumption of HPSG is that all linguistic information is organized monotonically, so that information from a more general category is always inherited by a more specific category. An inheritance hierarchy is monotonic if every node lower on the hierarchy inherits all properties of all nodes higher on the hierarchy.

9. a. MONOTONIC INHERITANCE
- ```

 A
 / \
 A+ A-

```
- b. NON-MONOTONIC INHERITANCE
- ```

      A-
     / \
    A+  A-
  
```

Any non-monotonic information structure is ungrammatical in HPSG.

The organization of feature declarations for sorts is also monotonic in HPSG. The features PHON and SYNSEM are appropriate for *sign*, so they are both necessarily appropriate for *word* and *phrase*, which are subsorts of *sign*. But the attribute DTRS is also appropriate for *phrase*.

- 10.
- ```

 sign
 [PHON]
 [SYNSEM]
 / \
 word phrase
 [DTRS]

```

(10) is a monotonic inheritance hierarchy which represents the information that the features PHON, SYNSEM, and DTRS are appropriate for *phrase*, while only PHON and SYNSEM are appropriate for *word*.

Since HPSG assumes that monotonicity characterizes the organization of all linguistic information, I assume initially that the requirement that all linguistic information be organized monotonically is imposed on the learner by UG. However, in Chapter 5, I will argue that the existence of exceptional lexical subclasses and lexical

items constitutes evidence that at least some lexical information is not organized monotonically.<sup>7</sup>

### 2.3 SYNTACTIC CONSTRAINTS AND DECLARATIVITY

The central mechanism by which constraints are imposed on syntactic trees in HPSG is structure-sharing, in which two (or more) features (or paths) share an identical value. For example, the Head Feature Principle, which ensures that phrases are projections of their head daughters, is a universal principle which is stated in terms of structure-sharing:

11. a. HEAD FEATURE PRINCIPLE (HFP) (P&S-94:1(16))  
The HEAD value of any headed phrase is structure-shared with the HEAD value of the head daughter.
- b.           XP[SYNSEM|LOCAL|CAT|HEAD[1]]  
              | (head)  
              Y[SYNSEM|LOCAL|CAT|HEAD[1]]

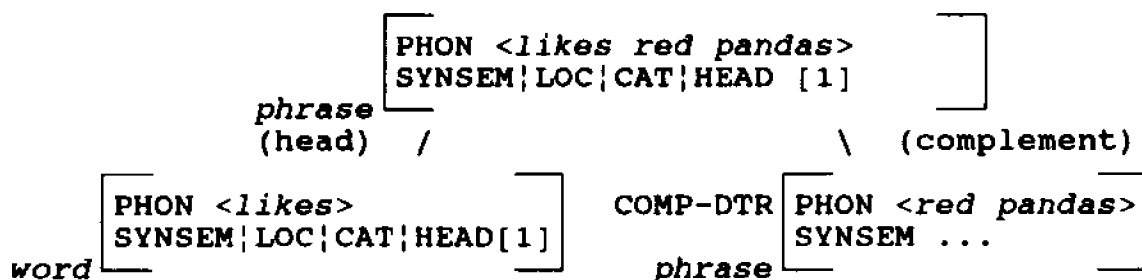
The tag [1] on the mother and the head daughter in (11.b) represents the requirement that the values of the two HEAD features are the same, that is token-identical, rather than type-identical. For example, the HFP is satisfied in (12.a) because both occurrences of [1] are understood as pointing to the HEAD features of *likes*. The value of the token to which the tag [1] points is shown in (12.b). Pointer values

---

<sup>7</sup>Green (1994:fn.2) suggests that "the available notion of inheritance (for children) is monotonic because it requires a much simpler calculus for inferences". But since children must acquire specific knowledge about the properties of particular nouns before they can acquire more general knowledge about the typical properties of nouns in their language, I do not assume that learners necessarily acquire all linguistic information monotonically.

can be conceived of as dangling from the phrase structure. A pointer is a kind of thread which connects a tag to the token that represents its value; each token will be tied to as many threads as there are pointers to that token.

12. a. (cf.(4))



b. [1] [VFORM fin,AUX-,INV-]

The use of structure-sharing, as opposed to movement, means that syntactic principles such as the HFP can be stated as constraints on well-formed trees. A local tree which satisfies all (relevant) constraints is grammatical.

Technically, all relevant constraints must unify.

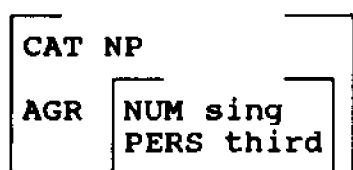
Unification is a single mathematical operation that merges feature structures if all of the information in them is consistent. For example, (13.a) unifies with (13.b) to form (13.c), which is then a more specific version of (13.a) and (13.b), and which is said . (This example was adapted from Shieber (1986a); the features are AGR(eement), NUM(ber), PERS(on), and GEN(der).)<sup>8</sup>

---

<sup>8</sup>This example provides a general illustration of how agreement works in HPSG. If two constituents, such as a subject and verb, agree, their agreement features must be consistent, or unify. A noun that is [NUM sing,GEN fem] may agree with a VP that is [NUM sing, PERS third] but not with a VP that is [NUM pl]. (Of course, the set of features of an N and a VP cannot

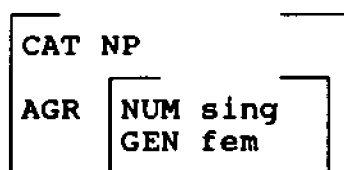
13.

a.



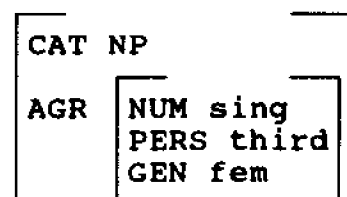
⊔

b.



=

c.



Formally, unification is a single operation, but conceptually it can be decomposed into two sub-processes; one checks that all information that is being unified is consistent, and the other copies information from each of the partially specified feature structures into the fully specific unified structure.<sup>9</sup> So the operation of unification can be understood as simultaneously checking that (13.a) and (13.b) are consistent and constructing (13.c), from which all information in (13.a) and (13.b) has been copied. Because unification formally combines consistency checking and feature building into one operation, unification will build (13.a), if only (13.b) and (13.c) are known; if only (13.a) and (13.c) are known, unification will build (13.b). In (14.a) and (14.b), the values of NUM are inconsistent. Unification fails because inconsistent information cannot be merged into a single fully specific feature structure, and so (14.c) is undefined.

---

unify because the features are inconsistent.)

<sup>9</sup>This conceptualization of unification was suggested by Mitchell Marcus (p.c.).



15. The Immediate Dominance (ID) Principle (P&S-94:399)  
Every headed phrase must satisfy exactly one of the ID schemata.

These ID Principle includes six schemas, which are disjunctive in that each local tree must be licensed by exactly one of the schemas. The number of schemas is relatively small because ID schemas are stated without regard to the linear order of constituents, which is determined independently by Linear Precedence specifications, and in most cases, without reference to syntactic category.<sup>11</sup>

Before introducing the six schemas, I turn to the Subcategorization Principle, which ensures that each complement for which a head is subcategorized is realized exactly once in a sentence structure. In HPSG, the value of SUBCAT consists of a list of the SYNSEM values of the head's complements. The SUBCAT value of *likes* can be represented as in (16), which shows that *likes* is subcategorized for a nominative subject NP and an accusative direct object NP; the tags stand for the SYNSEM value of the constituents for which the verb is subcategorized.<sup>12</sup>

---

<sup>11</sup>Of course, the schemas proposed in P&S-94 may not be the only ones that UG specifies; other schemas might be attested in languages not yet studied, and still others might be licensed by UG but not be used in any existing natural language.

From the point of view of learnability theory, it doesn't matter how many innate schemas there are (as long as they are disjunctive), since they don't need to be learned. But considerations of economy suggest that the number of innate schemas is probably quite limited.

<sup>12</sup>HPSG does not employ a notion of case assignment or theta assignment like that of GB; most or all of this work is done by subcategorization in HPSG.

16. likes [SUBCAT<[1],[2]>]  
 [1] NPnom; [2] NPacc

For expository convenience, it is a frequent practice to list only relevant features of SYNSEM, as in (17).

17. likes [SUBCAT<[NPnom],[NPacc]>]

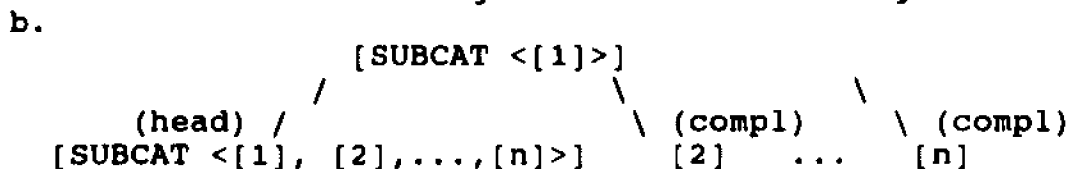
The Subcategorization Principle in essence checks off arguments as they are satisfied in local trees.

18. Subcategorization Principle (=P&S-94 1(15))  
 In a headed phrase (i.e. a phrasal sign whose DTRS value is of sort *head-struct*), the SUBCAT value of the head daughter is the concatenation of the phrase's SUBCAT list with the list (in order of increasing obliqueness) of SYNSEM values of the complement daughters.

The Subcategorization Principle thus operates in conjunction with the schemas, which license specific configurations of heads and complements in local trees.

For example, Schema 2, the Head-Complement Schema, licenses configurations with a lexical head daughter and with zero or more complements, thus eliminating the need for multiple phrase structure rules such as VP --> V; VP --> V,NP; N'--> N,PP; and PP --> P,NP.

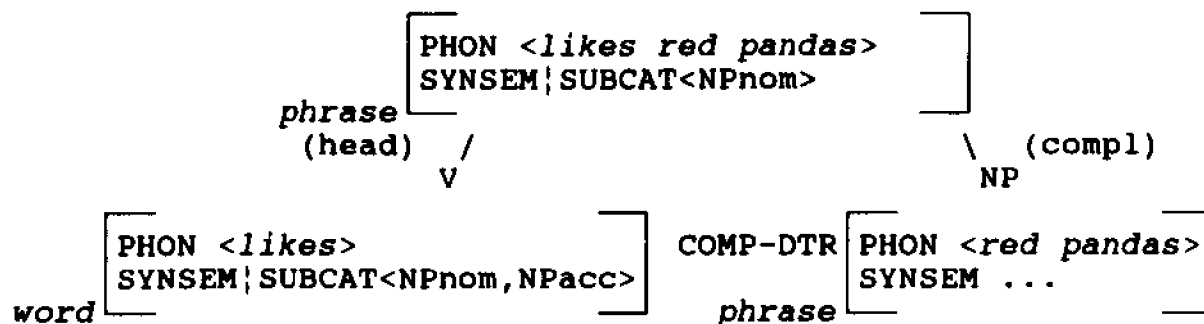
19. a. Schema 2 (= P&S 1(22))  
 An almost-saturated (SUBCAT list of length one) phrase with DTRS value of sort *head-comp-struct* in which the head daughter is a lexical sign.



The structure of the VP *likes red pandas* is licensed by Schema 2.

20. (cf. (12.a))

VP



The fact that the SUBCAT value of the mother node lists only the subject and not the direct object comports with the fact that the subcategorization requirement for the direct object is satisfied in this further specification of Schema 2.

Schema 1, the Head-Subject Schema, licenses configurations in which the subcategorization requirements of the mother node are fully satisfied (as indicated by SUBCAT<>), and a phrasal daughter which is subcategorized for a single phrase, as in old-style phrase structure rules such as  $S \rightarrow NP, VP$ , and  $NP \rightarrow DET, N'$ .<sup>13</sup>

21. a. Schema 1 (P&S 1(22))  
 A saturated ([SUBCAT <>]) phrase with DTRS value of sort *head-comp-struct* in which the HEAD-DTR value is a phrasal sign and the COMP-DTRS value is a list of length one.

b.

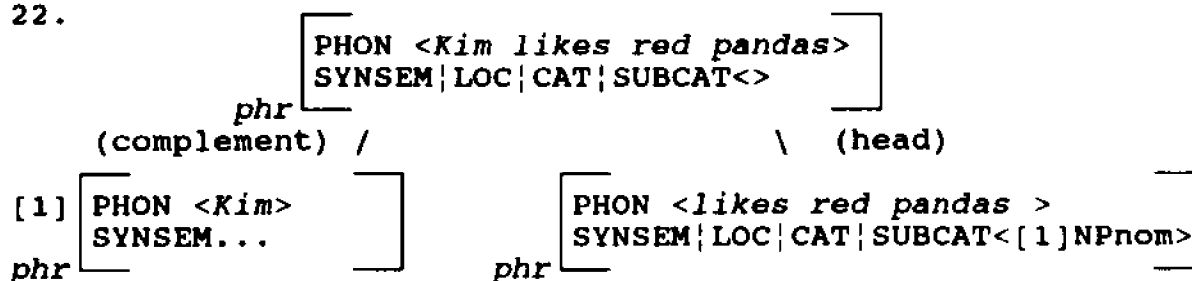
|             |                |
|-------------|----------------|
|             | [SUBCAT <>]    |
| (compl) /   | \ (head)       |
| [SYNSEM[2]] | [SUBCAT <[2]>] |

So the local tree in (22) is licensed by Schema 1.

---

<sup>13</sup>In P&S-94:Ch.9, two new attributes, SUBJ and COMP, are introduced (as suggested by Robert Borsley), thereby making the special treatment of subjects natural. Schemas 1 and 2 are accordingly revised, and the Subcategorization Principle is replaced by the Valence Principle.

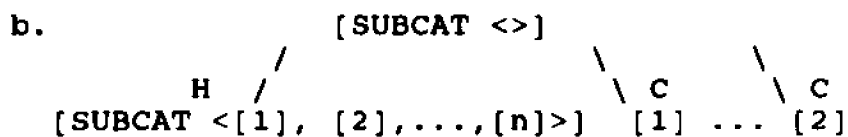
22.



The saturated SUBCAT value on the mother node is grammatical because all subcategorization requirements of the head daughter have been realized.

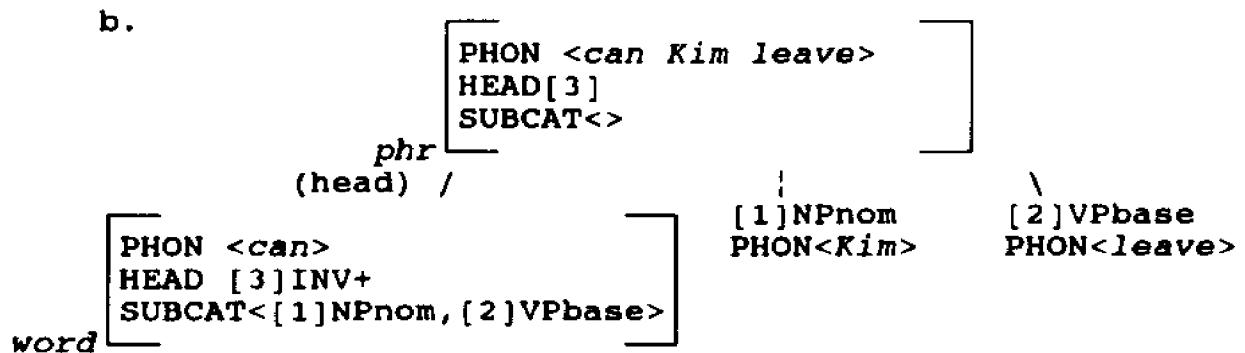
Schema 3 embodies P&S-94's claim that a single schema licenses verb-peripheral structures such as English subject-auxiliary inversion, scrambling in free word order language, and verb-final embedded clauses in German.

23. a. Schema 3 (=P&S-94 1(25))  
 A saturated ([SUBCAT <>]) phrase with DTRs value of sort *head-comp-struct* in which the HEAD-DTR is a lexical sign.



In English, Schema 3 licenses only subject-auxiliary inversion, for lexical items which are specified as [INV+]. The configuration in (24.b) is licensed by Schema 3.

24. a. Can Kim leave?



On this analysis, *can leave* is not a constituent of this sentence.<sup>14</sup>

P&S-94 propose that universal grammar licenses three other configurations, the Head-Marker Schema, for complementizers and heads, the Head-Adjunct Schema, for adjuncts and heads, and the Head-Filler Schema, for fillers and heads).<sup>15</sup> Schema 6, the Filler-Head Schema (which will also be discussed in Chapter 6), is an important component of the way that unbounded dependency relations, which are treated as movement in GB, are handled. One of the important innovations of GPSG was that long-distance relations could be represented by featurally marked path of nodes between a 'filler' (e.g., an interrogative or topicalized constituent) and its associated 'gap'.<sup>16</sup>

<sup>14</sup>See Appendix to Ch.6 for further discussion of inversion.

<sup>15</sup>Schema 5 will be discussed at the end of this chapter, and the Head-Marker Schema will be introduced at the beginning of Chapter 3.

<sup>16</sup>This notion is originally due to Gerald Gazdar (e.g., Gazdar 1981). See Jacobson (1987) for a comparison of this approach to the movement approach.

HPSG has three *nonlocal* features, QUE(stion) for interrogative expressions, and REL(ative) for relativizers, and SLASH, a feature which carries information about the LOCAL value of the filler between the filler and the associated gap. The value of SLASH is a variable whose value is the same as the LOCAL value of the gap. The LOCAL value is imposed, for example, by a verb on its object position.

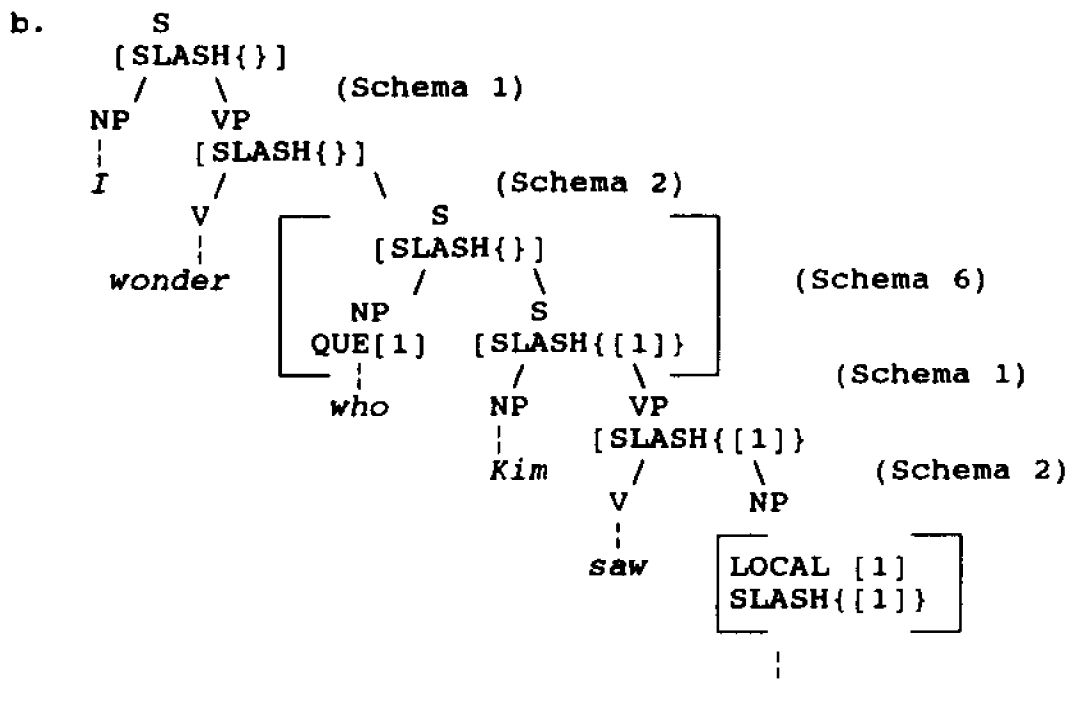
The local tree at the top of the UDC, a configuration with a filler whose LOCAL value is identical to the SLASH value on the head daughter is licensed by the version of Schema 6 in (25).

25. Schema 6 (simplified version)

$$\begin{array}{c} \text{YP} \\ / \qquad \qquad \qquad \backslash \text{ (head)} \\ \text{filler}\{\text{LOCAL}\{[1]\}\} \quad \text{XP}\{\text{SLASH}\{[1]\}\} \end{array}$$

All NONLOCAL features are set-valued because in some languages, multiple fillers and/or gaps must be carried on a single node. The empty set value of SLASH, {}, indicates that a node contains no gap. The bracketed tree in (26) is thus licensed by Schema 6.

26. a. I wonder S[who Kim saw \_\_\_]?



Technically, fillers structure-share the SYNSEM|LOCAL value of their sister, represented by [1]. So SLASH carries the LOCAL value of *who* between the filler and the gap.<sup>17</sup>

The local features of the trace position and those of the filler are both identical with the local structure passed up as a SLASH value, and therefore by the transitivity of equality they are identical with each other. In other words, the filler and the trace are the same thing, at least as far as their local structure is concerned. Thus an effect similar to that of *wh*-movement is achieved, without actually moving anything.

P&S-94 pp.162-163

To prevent NONLOCAL features from passing indefinitely up the tree, P&S must ensure that the SLASH is 'bound off' when it is instantiated on the sister of the associated filler. This rather simple requirement is complicated

<sup>17</sup>P&S-94:Ch.9 introduces the notion of UDCs without traces; this idea is developed further in Sag and Fodor (1994).



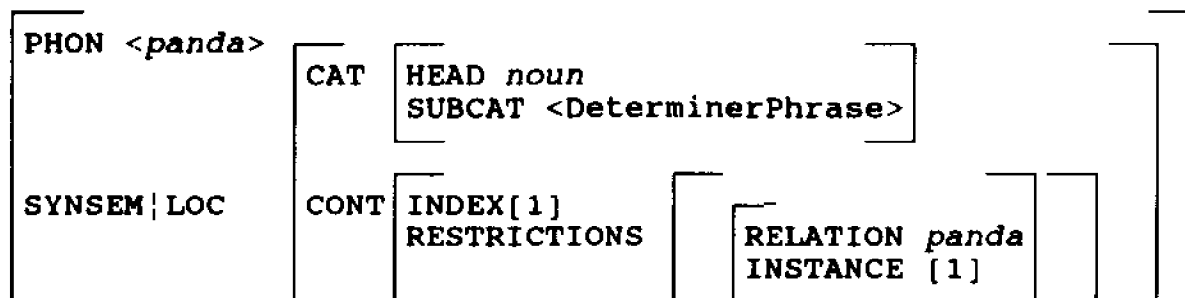
NONLOCAL Feature Principle, another universal constraint on all LOCAL trees.

29. NONLOCAL FEATURE PRINCIPLE (P&S-94,4(16))  
 For each nonlocal feature, the INHERITED value on the mother is the union of the INHERITED values on the daughters minus the TO-BIND value on the head daughter.

This principle licenses local trees in which the mother node structure-shares the (non-empty) SLASH values of a non-head daughters.<sup>18</sup>

In order to round out this discussion of the universals of HPSG, I introduce briefly P&S-94's treatment of semantics, which is based upon the theory of situation semantics (P&S-94,Ch.8). The CONTENT of the noun *panda* might be represented as follows, assuming that nouns are subcategorized for determiners; INDEX is a reference marker or a parameter introduced by an NP in situation semantics; the tag [1] indicates that a particular panda is being referred to.

30.



Adjectives are treated as adjuncts in P&S-94; they specify the SYNSEM value of the head that they modify via the

<sup>18</sup>I will now ignore the distinction between NONLOCAL|TO-BIND and NONLOCAL|INHERITED until Ch. 6, where I will propose a revision to Schema 6 that eliminates certain non-universal aspects of P&S-94's formulation.

feature MOD(IFY). Schema 5 licenses the combination of an adjunct with the head that it modifies.

31. Schema 5: Adjunct-Head Schema (P&S-94(Ch.1(60)))  
 A phrase with DTRS value of sort *head-adjunct-structure* (*head-adj-struct*) such that the MOD value of the adjunct daughter is token-identical to the SYNSEM value of the head-daughter.

[MOD N'] embodies the requirement that an adjective combine with a head noun; in contrast, adverbs modify (at least) verbs and/or perhaps VPs, adjectives and adjectives, as in (32).

32. a. Kim ADV[happily VP[went to the zoo]].  
 b. A ADV[very ADJ[happy]] Kim went to the zoo.  
 c. Kim went to the zoo ADV[very ADV[happily]].

The feature MOD will be discussed further in Chapter 7.

The lexical entry for the adjective *red* is in (33); I include part of the CONTENT value in order to introduce P&S-94's general approach to semantics.

33.

|            |      |                     |        |                 |  |  |  |  |
|------------|------|---------------------|--------|-----------------|--|--|--|--|
| SYNSEM LOC | CAT  | HEAD                | MOD N' | INDEX [1]       |  |  |  |  |
|            |      | <i>adj</i>          |        | RESTRICTIONS[2] |  |  |  |  |
|            | CONT | RELATION <i>red</i> |        |                 |  |  |  |  |
|            |      | ARG [1]             |        |                 |  |  |  |  |

(33) specifies that the adjective *red* must combine syntactically with an N', whose INDEX is [1]. In terms of content, *red* must combine with that [ARG [1], the token-identical N', thus imbuing redness on it.

The Semantics Principle specifies how semantic information from different constituents is combined.



addition, language-particular constraints must also be satisfied. The central problem of this dissertation is how learners can acquire these language-particular constraints. I noted above (p.32) that P&S-94 specify three components of language-particular grammar: the lexicon, further specification of the universal ontology (linguistic entities), and further specification of the universal ID schemas. As also noted above, I assume that there is only a universal linguistic ontology (a set of features and sorts) so that learners do not face the problem of determining what language-particular features their language uses. Therefore language-particular lexical entries and language-particular further specifications of the schemas are what must be acquired by learners. In what follows, I present a model of language-learning in which acquiring the apparently syntactic language-particular further specifications of the schemas is an integral part of the process of acquiring lexical entries.

#### 2.4 CONCLUSION

One of the methodological assumptions underlying this research is that in the absence of full knowledge of UG, a carefully conceptualized model for organizing linguistic information is the most fruitful starting place for the development of a learning model. The systematic and formally explicit architecture of HPSG makes it an optimal framework for my research.

In what follows, I assume the syntactic theory of P&S-94 (Chs.1-8) unless otherwise noted. I do this as a practical matter, to make my model as accessible to readers as possible, and I modify this theory only when necessary for successful learning. In this chapter, I make two initial modifications: first, I assume that all universal principles are innate, while P&S-94 do not; second, I assume the UFAP, which guarantees that the learner's feature structures are totally well-typed. The general approach that I propose in this dissertation will continue to work, even if modifications are made to HPSG's syntactic theory, unless the revisions render HPSG unlearnable in the ways discussed in Chapter 1.

**CHAPTER 3**  
**ACQUIRING THE SELECTIONAL PROPERTIES OF COMPLEMENTIZERS**

In this chapter, I show how learners can acquire the selectional properties of complementizers from positive evidence. I start by reviewing the learning model developed by Fodor for GPSG; in this model, GPSG's syntactic defaults play a crucial role in acquisition of correct syntactic rules. Then I turn to the question of how this model can be adapted to HPSG, in which by hypothesis all language-particular constraints are lexical, and in which there are no syntactic defaults. I first consider the problem of acquiring language-particular further specifications of Schema 4, the Marker-Head Schema. In particular, I consider the question of how the selectional properties of complementizers, which P&S-94 characterize as a kind of marker, can be acquired. P&S-94 postulate a lexical feature SPEC, by which complementizers specify properties of the head node to which they are sensitive. I show that Fodor's approach to acquisition of syntactic rules can be naturally adapted to lexical learning in HPSG.

Then I address the question of what *requires* the learner to acquire the language-particular further specification of Schema 4. Conservative learning suggests that learners would never bother; crucially, it is the fact that SPEC, whose value is structure-shared with the SYNSEM value of the head daughter, is obligatory for markers that

is responsible for the fact that learners do not fail to acquire its value.

I turn next to the problem of how the learner can generalize correctly over lexical entries in a default-free learning model. If the learner omits features before its learning is complete, the acquired lexical entries will license ungrammatical constructions. This means that the learner must store every feature value that is in the input; however, the cost of doing this is undergeneralization. However, omission of default values produces more concise, linguistically plausible generalizations.

Finally, I consider the problem of generalizing over SPEC values which are stored in the lexical entries for complementizers. Even on the (false) assumption that the learner waits until its knowledge is complete, generalization by collapsing of feature structures licenses ungrammatical constructions. I show that part of the disjunction problem can be solved by what I call 'conservative disjunction'; however, even with conservative disjunction, the stored SPEC values will still be duplicative, and much of the duplication can be attributed to default values.

I conclude that Fodor's syntactic learning model can be easily adapted to lexical learning of the selectional properties of complementizers in HPSG. However, the absence of defaults leaves the learner unable to acquire correct generalizations.

### 3.1 THE GPSG MODEL

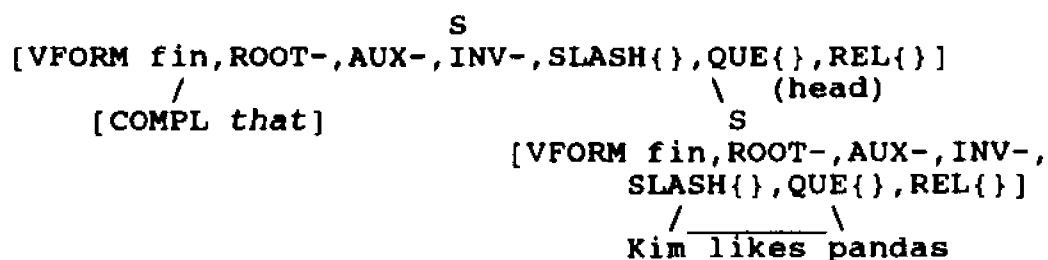
In this section, I discuss the model of learnability developed for GPSG, HPSG's predecessor theory, in Fodor (1989a, 1990, 1992a, 1992b), Fodor and Crain (1990) and Maxfield (1990). One of Fodor's pioneering insights is that if the learner's syntactic knowledge is conceived of as being represented by feature structures, the learner can derive syntactic generalizations directly from the feature structures of fully specific local trees. To see how the learning process works, consider (1.b), which I assume is the fully specified local tree of the embedded clause in (1.a).<sup>1</sup>

1. a. I think S[*that* Kim likes pandas]

---

<sup>1</sup>I have made no attempt to represent all relevant features; rather I have selected features that show the nature of the learning problem.

Also, since most of the technical differences between features in GPSG and HPSG are not relevant here, I use the features of P&S-94 in a GPSG-like structure. Except for ROOT, which is not used in GPSG or HPSG, all other features in this section, and elsewhere in this dissertation, are used in P&S-94 unless indicated otherwise. The need for a feature such as ROOT is mentioned in Maxfield (1990).

b.<sup>2</sup>

The feature structure of (1.b) is exactly equivalent to the phrase structure rule in (2).

2. S[VFORM fin,ROOT-,AUX-,INV-,SLASH{}],QUE{}],REL{}]  
 --> [COMPL that],S[VFORM fin,ROOT-,AUX-,INV-,  
 SLASH{}],QUE{}],REL{}]

(2) constitutes what I term 'generalization by type' in that it licenses not only the embedded clause in (1.a), but also a potentially infinite number of clauses of the same type. If the learner stores (2), then it has acquired a syntactic generalization directly from positive evidence. But although (2) is expressed as a phrase structure rule, it is important to note that generalizations need not be rules. The word *apple* does not constitute a generalization, but the information that *apple* is a singular count noun does constitute a generalization: the singular count noun *apple* can be used in any context that allows a singular count noun.

---

<sup>2</sup>As noted in Ch.2, fn.6, AUX and INV are not universal properties because they are characterized by different properties in different languages. In order to be learnable, they should be replaced by features that represent identical properties in all languages. However, since my goal is to model the acquisition of linguistic generalizations represented by complex feature structures, the question of what features would be linguistically preferable, I will not speculate on this topic; I note only that it is not clear whether treating AUX and INV should be treated as head features.

Fodor also pointed out that a fully specific (correct) local tree of the target language, such as (1.b), and its equivalent rule (2), can be factored into two components, the universal and the language-particular.

3. FULLY SPECIFIED LOCAL TREE IN TARGET LANGUAGE

/                    \  
UG                    LANGUAGE-PARTICULAR GRAMMAR

Given the working assumption that universal principles are innate, and assuming that the set of universal specifications and language-particular specifications are disjoint, the learner's task is to perform a second kind of generalization which I call 'generalization by feature omission': the learner must factor out those properties of the local tree which are predictable because they are specified by UG; the residue, which is language-particular, must be stored.

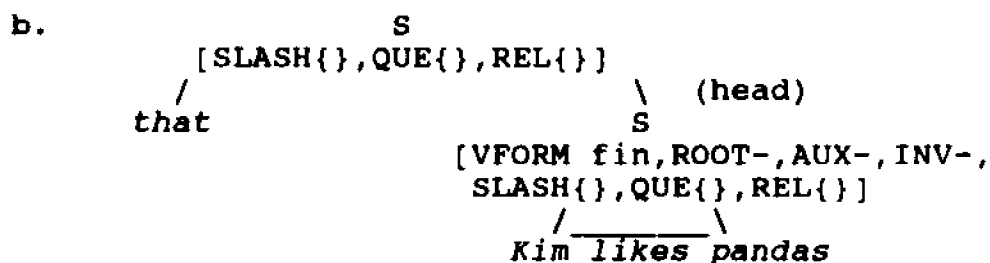
4. Fodor's formula (=Ch.1(8))  
FULLY SPECIFIC TREE - UG = LANGUAGE-PARTICULAR RULE

The HFP<sup>3</sup> (discussed in Chapter 2) allows the learner to strip off of the local tree in (1.b) those features of the mother node that are identical to those of the head daughter, as in (5.b).

5. a. I think S[that S[Kim likes pandas]]

---

<sup>3</sup>The Head Feature Principle was called the Head Feature Convention in GPSG.



I list the features on the head daughter rather than the mother node because I assume that these head features are in fact stored in the lexicon, and projected up through the tree.

On the assumption that there are default feature values which are universal and are pre-specified in UG, they too can be stripped off of the local tree. The reasons for stripping off default values, and some problems that arise in the process are discussed in Maxfield (1990), and will also be discussed below. For the moment I assume the default values in (6). (Criteria for determining default values will be discussed in Ch.4.)

6. Tentative Default Values (u = unmarked)
- |                         |                         |
|-------------------------|-------------------------|
| [u VFORM = VFORM fin]   | [u SLASH = SLASH {}]    |
| [u ROOT = ROOT +]       | [u QUE(ATION) = QUE {}] |
| [u AUX(ILIARY) = AUX -] | [u REL(ATIVE) = REL {}] |
| [u INV(ERSION) = INV -] |                         |

(7.a), which can also be expressed as the phrase structure rule in (7.b), is equivalent to (5.b), except that predictable default values have been stripped off.

Generalization by feature omission allows the reader to store rules that are very compact relative to the fully specified rule because predictable information, including

that specified by universal principle and by default values, has been stripped off.<sup>4</sup>

7. a.           XP  
       /                               \  
      [COMPL that]                   S (head)  
                                           \  
                                           [ROOT-]
- b.   XP --> [COMPL that], S[ROOT-]

(7) is a language-particular syntactic rule which the learner must store. This rule was acquired directly from the local tree without the need for evaluation of preferable grammars, and without the need to compare the sentences produced by alternative grammatical generalizations.

Note that if the defaults were not universal, but learned, there is no guarantee that the learner would acquire defaults that would produce a compact syntactic rule. Language-particular generalizations are actually learner-particular unless some mechanism ensures that all learners will acquire identical defaults.

However, by GPSG convention, (7) licenses all values of all features not mentioned in the rule. For example, since (7) does not mention VFORM, it licenses [VFORM inf] as well as [VFORM fin]. Nothing in (7) makes (8) ungrammatical.

8. \*I think that Kim to like pandas.

To solve this problem, Fodor (1992b) proposed the Specific Defaults Principle.

---

<sup>4</sup>In Ch.4, I will argue that the default value for the feature ROOT is [-] in the context of a complementizer, so that in fact [ROOT-] does not need to be stored either.

9. **Specific Defaults Principle (Fodor 1992b:17)**  
 A specific (i.e. non-disjunctive) default value must be assigned by UG to every feature in every context, unless the value in that context is universally fixed, or is universally free.

The Specific Defaults Principle forces the learner to obey the Subset Principle by providing a specific interpretation of the omission of a feature value from a rule: if a generalization does not mention a feature, only the default value is licensed. The Specific Defaults Principle allows the learner to acquire a grammar from positive evidence because it supplies the learner with an initial expectation of what the grammar will be like. If the default value is incorrect, the input sentence will not be licensed by the learner's grammar. Positive evidence thus informs the learner that it needs to modify its grammar, and it replaces the value that does not license the construction with one that does in accord with the principles of conservative learning discussed in Fodor (1992).<sup>5</sup> I frequently refer to a correct feature value that licenses the target construction as an 'attested' value.

The Specific Defaults Principle, which supplies the learner with a specific hypothesis about what the grammar is like, works in conjunction with the Uniqueness Principle, which requires that the learner interpret the absence of a

---

<sup>5</sup>This is not to imply that the first replacement value that the learner tries is necessarily the correct one. Recall from Ch.1 that I set aside the problem of how string-to-structure learning is achieved.

feature in a phrase structure rule as signifying that *only* the default value is grammatical. This is guaranteed by the Uniqueness Principle.

The principle is that the need for negative evidence in language acquisition can be eliminated if the child knows that when he or she is faced with a set of alternative structures fulfilling the same function, only one of the structures is correct unless there is direct evidence that more than one is necessary. ...the principle... allows the learner to discard structures even when there is no negative feedback indicating that they are grammatical...

(Pinker 1984:113)

If the learner encounters positive evidence that the specific default value is incorrect, the Uniqueness Principle requires the learner to replace the default value with exactly one marked value: the learner cannot interpret the non-specification of a feature value in a rule as equivalent to a disjunctive value. If the learner mistakenly assumes that the marked value is grammatical, positive evidence will tell the learner that it has made a mistake and that the unmarked value should be replaced by a single marked value. The learner may list disjoint values in the rule only if it has encountered positive evidence that both (or all) values specified in the disjunction are grammatical. The requirement that learners prefer storing singleton values to storing disjunct values penalizes disjunction, which is representationally more costly than singleton values are.

An advantage of this model is that the link between the evidence set -- the sentences of the language -- and the

learner's syntactic knowledge is simple and direct. Moreover, since all language-particular constraints are read directly off of fully specific trees, these learned constraints take the form of positively stated feature structures, as required by the Subset Principle. Note that this learning model accounts naturally and simply for dialect differences: the learner will acquire grammatical rules that license exactly those constructions it encounters in the target language.<sup>6</sup>

The general picture developed by Fodor for GPSG can be adapted for HPSG, as I show below. However, two important differences between GPSG and HPSG have major implications for the HPSG learning model. First, the GPSG model involves acquisition of syntactic rules; in the move from a rule-based approach to a principle-based approach, HPSG eliminated language-particular syntactic rules.

(In both HPSG and GB) structure is determined chiefly by the interaction between highly articulated lexical entries and parameterized universal principles of grammatical well-formedness, with rules reduced to a handful of highly general and universally available phrase structure (or immediate dominance) schemata.

P&S-94:2

Since language differences are essentially lexical, a learning model consistent with HPSG theory will include lexical learning rather than syntactic learning. The learner will need to acquire lexical entries and language-particular further specification of the schemas, but it will

---

<sup>6</sup>See fn.27 for an example.

not need to construct phrase structure rules from scratch as in GPSG. Second, defaults were used in GPSG to account for semi-predictable feature values which could not be imposed by absolute universal principles because of the existence of marked exceptions. However, certain computational problems with defaults, which I will discuss in Chapter 4, led to the elimination of syntactic defaults from HPSG.<sup>7</sup>

In what follows, I show how Fodor's model can be adapted to HPSG, with lexical learning and without defaults. I start by showing, in the next section, that the further specification of Schema 4 is stored lexically in HPSG and therefore can be acquired by lexical learning. Then, I consider the problem of acquiring correct generalization in a feature-based model in which there are no defaults.

### 3.2 PROPERTIES OF COMPLEMENTIZERS IN HPSG

To show how the selectional syntactic properties of complementizers can be stored in the lexicon, I begin with a discussion of the properties of complementizers in HPSG. P&S-94 treat complementizers as a subsort of a part of speech called 'marker', "so-called because it formally marks the constituent in which it occurs".<sup>8</sup> P&S-94:45

---

<sup>7</sup>Green and Morgan (to appear) question the need for defaults in GPSG and HPSG.

<sup>8</sup>To avoid undue confusion with concepts in the theory of markedness (defaults), I use [MARKING null] rather than [MARKING unmarked] to refer to a constituent which is not a marker, and [MARKING lex] rather than [MARKING marked] to refer to a constituent which is a marker. [MARKING lex] refers to a marker with a non-null value, without specifying which marker it is. I follow

characterize markers as "'functional' or 'grammatical' as opposed to substantive, in the sense that (their) semantic content is purely logical in nature, perhaps even vacuous". As such, complementizers are prototypical examples of functional categories whose syntactic properties have to be learned independently of meaning.<sup>9</sup> In addition to complementizers, P&S suggest that the comparative words *than* and *as* are also markers.<sup>10</sup>

The properties of complementizers are captured by two universal constraints: Schema 4 and the SPEC Principle.

Schema 4, which is the only schema that licenses a marker

---

P&S 94 in using [MARKING *compl*] to refer to a marker of subsort *complementizer*. Robert Levine (p.c.) points out that the ontological status of the value of MARKING is unclear; I follow P&S-94 in representing the MARKING value of individual markers with their spelling.

<sup>9</sup>In some cases, complementizers do seem to make a logical contribution to the semantics of the sentence, as shown in (i).

- i. (=Bresnan 1970 (10))
- a. It may not distress John *for* Mary to see his relatives.
  - b. It may not distress John *that* Mary sees his relatives.
- (i.b) presupposes that Mary actually sees John's relatives, but (i.a) does not. However, the difference between the sentences may not be the result of the semantics of the complementizer but rather of the non-finiteness of the infinitive and the finiteness of the finite form.

But whether or not complementizers have meaning, Robert Levine (p.c.) points out other markers establish different truth conditions (on the assumption that *if*, *as*, and *whether or not*, are markers).

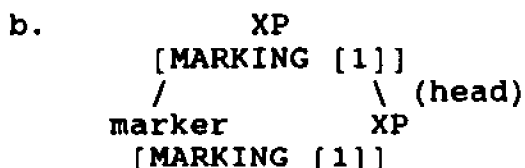
- ii. a. *If* Robin is a spy, we're all in big trouble.  
 b. *As* Robin is a spy, we're all in big trouble.  
 c. *Whether or not* Robin is a spy, we're all in big trouble.

<sup>10</sup>P&S-94 also mention the case-marking postclitics of Japanese and Korean as another kind of marker. These markers differ from complementizers in that they assign case to NPs rather than marking a set of properties of S or VP.

daughter, stipulates that the MARKING value of the (non-head) marker daughter is structure-shared with its mother.

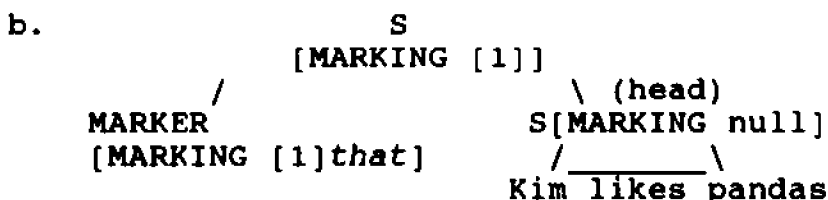
10. SCHEMA 4 (P&S-94 1.51)

- a. A phrase with DTRS value of sort *head-marker-structure* whose marker daughter is a marker with MARKING value token-identical to that of the mother.



Schema 4 requires that the value of MARKING be structure-shared by the mother node and the non-head (marker) daughter (whether that value is *lex* or *null*), as in (11.b).

11. a. I think [that [Kim likes pandas]]. (=1.a)



The value of MARKING on the lower S is null in (11.b).<sup>11</sup>

The feature MARKING is distinguished in that its value on the non-head daughter is structure-share their value with the mother node. A marker cannot be a head daughter because the head features of the marker's sister constituent must be subject to the HFP. For example, the finiteness of a clause is due to the finiteness of its verb, which must be represented on the higher S node so that it is accessible to the

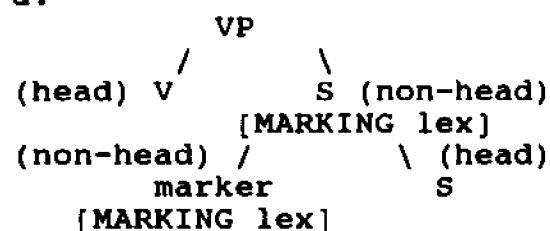
---

<sup>11</sup>P&S-94 (Ch.1,fn.51) speculate that [MARKING null] is introduced on the head-daughter either by a special principle or by each schema except Schema 4. In Ch.4, I argue that [MARKING null] is a context-free default value.

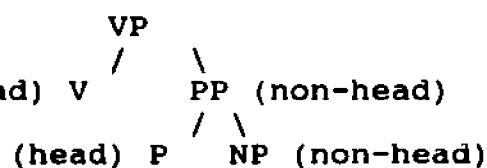
selecting matrix verb. The significance of Schema 4 can be shown by the contrast between (12.aa) and (12.ab).

12.

a.



b.



Schema 4 ensures that a matrix verb has access in its local tree to the MARKING value of the non-head daughter of a complement S in (12.aa).<sup>12</sup> That information as to whether or not the embedded clause has a complementizer is needed by the higher verb can be seen from (13).

13. I think *that*/\**whether*/\**where* you went.

The HFC ensures that the matrix verb has access to the head features of S. By contrast, a verb that is subcategorized for a PP, as in (12.bb), will have access only to LOCAL features of the head of PP; it will have no access to features of the PP's non-head daughter NP in the local tree.

In order to ensure that a marker takes as its sister only a node whose features match those that the marker is specified for, P&S-94 stipulate that a feature SPEC is appropriate for all markers. The SPEC Principle requires that the value of SPEC be the same as the SYNSEM value of the sister.

---

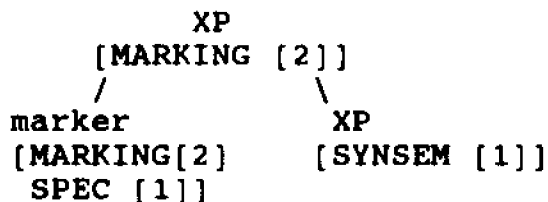
<sup>12</sup>As noted in fn.8, the ontological status of the value of MARKING is unclear.



their N' sisters via SPEC. As discussed in Chapter 2, adjuncts have an analogous feature, MOD, by which they select the SYNSEM value of the head that they modify (though MOD differs from SPEC in that it is borne by phrases as well as by lexical items).<sup>14</sup> Categories for which SPEC and MOD are appropriate thus resemble heads in one important respect: they impose selectional constraints on the SYNSEM value of their sister nodes (heads do so via SUBCAT) but they are unlike heads in that they do not project phrases or license arguments.<sup>15</sup>

Schema 4 and the Specifier Principle cannot be folded together into a single principle because not all constituents with the feature SPEC are markers. Although SPEC is appropriate for determiners, MARKING is not. Local trees with complementizer daughters must be licensed by both Schema 4 (10) and the Specifier Principle (14); these can be unified as in (17), which, in GB terms, is roughly equivalent to S' and S (or CP and IP).

#### 17. Unification of Schema 4 and the SPEC Principle




---

<sup>14</sup>On P&S-94's view, markers are different from adjuncts and heads in that they do not contribute to the semantics of the phrase (but see fn.9).

<sup>15</sup>MOD will be discussed in Ch. 7.

Given (17), the language-particular feature values that the learner needs to acquire are represented by the tags [1] and [2].

An important question then is this: what mechanism forces learners to refine Schema 4 by imposing on it the language-particular constraints represented by [1] and [2] in (17)? Recall from Chapter 1 that the tendency of learners to acquire compact generalizations that license ungrammatical constructions must be countered by some mechanism that forces the learner to prefer more specific feature structures over more general ones, in accordance with the Subset Principle. The particular question at hand is really part of a more general one: why don't learners violate the Subset Principle, and *never* refine the UG schemas? Something about the mechanism of syntactic learning must impel the learner to acquire arbitrary syntactic distinctions not associated with meaning.

Fodor (1989b) considered possible complications of the learning mechanism and concluded that a solution that requires learners to have a special piece of machinery for explicit calculation of subset relations is unconvincing.

Instead, where learners are observed to engage in interestingly restricted patterns of generalization, we should attempt to fix up our theory of UG.

(Fodor 1989b:140)

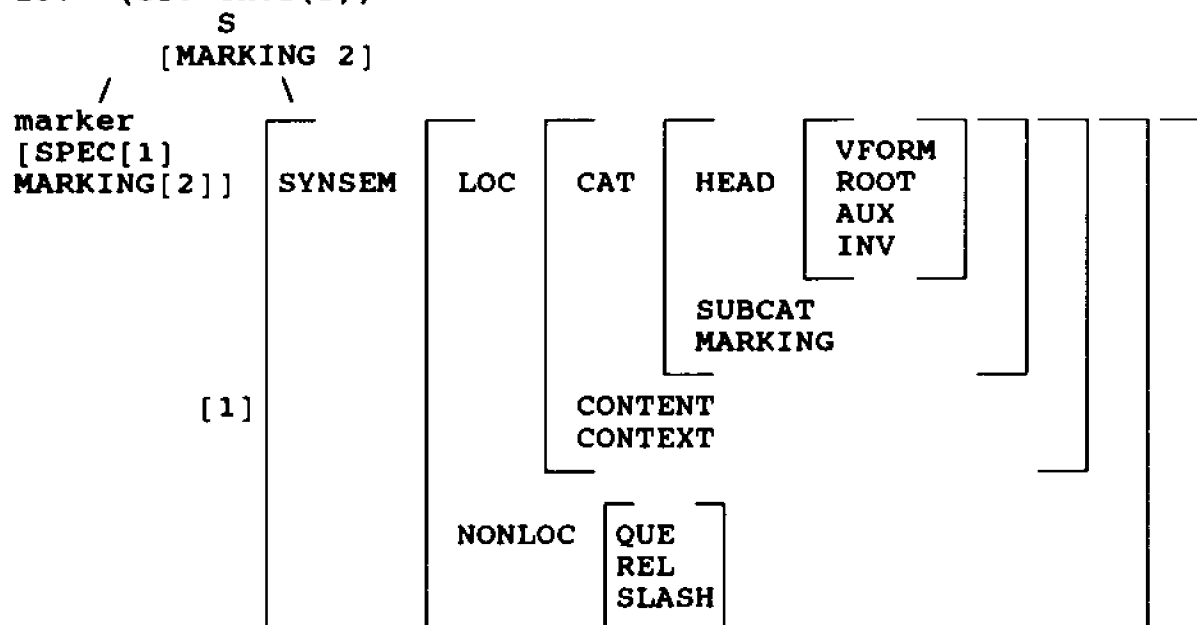
In HPSG, the mechanism that requires that the learner acquire the refinements of Schema 4 is in fact a linguistic one. Markers obligatorily have non-empty values for SPEC

and MARKING, as required by the UFAP (Ch.2(6)). Thus there needn't be any Schema-4-for-English at all. The features of the marker daughter that license properties of the head daughter are indeed stored in the grammar; but since the marker daughter is a lexical item, not a phrase, the language-specific properties of this construction will be recorded exclusively in the lexicon. And, as will be shown below, strictly local defaults that are confined to the lexicon are computationally tractable, and facilitate correct generalization over lexical entries.

### 3.3 FEATURES THAT COMPLEMENTIZERS ARE SENSITIVE TO

In this section I show that complementizers are sensitive to the SYNSEM values of the head daughter of Schema 4. The fact that a complementizer in even one language is sensitive to the value of some feature is evidence that the value must be included in SPEC. The fact that these values vary from language to language means that language-particular learning is required.

18. (cf. Ch.2(2))



Since all relevant information is available in the local tree and instantiated on the SYNSEM value of the sister of the complementizer, it is accessible to the SPEC value. A learner who acquires the value of SPEC for a particular complementizer will have obtained all necessary about the selectional properties of complementizers.<sup>16</sup>

For example, complementizers are sensitive to the value of ROOT; in English, complementizers specify [ROOT-] but not [ROOT+].

19. a. \*[That S[Tom saw Mary]]. S[ROOT+]  
 b. I think [that S[Tom saw Mary]]. S[ROOT-]  
 c. [That S[Tom saw Mary]] disturbed me. S[ROOT-]

<sup>16</sup>In my working examples, I limit myself primarily to syntactic features in SYNSEM|LOCAL|CAT and in SYNSEM|NONLOCAL. Markers may be also be sensitive to LOCAL|CONTENT (which contains information that contributes to semantic interpretation) and LOCAL|CONTEXT (which includes information that contributes to presupposition and conventional implicature), but I will consider these only when necessary, e.g., when discussing inversion.

20. a. \*[For S[Tom to see Mary]]. S[ROOT+]  
 b. I hope [for S[Tom to see Mary]]. S[ROOT-]  
 c. [For S[Tom to see Mary]] would disturb me. S[ROOT-]

But the restriction of complementizers to root clauses is not universal, as the example of Latin (mentioned in Chapter 1 and repeated here) and German show.<sup>17</sup> The Latin complementizer *utinam*, which introduces wishes, may be used in embedded and matrix clauses; it selects verbs which are subjunctive.

21. Latin (Atticus 3,3, cited in Lane (1903, par.1541)  
 (=Ch.1(6))  
 [Utinam illum diem [ROOT+]  
 Would that that-ACC-SING-MASC day-ACC-SING-MASC  
 videam]  
 see-PRESENT-SUBJUNCTIVE-1PERS-SING  
 'I hope that I may see that day.'

22. German  
 Dass er so gemein ist! [ROOT+]  
 'That he is so mean.'  
 '(I am surprised) that he is so mean.'

(23) shows that complementizers are sometimes grammatical in root clauses, given arguments by McCloskey (1979) that Irish *aL* is a complementizer.

23. Irish (McCloskey 1979 ch.1 1a)  
 [Cé aL dhíol an domhan]? [ROOT+]  
 who COMP sold the world  
 'Who sold the world?'

Complementizers are sensitive to the category of their sister: in standard English, *for* is specified for S[inf] and

---

<sup>17</sup>Alternatively, (22) could be analyzed as an embedded clause with an implicit root clause.

not VP[inf], but *that* is specified for S[fin] and not VP[fin].<sup>18</sup>

24. a. [For S[Tom to see Mary]] was a surprise. S[VFORM inf]  
 b. \*[For S[to see Mary]] was a surprise. VP[VFORM inf]  
 c. [To see Mary] was a surprise.
25. a. \*Who do you hope [that S[*she* saw]]? S[VFORM fin]  
 b. \*Who do you hope [that VP[saw her]]? VP[VFORM fin]  
 c. Who do you hope VP[went]? VP[VFORM fin]

Complementizers are also sensitive to the value of VFORM.

26. a. [That S[Tom saw Mary]] is disturbing. S[VFORM fin]  
 b. \*[For S[Tom saw Mary]] is disturbing. S[VFORM fin]  
 c. \*[That S[Tom to see Mary]] is disturbing. S[VFORM inf]  
 d. [For S[Tom to see Mary]] is disturbing. S[VFORM inf]  
 e. I demand [that S[Tom talk to Mary]]. S[VFORM base]  
 f. \*I demand [for S[Tom talk to Mary]]. S[VFORM base]

But in some dialects of Irish, the complementizer *aL* optionally selects a "relative" verb form that other dialects no longer have (McCloskey 1979, Ch.1(12a-b)). McCloskey presents convincing arguments that *a(L)* is a complementizer rather than a relative pronoun.

27. a. An t-iascaire a dhíolas a bhád  
 The fisherman COMP sells(+rel) his boat  
 b. An t-iascaire a dhíolann a bhád

---

<sup>18</sup>*whether* and *if*, which I assume are markers, also have different selectional properties.

- i. a. I wonder *whether* to go.  
 b. \*I wonder *if* to go.
- ii. a. I wonder *whether* Kim went.  
 b. I wonder *if* Kim went.

This is another example of the lexical variation that typifies SPEC.

The fisherman COMP sells(-rel) his boat

In standard English *that* is ungrammatical with inversion.<sup>19</sup>

28. a. I think [*that* Tom will see Mary]. [INV-]  
 b. \*I think [(*that*) will Tom see Mary]. [INV+]

But in Belfast English, inversion is optional in embedded interrogatives, but only if the complementizer *that* is not present.

29. Henry (1992(42))  
 I wonder what street does he live in. S[QUE{{1}},INV+]
30. Henry (1992(46))  
 \*I wonder S[what street *that* does he live in].  
 S[QUE{{1}},INV+]
31. I wonder S[what street *that* he lives in].  
 S[QUE{{1}},INV-]

Standard English allows only one marker per clause.

32. English  
 a. \*I think [*that* S[*that* Kim likes pandas]].  
 S[MARKING lex]  
 b. I think [*that* S[Kim likes pandas]].  
 S[MARKING null]

However, the requirement that there only be a single marker per clause is not universal, at least on the assumption that *bevor* in Swiss Bernese is best treated as a marker.<sup>20</sup>

---

<sup>19</sup>Since [INV-] is also obligatory without the complementizer, this may be due to some other factor besides the complementizer.

<sup>20</sup>In some dialects of English, (i.a), in which *but* is arguably a marker, is grammatical.

- i. a. I don't know *but that* we will win that race.  
 S[MARKING *that*]  
 b. \*I don't know *but* we will win the race. S[MARKING null]

33. a. Swiss Bernese (Penner and Bader (3.6d))  
 i gibe dir s Gäüt bevor dass du s Buech  
 nimmsch  
 I give you the money before that you the book take
- b. Swiss Bernese (Penner and Bader (3.6e))  
 \*i gibe dir s Gäüt bevor du nimmsch s Buech  
 I give you the money before you take the book

In English, that licenses both [SLASH{}] and [SLASH{[1]}]. (Recall that [SLASH{[1]}] indicates that the clause has a gap.)

34. a. I think [that S[Kim likes pandas]]. S[SLASH{}]  
 b. What do you think [that S[Kim likes]]? S[SLASH{[1]}]

But McCloskey (1979) notes that Irish has complementizers that are sensitive to extraction: the complementizer *goN* occurs only with clauses that do not have a gap; the complementizer *aL* occurs only with clauses that have a gap, including questions, relatives, comparatives and clefts, and it must be in a position between the antecedent and the trace.

35. McCloskey (1979 Ch.2:45-46)
- a. Deir siad *goN* síleann an t-athair *goN* bpósfaidh Síle é.  
 say they that thinks the father that will marry Sheila him  
 'They say that the father thinks that Sheila will marry him.'
- b. An fear *aL* deir siad *aL* shíleann an t-athair *aL* phósfaidh Síle  
 the man that say they that thinks the father that will marry Sheila  
 'the man that they say the father thinks Sheila will marry'

McCloskey suggested that *aL* occurs only in the binding domain of gaps; Fodor (1992a) recasts Zaenen's (1983) analysis of this data in terms of subcategorization by *goN*

for a sister with no extraction feature and subcategorization by aL for SLASH. So [SPEC SLASH{[1]}], aL specifies that there must be a non-empty SLASH value on the head daughter, so that it is located between a filler and a gap.

In English, *that* does not license an interrogative expression, either as a filler or in situ.

36. a. \*Kim thinks that what Leslie bought? S[QUE{[1]}]  
 b. \*Kim thinks that Leslie bought what? S[QUE{[1]}]  
 (\* as informational question)

However, Polish *że* selects QUE{[1]}.

37. Polish (Lasnik and Saito 1984(11))  
 Maria myśli, [że co Janek kupił.]  
 Maria thinks that what Janek bought  
 What does Maria think that Janek bought?

HPSG's treatment of markers assumes that markers are not sensitive to any features not instantiated on the SYNSEM value of the sister node. For example, it appears that markers are not sensitive to properties of the complements of S, or its adjuncts, at least in English.

- 38.
- |              |    |             |
|--------------|----|-------------|
|              | S2 |             |
| /            |    | \ (head)    |
| marker       |    | S1          |
|              |    | [SYNSEM[1]] |
| (complement) | /  | \ (head)    |
|              | NP | VP          |

Markers are also not sensitive to whether the head daughter of Schema 4 has a referential subject or an expletive one, at least in English.

39. a. I think [that S[Kim likes pandas]].  
 b. I think [that S[it is obvious that Kim likes pandas]].

So although the lexical entry of a complementizer records the fact that it may select a sister with a gap, I know of no cases where a complementizer specifies which constituents may be extracted.

This is captured in HPSG because the HFP specifies that only head features of head daughters are structure-shared with the mother. In (38) for example, the head features of VP will be structure-shared with nodes S1 and S2. But S1 and S2 contain no syntactic information about the subject NP (though they contain semantic information via the effects of the Semantics Principle) because features of a non-head daughter normally are not structure-shared with the mother node.

Given this evidence, I follow P&S-94 in assuming that the value of SPEC is token-identical to the SYNSEM value of the head daughter, but not to the SYNSEM values of the daughters of the head daughter. If a learner acquires the value of SPEC, it will have acquired all relevant selectional properties of complementizers.<sup>21</sup>

### 3.4 LEXICAL ACQUISITION OF SELECTIONAL PROPERTIES

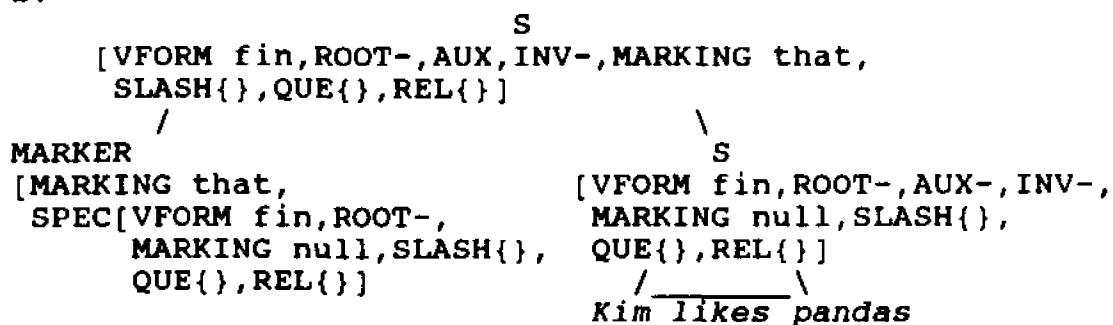
In this section, I give an example of how the learning process that Fodor proposed for GPSG works in HPSG, where

---

<sup>21</sup>Whether this is correct is an empirical question. Robert Levine (p.c.) says that in Kwakw'ala, complementizers carry agreement features for the subject of the embedded clause they introduce. If there is a discrepancy between the features that are available to SPEC and the features that complementizers are sensitive to, this question will need to be reconsidered.

the fact that the value of SPEC must be acquired ensures that the learner will acquire the language-particular further specification of Schema 4, in accordance with the Subset Principle. Suppose that the learner has encountered a sentence like (40.a), with the structure in (40.b), which is similar to the GPSG-type representation in (1.b), except that it includes the features MARKING and SPEC.

40. a. I think S[that Kim likes pandas].  
 b.



In a local tree, the combination of relevant universal constraints plus the relevant language-particular constraints plus the relevant lexical information (also a kind of language-particular constraint) unify to license a grammatical tree.

In HPSG, the HFP requires that the values of the head features of the mother node and the head daughter be token identical. Recall from Chapter 2 that the NONLOCAL Feature Principle of HPSG requires that the value of the NONLOCAL features on a phrasal sign be the union of the values of the nonlocal features of the daughters, except in the local tree where a NONLOCAL value is bound off. In addition, Schema 4 requires that the value of MARKING on the mother and the





because they are compatible with relevant properties of markers pre-specified in UG, are not in the learner's experience and therefore never become part of the learner's grammar.<sup>25</sup> Fodor's model for acquiring syntactic rules in GPSG is thus easily adapted to lexical learning in HPSG.<sup>26</sup>

### 3.5 ACQUIRING CORRECT GENERALIZATIONS

I now turn to the problem of how correct generalizations, that do not miss syntactic generalizations or license ungrammatical constructions, can be acquired in a default-free model. Omitting any feature has the potential to lead to over-generalization, as discussed in Chapter 1, so successful learning cannot be guaranteed unless the learner stores every appropriate feature on the SYNSEM node of the head daughter in the SPEC value of *that*. If even one feature in the local tree is not stored in the lexical entry for *that*, the lexical entry has the potential to license over-generalization. For example, if the lexical entry for *that* is identical to (42) except that the SPEC value does not include a value for VFORM, the SPEC value will license ungrammatical sentences such as (8), repeated here as (45).

45. \*I think that Kim to like pandas.

---

<sup>25</sup>Since they do not have to be acquired, universal constraints could be stated negatively.

<sup>26</sup>In Ch.5, I will explain in greater detail how the learning process works.

If the lexical entry for *that* were missing the value for INV, it would license ungrammatical sentences such as (46).

46. I think that *will* Kim like pandas.

To avoid a grammar which licenses ungrammatical constructions, the learner must store the value of every feature in the local tree. I call this the Store-all Principle.

47. Store-all Principle

The learner must store every value of every appropriate feature.

Note that the Store-all Principle precludes over-generalization by eliminating all generalization. If the learner obeys the Store-all Principle, its grammar will license the types of local trees that it encounter, and it will get the language right. In that sense, there is no learning problem. Unlike a hypothesis testing device, this model will not make errors. However, by the more stringent condition that the model should capture syntactic generalizations, this model fails. Therefore, this default-free model does not achieve correct generalization.

The problem of acquiring correct generalizations is still more complicated because the SPEC value in (42) is not the only such value for a complementizer that the learner will encounter. There will be other SPEC values for *that*; there will also be SPEC values for other complementizers, such as *for*. Recall that the Store-all Principle precludes over-generalization at the cost of eliminating all

generalization. The fact that a range of feature values are grammatical on the head daughter of Schema 4 means that a learner could encounter a set of local trees whose feature values differ by only a few features. For example, (42) is not the only SPEC value licensed by *that* in English. (48) requires a SPEC value like (42) but with [SLASH{[1]}].

48. a. What animal do you think *that* Kim likes \_\_\_?  
 b. [[MARKING *that*], [SPEC S[VFORM fin, ROOT-, AUX-, INV-, MARKING null, SLASH{[1]}, QUE{}, REL{}]]

In addition to [AUX-], *that* can also be specified for [AUX+].

49. a. I think *that* Kim might like pandas.  
 b. [[MARKING *that*] [SPEC S[VFORM fin, ROOT-, AUX+, INV-, MARKING null, SLASH{}, QUE{}, REL{}]]
50. a. What animal do you think *that* Kim might like \_\_\_?  
 b. [MARKING *that*], SPEC S[VFORM fin, ROOT-, AUX+, INV-, MARKING null, SLASH{[1]}, QUE{}, REL{}]]

The complementizer *that* can also be specified for [VFORM base], both with [SLASH{}] and [SLASH{[1]}].<sup>27</sup>

51. a. I demand *that* he throw away that paper.  
 b. [[MARKING *that*], [SPEC S[VFORM base, ROOT-, AUX-, INV- MARKING null, SLASH{}, QUE{}, REL{}]]
52. a. What do you demand *that* he throw away now?  
 b. [[MARKING *that*], [SPEC S[VFORM base, ROOT-, AUX-, INV-, MARKING null, SLASH{[1]}, QUE{}, REL{}]]

---

<sup>27</sup>Some speakers of English might never hear sentences like (51.a) or (52.a) and so they might not have the SPEC values for those sentences in their grammar. This is an example of how that this model accounts simply for dialectal differences, as noted in Section 3.1.

There are also at least two closely related SPEC values for, as sentences like (53.a) and (54.a) show.<sup>28</sup>

53. a. I would like for S[Kim to support Leslie].  
 b. [[MARKING for], [SPEC S[VFORM inf,ROOT-,AUX+,INV-, MARKING null,SLASH{}],QUE{}],REL{}]]
54. a. Who would you like for S[Kim to support \_\_\_]?  
 b. [[MARKING for], SPEC S[VFORM inf,ROOT-,AUX+,INV-, MARKING null,SLASH{[1]}],QUE{}],REL{}]]

The SPEC values in (42) and (48)-(54) are bulky and redundant. Each SPEC value licenses a type of local tree that is grammatical in English. If the learning mechanism requires that each SPEC value be stored individually, the grammar will under-generate: it will miss generalizations and it will contain duplicative information that is awkward to store.

However, most of the repeated values in (42) and (48)-(54) are the tentative default values proposed in (6) which Fodor suggested could be stripped away because they are predictable.

55. Tentative Default Values (*u* = unmarked) (=6)  
*u* VFORM = VFORM fin      *u* SLASH = SLASH{}  
*u* ROOT = ROOT+          *u* QUE(STION) = QUE{}  
*u* AUX(ILIARY) = AUX-      *u* REL(ATIVE) = REL{}  
*u* INV(ERSION) = INV-

This can be seen by comparing the lexical entries in (56)-(61), from which the tentative default values in (55) have been stripped off, with those in (42) and (48)-(52).

---

<sup>28</sup>I follow P&S-94 and P&S-87 in treating *to* as an auxiliary, as proposed by Pullum (1982).

56. (cf.(42))  
 [[MARKING *that*], [SPEC [ROOT-]]]
57. (cf.(48.b))  
 [[MARKING *that*], [SPEC [ROOT-,SLASH{[1]}]]]
58. (cf.(49.b))  
 [[MARKING *that*] [SPEC [ROOT-,AUX+]]]
59. (cf.(50.b))  
 [[MARKING *that*], SPEC [ROOT-,AUX+,SLASH{[1]}]]]
60. (cf.(51.b))  
 [[MARKING *that*], [SPEC [VFORM base,ROOT-]]]
61. (cf.(52.b))  
 [[MARKING *that*], [SPEC S[VFORM base,ROOT-,SLASH{[1]}]]]

All of the lexical entries in (56)-(61) have the value [ROOT-]; in Chapter 4, I argue that [ROOT-] is a context-sensitive default associated with complementizers and therefore it too does not need to be stored. Stripping off of defaults therefore produces linguistically perspicacious generalizations that correspond roughly to intuitions about relative complexity.

In addition, defaults have been assigned several other important roles in learning.

- Marked feature values guide the initial parsing process, called the "string to structure process" (Fodor 1989). If the learner doesn't know which value is appropriate for some feature, it selects or assumes the default value. If the default value doesn't work, the learner replaces it with a marked value. So defaults guide

selection of features, thus helping to make learning automatic.<sup>29</sup>

■ Markedness specifications guide the learner when it is stuck, either because the grammar is under-determined by the evidence or because the learner has made an error in constructing its grammar that it is having difficulty recovering from. Pinker (1989) proposes a notion of default-markedness, where the child adopts the unmarked form in the absence of evidence to adopt a marked form. This is similar to the proposal of Freidin and Quicoli (1989 BBS) who argue that defaults are necessary to set parameters under conditions of "zero-stimulation", i.e. in the absence of linguistic evidence. Lebeaux (1990) suggests that defaults are available to the learner "in times of computational difficulty", when the child is not able to determine the target grammar or when the child has made a mistake that it doesn't know how to fix.

■ Since markedness specifications determine the units by which languages may vary, they also determine the steps that learners may take during their progress from their immature initial grammar -- the default grammar -- to their mature

---

<sup>29</sup>From this point of view, there is an important distinction between systems in which all marked values are equally marked and systems in which the marked values are ranked by degree of markedness. A system with markedness ranking provides more guidance to the learner than a system in which all values are equally marked.

adult grammar. Williams (1981) equates markedness with the number of steps from the initial state to the final state.

The "unmarked case" can be understood as the child's initial hypothesis about language (in advance of any data), and a dimension of variation can be understood as an ordered set of hypotheses about a particular aspect of linguistic structure that the child successively submits to empirical test in learning a language.

Williams (1981:8)

(See also Berwick (1985), Fodor (1990).)

I therefore conclude that defaults play a crucial role in acquiring linguistically correct generalizations.

### 3.6 GENERALIZATION BY COLLAPSING

The duplication of values in (56)-(61) makes it clear that still a third kind of generalization, which I term 'generalization by collapsing', is needed. However, note that this type of generalization is required not on linguistic grounds but rather for compact storage. In this section, I discuss aspects of generalization by collapsing that pertain to conservative learning. Two standard ways of handling what I will call generalization by collapsing of feature structures are 'intersectification' and disjunction.<sup>30</sup> Intersectification is generalization by omission of features common to more than one feature structure. Intersectification resembles feature stripping in one important respect: it is also a kind of

---

<sup>30</sup>For a useful discussion of different ways of generalizing over typed feature structures, see Kasper (1993).

generalization by feature omission.<sup>31</sup> Disjunction is generalization by collapsing of closely related feature structures so that redundant information is listed only once. But, as noted in Chapter 1, the learner has no way of knowing when it has encountered all relevant occurrences of the complementizer and therefore can safely collapse feature structures. So in the absence of negative data, both operations carry the potential for licensing ungrammatical constructions, as will be shown below.

### 3.6.1 INTERSECTIFICATION

Carpenter (1993:12) defines the intersectification of two feature structures (which he calls 'generalization') as "the most specific feature structure which contains only information found in both feature structures", and he gives the example in (62).

62. Carpenter (1993(10))

$$\begin{array}{ccc}
 \text{a.} & & \text{b.} & & \text{c.} \\
 \boxed{\begin{array}{l} \text{F : a} \\ \text{G : b} \end{array}} & \cup & \boxed{\begin{array}{l} \text{G : b} \\ \text{H : c} \end{array}} & = & \boxed{\text{G : b}}
 \end{array}$$

This operation is well-defined, and the feature structure in (62.c) offers a perfectly acceptable abbreviation for the feature structures in (62.a) and (62.b) as long as (62.c) is only a descriptive generalization of [F:a, G:b] and [F:b,

---

<sup>31</sup>The term *intersectification* was suggested by Robert Levine. It captures the fact that intersectification is to intersection as unification is to set union.

H:c].<sup>32</sup> But recall from Chapter 1 that simplicity and generality go hand in hand: a feature description that mentions fewer features is more general than one that mentions more features. So (62.c), which is more general than (62.a) and (62.b), licenses both of them; it also licenses all of the feature structures in (63), and many others (an infinite number if the number of features is infinite).

63. a.

|   |   |   |
|---|---|---|
| G | : | b |
| F | : | b |

b.

|   |   |   |
|---|---|---|
| F | : | a |
| G | : | b |
| H | : | c |

c.

|   |   |   |
|---|---|---|
| G | : | b |
| H | : | x |

However, consider now the problems that intersectification can create for the learner. Imagine that the feature structures in (63) are ungrammatical in the target language. Even if the feature structures (62.a) and (62.b) occur in the target language, (62.c) is an overly strong generalization over them in that it licenses all of the feature structures in (63). This is exactly the type of learning that violates the Subset Principle: the learner has acquired a syntactic generalization that licenses a set of local trees that is a superset of those trees which are actually grammatical and so the learner has made an error that it may never recover from. Since the more powerful generalization

---

<sup>32</sup>Technically, (62.c) is a feature description and not a feature structure because it licenses more than one feature structure.

licenses all of the grammatical constructions and is simpler than the generalization that licenses only grammatical constructions, positive evidence does not require (or even hint) that the learner should narrow down its initial generalization.

### 3.6.2 CONSERVATIVE DISJUNCTION

Collapsing feature structures by disjunction in order to avoid proliferation of rules and redundant feature structures, also carries the potential for over-generalization.<sup>33</sup> With disjunction, the SPEC values in (64)-(68) can be combined into (69).

64. (=42)  
 a. I think that Kim likes pandas.  
 b. [[MARKING *that*], [SPEC S[VFORM *fin*, ROOT-, AUX-, INV-, MARKING null, SLASH {}, QUE {}, REL {}]]]
65. (=48)  
 a. What animal do you think *that* Kim likes \_\_\_?  
 b. [[MARKING *that*], [SPEC S[VFORM *fin*, ROOT-, AUX-, INV-, MARKING null, SLASH {{1}}, QUE {}, REL {}]]]
66. (=49)  
 a. I think that Kim might like pandas.  
 b. [[MARKING *that*], [SPEC S[VFORM *fin*, ROOT-, AUX+, INV-, MARKING null, SLASH {}, QUE {}, REL {}]]]
67. (=50)  
 a. What animal do you think *that* Kim might like \_\_\_?  
 b. [[MARKING *that*, [SPEC S[VFORM *fin*, ROOT-, AUX+, INV-, MARKING null, SLASH {{1}}, QUE {}, REL {}]]]
68. (=51)  
 a. I demand that he throw away that paper.  
 [[MARKING *that*], [SPEC S[VFORM *base*, ROOT-, AUX-, INV-, MARKING null, SLASH {}, QUE {}, REL {}]]]

---

<sup>33</sup>Carpenter (1992:194) says that without disjunction, the number of rules would increase exponentially.

b. What do you demand that he throw away now?  
 [[MARKING *that*], [SPEC S[VFORM base,ROOT-,  
 AUX-,INV-,MARKING null,SLASH[{1}],QUE{}],REL{}]]

69. [[MARKING *that*], [SPEC S[VFORM finVbase,ROOT-,  
 AUX+V-,INV-,MARKING null,SLASH[{{1}}V{}],QUE{}],  
 REL{}]] (V represents 'logical or')

(69) is simpler by the criterion of number of features, but it is also overly powerful; for example, it licenses base verbs with auxiliaries, even though this type of structure was not in the evidence set and in fact is not grammatical in English. So if learners collapse feature structures by disjunction, the collapsing must be constrained in a way that precludes this type of over-generalization.

Berwick (1985:234) says that are "two basic solutions to the problem of disjunction. One is to admit negative evidence...The only other possibility is to eliminate disjunctive statements." However, there is still a third possibility that preserves conservative learning in collapsing of lexical entries: the problem of over-generalization resulting from collapsing by disjunction can be solved by what I call 'conservative disjunction', where disjunction is allowed only if the learner has encountered evidence that every combination licensed by the disjunction is grammatical. If two feature structures have all of the same features but the values of two or more features in those structures differ, then the feature structures can be collapsed by nested disjunction. For example, the feature structures in (70) must be collapsed as in (71).

70. a.  $[\alpha F, \alpha G, \alpha H]$   
 b.  $[\alpha F, \beta G, \beta H]$

71.  $[\alpha F, ([\alpha G, \alpha H] \vee [\beta G, \beta H])]$

The disjunction is nested so that only features which co-occur are combined by disjunction. (71) licenses (70) but not (72), though (72) would have been licensed after standard disjunction.

72. a.  $[\alpha F, \beta G, \alpha H]$   
 b.  $[\alpha F, \alpha G, \beta H]$

Nested disjunction allows (64)-(68.b) to be collapsed into (73).

73. *that* SPEC S[(((fin, (AUX-VAUX+), INV-))  $\vee$  (BASE, INV-)),  
 ROOT-, MARKING null, SLASH{}  $\vee$  {1}, QUE{}, REL{}]<sup>34</sup>

This type of disjunction produces the desired results, but in a very unwieldy form -- the resulting feature descriptions will be very complex if there are multiple feature structures with complex disjunction.

Alternatively, we might suppose that disjunction is licensed only when two feature structures differ by exactly one feature value.<sup>35</sup> So (70.a) and (70.b) could not be combined in a single step. The learner could combine (74.a) and (74.b), if it encountered them, as (75).

---

<sup>34</sup>I mention INV with BASE because in some languages, verbs other than auxiliary may be inverted; I do not mention AUX because I assume that [+BASE] implies [-AUX]. Fodor (p.c.) notes that this is true only if *be* is not an AUX, as in *I require that you be sleeping by the time I return*.

<sup>35</sup>This type of disjunction embodies Pinker's one-difference constraint (1984 Ch.3, Procedures 1-6) which was introduced to solve this problem.

74. a. [ $\alpha$ F,  $\alpha$ G,  $\alpha$ H]  
 b. [ $\alpha$ F,  $\alpha$ G,  $\beta$ H]

75. [ $\alpha$ F,  $\alpha$ G,  $\alpha$ V $\beta$ H]

Then, if it also encountered (76), it could combine it with (75) as (77).

76. [ $\alpha$ F,  $\beta$ G,  $\alpha$ V $\beta$ H]

77. [ $\alpha$ F,  $\alpha$ V $\beta$ G,  $\alpha$ V $\beta$ H]

(64) and (65) can be collapsed into (78) because they differ by only one feature structure.

78. [[MARKING that], [SPEC S[VFORM fin,ROOT-,AUX-,INV-, MARKING null,SLASH{}V{[1]},QUE{},REL{}]]

Similarly, (66) and (67) could be combined into (79).

79. [[MARKING that], [SPEC S[VFORM fin,ROOT-,AUX+,INV-, MARKING null,SLASH{}V{[1]},QUE{},REL{}]]

Then single-value disjunction collapses (78) and (79) into (80).

80. [[MARKING that], [SPEC S[VFORM fin,ROOT-,INV+V-,AUX+, MARKING null,SLASH{}V{[1]},QUE{},REL{}]]

The SPEC values for in (53.b) and (54.a) can also be combined via disjunction, as in (81).

81. [[MARKING for], [SPEC S[VFORM inf,ROOT-,AUX+,INV-, MARKING null,SLASH{[1]}V{}},QUE{},REL{}]]

However, (80) cannot be collapsed with (81) because the two structures differ by two feature values, VFORM and INV.

So single-value disjunction produces the following generalizations for that:<sup>36</sup>

---

<sup>36</sup>Actually, if the learner first combined [VFORM fin V base], its generalizations would be somewhat different than these.

82. [[MARKING that], [SPEC S[VFORM fin,ROOT-,AUX+V-,INV-,  
MARKING null,SLASH{}V{[1]}],QUE{},REL{}]]
83. [[MARKING that], [SPEC S[VFORM base,ROOT-,AUX-,INV-,  
MARKING null,SLASH{}V{[1]}],QUE{},REL{}]]

These two cannot be combined any further. Notice that it would also have been possible to combine other pairs by single-value disjunction: for example, (64) could have been combined initially with (68.a) and (65) with (68.b). Further single-value disjunction would also collapse the six SPEC values into two; the two would be slightly different than (82) and (83), but they still would not license ungrammatical constructions.

Nested disjunction produces fewer, more complex generalizations and a consistent final result; single-value disjunction produces more, simpler generalizations but the resulting generalizations depend upon the order in which they were created. Since representing nested disjunction is very complex, I assume that single-value disjunction is preferable, and so I adopt the Principle of Conservative Disjunction in (84).

84. PRINCIPLE OF CONSERVATIVE DISJUNCTION  
The learner may combine two feature structures disjunctively only if they differ by exactly one feature value.

This principle will be needed in any learning model that has disjunction. Since they require extra learning and more complicated rules, disjunction and optionality are punished in this model.

### 3.7 CONCLUSION

The model of syntactic learning developed by Fodor for GPSG can be adapted to lexical learning in HPSG naturally and simply. The learning procedure proposed in this chapter will succeed for any syntactic phenomenon where there is a lexical feature like SPEC which can store 'sideways' selectional properties. In P&S-94 this includes SUBCAT (which is appropriate for heads) and MOD (which is appropriate for adjuncts). In Chapter 6, I will propose a similar feature for fillers, SELECT; in Chapter 7 I will discuss the implications of the fact that each schema is associated with a selectional feature.

I have also shown in this chapter that it is possible to conceive of a procedure for successful learning of these selectional properties from positive evidence, without defaults. However, the default-free model means that the learner will acquire a grammar that misses important generalizations. In particular, a grammar that is learned conservatively, from only positive evidence, will include many syntactic generalizations that list the same features over and over again. This can be seen in the values which are repeated in (81), (82) and (83).

85. (= (81))  
 [[MARKING for], [SPEC S[VFORM inf,ROOT-,AUX+,INV-,  
 MARKING null,SLASH{}V{{1}},QUE{},REL{}]]
86. (= (82))  
 [[MARKING that], [SPEC S[VFORM fin,ROOT-,AUX+V-,INV-,  
 MARKING null,SLASH{}V{{1}},QUE{},REL{}]]

87. (= (83))  
 [[MARKING that], [SPEC S[VFORM base, ROOT-, AUX-, INV-,  
 MARKING null, SLASH{}V{[1]}, QUE{}, REL{}]]

Stripping off the defaults in (55) from (81), (82) and (83) produces considerable economy of storage.

88. [[MARKING for], [SPEC S[VFORM inf, ROOT-, AUX+,  
 SLASH{}V{[1]}]]  
 89. [[MARKING that], [SPEC S[VFORM fin, ROOT-, AUX+V-,  
 SLASH{}V{[1]}]]  
 90. [[MARKING that], [SPEC S[VFORM base, ROOT-,  
 SLASH{}V{[1]}]]<sup>37</sup>

But in the default-free model, the Store-all Principle and the Principle of Conservative Disjunction, in conjunction with the Universal Feature Appropriateness Principle (Ch.2(6)) militate against any generalization at all: any feature licensed by UG in some context needs to have a specific value in order to prevent over-generalization by under-specification. So a grammar with the Feature Appropriateness Principle and the Store-all Principle is a grammar that misses generalizations, because every feature that is specified by UG as being appropriate for a context must have a value in that context.

However, the switch from GPSG, in which defaults were central, to HPSG, in which defaults have no theoretically sanctioned role means that a learning model with defaults is not consistent with P&S-94. In Chapter 4, I will discuss

---

<sup>37</sup>Note that with single-value disjunction there can be no more collapsing because each pair of lexical entries differs by two or more values.

the reasons that defaults were a central component of GPSG but were eliminated from HPSG. In Chapter 5, I will present a learning model in which defaults are restricted to the lexicon and constrained in a linguistically justified manner that avoids the computational problems they caused in GPSG.

**CHAPTER 4  
DEFAULTS AND LINGUISTIC THEORY**

Markedness theory arose to explain the existence, across languages, of phenomena which are pervasive but not universal; these asymmetries would be unexpected if the distribution of features across languages were not biased. The need to account for recurring syntactic patterns is also the reason for the continued reliance by linguists on the notion of markedness, in spite of a mixed attitude toward its place in linguistic theory. In recent years, this variable attitude has been apparent both in GB and in phrase structure grammar. In 1981, Chomsky made a clear statement about markedness and UG:<sup>1</sup>

...we assume that the child approaches the task equipped with UG and an associated theory of markedness that serves two functions: it imposes a preference structure on the parameters of UG, and it permits the extension of core grammar to a marked periphery.

Chomsky (81:8-9)

However, by the 1990's, Chomsky defends the 'minimalist' view that principles are universal, and that apparent

---

<sup>1</sup>Chomsky (1986) characterized three types of markedness: "The distinction between core and periphery leaves us with three distinct notions of markedness: core versus periphery, internal to the core, and internal to the periphery. The second has to do with the way parameters are set in the absence of evidence. As for the third, there are, no doubt, significant regularities even in departure from core principles (for example in irregular verb morphology in English) and it may be that peripheral constructions are related to the core in systematic ways, say, by relaxing certain conditions on core grammar."

(Chomsky 1985:147)

language-particular rules can be reduced to choices of parameter settings.

In early work, economy considerations entered as part of the evaluation metric, which, it was assumed, selected a particular instantiation of the permitted format for rule systems, given PLD (primary linguistic data). As inquiry has progressed, the presumed role of an evaluation metric has declined, and within the principles-and-parameters approach, it is generally assumed to be completely dispensable: the principles are sufficiently restrictive so that PLD suffice in the normal case to set the parameter values that determine a language...With a proper formulation of such principles, it may be possible to move toward the minimalist design: a theory of language that takes a linguistic expression to be nothing other than a formal object that satisfies the interface conditions in the optimal way. A still further step would be to show that the basic principles of language are formulated in terms of notions drawn from the domain of (virtual) conceptual necessity.

Chomsky (1993:4-5)

Thus Chomsky takes it as the goal of 'minimalist' theory to eliminate the need for the evaluation metric; however, he reserves the possibility that markedness (presumably in a subset relation) will still be needed for ordering parameter settings.

Markedness of parameters, if real, could be seen as a last residue of the evaluation metric."

Chomsky (1993,fn.6)

Moreover, Fodor (1989b) argues that even if general principles can be found to replace all parameters proposed for the core, the problem of explaining how children can learn the body of exceptional information contained in the periphery will remain.

Recent work in the principles and parameters approach continues to rely on markedness. In general, the two kinds

of economy discussed in Chomsky (1993:15), "shortest move" and "fewest steps in a derivation" can be viewed as modern incarnations of traditional markedness principles. GB theorists also continue to rely on the notion of defaults in their descriptive work. For example, Rizzi (1994:268) says that the unmarked expletive is [NULL+]: "If the unmarked case for an expletive is to be null, then it is natural that the child will take the null constant option for the expletive." Rothstein (1995:508) suggests that "pronouns that have a pleonastic use are the most minimal of the pronouns, the realization of the least marked, most neutral features". Non-pleonastic pronouns are marked relative to pleonastic ones. Farkas and Zec (1995) argue that evidence from Roumanian supports the desirability of agreement systems relying on defaults.

A shift in the attitude toward defaults also took place in phrase structure grammar: syntactic defaults were a central component of GPSG, but HPSG has no syntactic defaults. In this chapter, I discuss the reasons for the mixed attitude toward defaults in phrase structure grammar. GKPS introduce Feature Specification Defaults (FSDs) by citing the precedent of the markedness conventions in the feature system of SPE (Chomsky and Halle 1968).<sup>2</sup> Although syntactic markedness has been studied by Greenberg (1966a

---

<sup>2</sup>Defaults were originally developed for phonology by TrubeTrubetzko (1939).

and 1966b), Hawkins (1983) and others, less work has been done on syntactic default than on phonological ones. It was an innovation of GPSG to assign to defaults a central role in a formal syntactic theory.<sup>3</sup> As in phonology, defaults provide a principled way to account for the presence of feature values that represent the typical or neutral case; in the absence of defaults, each typical value that is not universal must be instantiated by stipulation in a rule or special principle. Defaults build into the syntax the notion that typical syntactic patterns exist, but there can be exceptions to these patterns.

I begin with a discussion of the linguistic basis for defaults: defaults are feature values that represent neutral states, either in that a particular kind of complexity is absent or predictable, or in that they capture typical feature associations that are somehow more expected than others. Although the reasons for the asymmetry may not always be apparent, it does appear to be a basic fact about natural language that not all feature combinations are of equal linguistic status. I discuss criteria for recognizing default values and then turn to a consideration of the default values for the feature ROOT. I first propose a context-free default value for ROOT; I then consider the association of complementizers with embedded clauses, and

---

<sup>3</sup>See also the Hierarchy of Noun Phrase Accessibility of Keenan and Comrie (1977).

show that if defaults are to do any real linguistic work, they must be context-sensitive. I then list other plausible defaults, and conclude that in general, linguistic evidence supports the hypothesis that markedness plays an important role in accounting for recurring linguistic patterns that are not universal.

In Sections 4.3 and 4.4, I examine some computational problems with defaults. I consider problems with context-sensitive defaults of the sort used in GPSG, (which were discussed in Maxfield (1990)), including conflicting defaults, ambiguity, and circularity. I also review the reasons that defaults were discarded in the shift from GPSG to HPSG. Because of the architecture for defaults adopted in GPSG, it was necessary to check each tree admitted by a phrase structure rule in order to determine whether default values were instantiated; in constructions such as raising and agreement, where the value in one tree was dependent on the value in another tree, this comparison was trans-derivational, and potentially infinite.

In spite of these computational problems, some sympathy for defaults persists in HPSG analyses. Although HPSG has 'officially' eliminated defaults (P&S-94 Ch.1, fn.22), defaults are a central mechanism in the lexical model presupposed by P&S-87 and P&S-94. Moreover, HPSG linguists commonly rely on lexical defaults in practice to account for

neutral or predictable feature values that follow from no other principles.

The general picture that emerges is that evidence from syntactic theory and learnability theory supports the hypothesis that defaults are essential; however, a learning model that relies on defaults that are indeterminate or require infinite search is computationally unacceptable. I will argue that the potential explanatory value of defaults in syntax and learnability theory motivates a reconsideration of whether defaults can be defined and regimented in a way that avoids the potential computational problems.

In Chapter 5, I propose a solution to the problem of syntactic learning without negative evidence that employs only lexical defaults. Linguistically motivated architecture under-writes the distinction between syntactic and lexical defaults. And the elimination of syntactic defaults is shown to be successful in precluding indeterminate defaults, defaults with insufficiently specific context, and infinite search problems.

#### 4.1 THE LINGUISTIC BASIS FOR SYNTACTIC DEFAULTS

The use of defaults in the learning model developed for GPSG by Fodor (discussed in Chapter 3) was uncontroversial because defaults were a central mechanism in GPSG.<sup>4</sup>

Feature Specification Defaults (FSDs) form an important part of the link between the highly schematic rules

---

<sup>4</sup>However, GPSG theory did not obey Fodor's Specific Default Principle; defaults were sporadic.

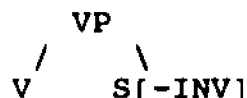
listed in the grammar and the fully specified structural description they induce.

As the first illustration, consider rule (18).

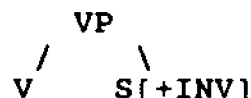
(18) VP --> V S

Consider the rule in relation to the feature INV (mnemonic for INVerted sentence). Two of the tree fragments that we might expect to get from (18) are displayed in (19).

(19) a.



b.



That is, there is no particular reason for having [-INV] rather than [+INV] on the S daughter. But if we allow the grammar to admit fragments like (19b), then we will end up generating examples like \*Lee believes will the children be late. So INV needs to have a default specification, [-INV]. ...

Since there is no reason for INV not to have that specification, it must have it, according to the approach to defaults that we adopt in this book. There is, of course, one class of structures in which [+INV] is obliged to be present, namely those arising from rules such as (21).

(21) S[+INV] --> V[AUX] NP VP[BASE]

But, since this rule stipulates the presence of [+INV], the default will not be invoked.

GKPS (1985:30-31)

This brings out a basic idea underlying the use of defaults (e.g., Fodor 1992b): defaults can be over-ridden, but only at a cost of added complexity. Sells (1985:104) says that defaults are "something that all (linguistic) theories use in some way or other", and that a default is by definition able to be over-ridden, but only if the presence of the non-default value is "specifically sanctioned" (Sells 1985:104). Calder (1994) "characterize(s) a default statement as one which holds in the absence of information to the contrary."

Carpenter (1993) says that defaults "provide a method for allowing information to be deduced about an object if it is consistent with what is already known about that object."

Because I am concerned with the role default values play in the grammar, I propose the definition of defaults in (1):

1. **Default values**  
Default values are values that are licensed whenever no rule or principle imposes a non-default value in some context.

The FSDs of GKPS include those in (2).

2. FSD 1: [-INV]           The default value for INV is [-].  
FSD 2:  ~[CONJ]        The default is for CONJ not to have a specification.  
FSD 3:  ~[NULL]        The default is the absence of null constituents.  
FSD 6:  [+ADV] > [BAR 0]  
FSD 9:  [INF, +SUBJ] > [COMP for]  
FSD 10: [+N, -V, BAR 2] \* ACC  
                          The default case for NPs is accusative.  
FSD 11: [+V, BAR 0] > [AGR NP[NP NFORM NORM]]  
                          The NP subject of a verb is referential by default.

The FSDs of GPSG, modified for learnability reasons, defined the default values that were stripped off in Maxfield's algorithm for acquiring compactly stored syntactic rules, as discussed in Chapter 3. The default values were eliminated because they were predictable, so that only marked, non-predictable values were left to be stored in acquired rules.

GKPS provide no systematic criteria for identifying linguistically perspicacious defaults. Arbitrarily selected defaults are not linguistically interesting because they provide no explanation for the asymmetrical distribution of

syntactic phenomena. In the next section, I propose some criteria.

#### 4.1.1 CRITERIA FOR RECOGNIZING DEFAULT VALUES

Since default values are by definition not manifested in every node of every tree, determining which value of some feature or attribute is unmarked is not always easy. An unfortunate side-effect of markedness theory is that the claim that there exist defaults forces a loosening of the nature of syntactic evidence. In the simplest form, the hypothesis that all universals are innate is constrained by the commitment to the empirical claim that all universals are exhibited in all natural languages. Once defaults are admitted, the empirical evidence for universals is weakened because a linguistic property can be innate, but not universal.

However, linguists' judgements tend to agree about cases where there is a clear asymmetry of values. The clearest instances of asymmetrical markedness occur where one state or value is neutral in that it represents the absence of some property that adds complexity. For example, GKPS suggest that the default for INV(ersion) is [-]; at least in English, there is an intuitive sense that inversion adds complexity relative to the absence of inversion which is the neutral case. Similarly, the default value for CONJ(unction) is [-], and not [+], because a conjoined phrase is more complex than a non-conjunctive one.

Linguists tend, perhaps without realizing it, to use [-] as the value that their implicit knowledge tells them is the default value.<sup>5</sup> The consensus on minus values attests to agreement the substantive issue of what is marked and unmarked. This intuition is supported by facts such as that inverted clauses are more restricted in their distribution than non-inverted clauses (e.g., in most dialects of English, subject-auxiliary inversion is limited to root clauses.)

Most of the tentative defaults that I assume in Ch.3(6) are of the sort that is characterized by absence of complexity.

3. Tentative Default Values (*u* = unmarked)
- |                     |                         |
|---------------------|-------------------------|
| [u ROOT = +]        | [u SLASH = {}]          |
| [u AUX(ILIARY) = -] | [u QUE(STION) = {}]     |
| [u INV(ERSION) = -] | [u REL(ATIVE) = REL {}] |

The default value for AUX is [-] because a clause without an auxiliary is simpler than a clause that has an auxiliary. A clause without a SLASH, an interrogative expression or a relative expression is simpler than one with each of these respectively. On this basis, I assume the context-free defaults in (3) throughout this dissertation. Still, it is

---

<sup>5</sup>The fact that unmarked values are typically characterized by the absence of complexity has resulted in confusion of the unmarked value with the negative value of Boolean features. For example, Greenberg (1966b:25-26) suggested that the negative value is always unmarked. But since which value is negative and which value is positive depends upon the name of the feature and therefore is arbitrary, unmarked values need not be negative. This can be seen clearly from the fact that [ROOT+] is equivalent to [EMBEDDED-] and [ROOT-] to [EMBEDDED+].

important to note that further research is necessary to determine exactly what counts as linguistic complexity.

For the NONLOCAL features, I suggested above that the empty set is the context-free default; this appears to be correct because a clause without a gap is less complex than a clause with one; similarly a clause with an interrogative expression is complex relative to clause without one.

For phenomena not characterized by the clear asymmetry of the absence versus the presence of some particular phenomenon, which state is unmarked is more difficult to determine. Since the values of these attributes are not characterized by an unambiguous asymmetry, no single value can be characterized as clearly less complex than any other.<sup>6</sup> The only example of an attribute like this which I propose in (Ch.3,(6)) is (4):<sup>7</sup>

4. [u VFORM = fin]

The reason for selecting (4) is that *fin* has by far the widest distribution relative to VFORM's other values, which

---

<sup>6</sup>Assuming that these symmetrical features have a default, the question arises as to whether there is a hierarchy of markedness, or whether all values other than the default are equally ranked. For CASE, there are two possibilities, where (ii) is intended only as an illustrative example and not a hypothesis about the marked values of CASE.

i. [m CASE = -CASE acc]  
 ii. [m CASE = [CASE nom < genitive < oblique] (i.e. the marked values of CASE, in order of increasing markedness, are nominative, genitive, oblique.

For binary features, of course, whichever value is not the default must be marked.

<sup>7</sup>An attribute is a feature with symmetrical values.

include at least *infinitive*, *base*, and *perfective*. The attribute CASE has at least three symmetrical values, *nominative*, *accusative*, and *oblique*. Keenan and Comrie's hierarchy of extraction (Keenan and Comrie 1977) seems to suggest that the default for CASE should be *nominative* because subjects are more accessible to extraction than other arguments. But if P&S-94 are right, this is attributable to the fact that *nominative* interrogatives are not associated with gaps.<sup>8</sup> I follow GKPS in assuming that the default for CASE is *accusative* because of examples like (5) (among others).

5. a. It is me[acc].  
b. Jimmy and me[acc] went home.

In spite of the difficulty in establishing formal criteria for recognizing default values when the default value is clearly characterizable as the least complex value, some have been proposed.<sup>9</sup>

- i) Neutralization (Trubetzkoy 1939): The intuition behind the neutralization criterion is that an unmarked class tends to be divided into more subclasses than a marked class does. If the unmarked value for NUMBER is *singular* and the marked

---

<sup>8</sup>The cross-linguistic variation among constituents may be extracted in different languages as reflected in Keenan and Comrie's hierarchy of extraction of relativizers is another example of the kind of differences between languages that would be likely to be lost in the absence of conservative learning.

<sup>9</sup>See also Moravcsik and Wirth (1986), Tomić (1989), and Klein (1990) for discussion of criteria for establishing relative markedness.

value is *plural*, it is expected that there will be more forms for singular than for plural. In English, gender distinctions are made among singular pronouns (*he, she, it*) but not among plural pronouns (*they*). In classical Latin the dative and the ablative typically have different case markings in the singular but the same markings in the plural.

The concept of neutralization can be translated into feature co-occurrence. If there are two binary features, F and G, and if [F-,G-], [F+,G-], and [F-,G+] are all observed, but [F+,G+] does not occur, the absence of [F+,G+] is evidence for the markedness of [F+] relative to [F-] and [G+] relative to [G-]. In other words, either there is no [F+V-] distinction in the presence of [G+], or, there is no [G+V-] distinction in the presence of [F+].

ii) Degree of variation in context (Koster 1978): Marked phenomena tend to be subject to lexical idiosyncrasy and variation in grammaticality judgement while unmarked phenomena do not. According to this criterion, it is unmarked for movement to be clause-bounded, and marked for movement to cross a clause boundary.

iii) Syntactic judgements about marked constructions tend to vary, while syntactic judgements about unmarked constructions are more uniform (Koster 1978): Koster's example of a construction that is marked according to this criterion is topicalization; syntactic judgements about the

grammaticality of topicalization in embedded clauses are both more sensitive to context and more likely to be subject to individual judgement than syntactic judgements about topicalization in root clauses.

6. Koster 1978(44)
- a. Each part he has examined carefully.
  - b. He explained that each part he had examined carefully.
  - c. \*I resent that each part he examined carefully.
  - d. \*Bill's claim that each part he examined carefully is clearly false.
  - e. \*After each part he had examined carefully, he left.

iv) Conditional distribution (Greenberg 1966a:73): "Given  $x$  in a particular language, we always find  $y$ ." Crucially, the converse does not apply: the presence of  $y$  does not imply the presence of  $x$ . This asymmetry implies that  $y$  is unmarked relative to  $x$ . That is, some languages, or contexts, may permit only the unmarked form. For example, overtly realized subjects are unmarked relative to non-overt subjects; conditional distribution implies that languages with null subjects will also permit overt subjects, but the fact that a language has overtly realized subjects does not imply that it will license null subjects.

The criteria for markedness are not mutually exclusive, as Moravcsik and Wirth (1986) point out.

The central claim is that the various tests that demonstrate the asymmetry between the two members of the opposition will have converging results. Once one of the two members has been shown to be marked by one criterion - let us say, it has been shown to be structurally more complex than the other, or paradigmatically poorer, or more restricted in its distribution - all other relevant tests will also

converge to select that entity as the marked member of the opposition.

Moravcsik and Wirth  
(1986:3)

Since I showed in the previous chapter that specific values will do as well as specific defaults from the point of view of getting the language right from positive evidence (Section 3.5), the success of my learning theory does not depend on whether or not every feature has a default value.<sup>10</sup> That is, the Specific Defaults Principle is not essential. But in the rest of this dissertation, I will assume, for the sake of uniformity, that all features have defaults in all contexts. Note that this is stronger than the Specific Defaults Principle, which simply assumes that all features whose values need to be learned have defaults.

In order to make a concrete proposal, I will be forced to rely more heavily than is desirable on conjecture and reasonable assumption in my choice of default values. I will assume that the neutral or less complex value is the default whenever possible; in addition, I will use patterns of feature association and the formal criteria for markedness discussed above as evidence that some phenomenon is marked.

---

<sup>10</sup>However, the economy of generalization that defaults permit would be lost if most features did not have default values.

## 4.1.2 SYNTACTIC MARKEDNESS: A CASE STUDY

## 4.1.2.1 CONTEXT-FREE DEFAULT FOR ROOT

In this section, I present evidence that [-] is the context-free default for ROOT, but that [+] is the default value for ROOT in the context of a complementizer. The feature ROOT is especially interesting because the distinction between root and embedded clauses is unequivocally syntactic: the value of ROOT is determined by sentence structure rather than by meaning, morphology or phonology.<sup>11</sup> The clause *Kim loves pandas* is [ROOT+] in (7.a) and [ROOT-] in (7.b).

- |    |    |                                         |          |
|----|----|-----------------------------------------|----------|
| 7. | a. | S[Kim loves pandas].                    | S[ROOT+] |
|    | b. | Ruth thinks S[(that) Kim loves pandas]. | S[ROOT-] |

I assume that the highest finite verb in a sentence is [ROOT+] and that all other verbs are [ROOT-].

I list in (7.b)-(7.b) some examples that illustrate phenomena with different distribution in root and embedded clauses, which motivate this featural distinction in UG.

- Standard English: Subject-auxiliary inversion in interrogatives is generally grammatical only in matrix clauses.
- German: In matrix clauses, the finite verb is generally second; in embedded clauses, the finite verb is last.
- Romance (Kayne 1991:647): There are "extra possibilities" for order of clitics relative to the verbs in root sentences such as imperatives (see also Emonds 1976:

---

<sup>11</sup>ROOT is not a feature in GPSG or HPSG, but it was proposed in Maxfield (1990).

202, 236), and in root sentences of other varieties in Galician and Portuguese.

○ Romance: Subject-clitic inversion occurs only in matrix contexts in Fiorentino and French (Rivero 1994:111 cites Kayne 1983 and 1989 for French and Brandi and Kordin (1989) for Fiorentino, a Northern Italian dialect). In embedded clauses, the subject clitic must precede the auxiliary.

○ English: Topicalized constructions are less likely to be grammatical in embedded clauses than in root clauses.

(See (6)).<sup>12</sup>

○ English: Tag questions only attach to [ROOT+] (Emonds 1976 II, 15, 17).<sup>13</sup>

<sup>12</sup>Robert Levine points out that some verbs, such as *resent*, can occur with a topicalized clause:

i. I strongly resent that MY work he feels that he can appropriate whenever he feels like it, though he's always careful to cite Leslie's papers a dozen times in each article.

The issue here is the role of extra-syntactic conditions in determining the grammaticality of root constructions in embedded clauses. If special factors, such as particular verbs like *resent*, or pragmatics or semantics, are needed to license 'root' phenomena in embedded clauses, then the distinction in syntactic markedness between root and embedded clauses is supported. In this particular example, (i) is better than (ii) and (ii) is better than (iii).

ii. I strongly resent that MY work he feels that he can appropriate whenever he feels like it.

iii. I strongly resent that my work he feels that he can appropriate.

This contrast suggests that the topicalization in (i) is licensed by pragmatics and by the explicit contrast, highlighted by the emphasized 'MY', between the speaker and Leslie. (For arguments against Emonds' and Koster's claims, see Green (1976)).

<sup>13</sup>Robert Levine points out the existence of examples in which tag questions are grammatical in embedded clauses:

i. Leslie's attitude was that Robin was a spy, wasn't he? and we execute spies, period.

8. a. Mary had come, hadn't she.  
 b. \*Bill wanted to know whether Mary had come, hadn't she.  
 c. \*The idea that Bill knew whether Mary had come, hadn't she, is preposterous.

It is an interesting fact that natural languages widely (and perhaps universally) exhibit distinctions between matrix clauses and embedded ones, especially since the distinction complicates syntax. In a feature based system like HPSG, this complication is manifested in that [ROOT+V-] is needed to constrain many syntactic generalizations. Syntactic generalizations that license constructions like topicalization, right and left dislocation and subject-auxiliary inversion will not be properly constrained without the inclusion of the appropriate value of ROOT in the feature structure.<sup>14</sup> The feature ROOT is thus an empirically necessary complication of the grammar; without ROOT (or some other representation of the [ROOT+V-] distinction), the grammar of a language like English would be simpler but would license ungrammatical sentences.

---

I assume that the complex pragmatics of this sentence, in combination with the semantics of 'attitude', license the embedded topicalization.

<sup>14</sup>GB theory assumes an analysis of the distinction between root and embedded clauses that depends on the presence of an invisible complementizer in ROOT clauses (e.g., Rizzi 1990:42). Thus German V2 and English subject-auxiliary inversion is analyzed as resulting from movement of the finite verb to C<sup>0</sup> or to some other fronted position in the ROOT clauses. In German, C<sup>0</sup> is "intrinsically endowed with morphosyntactic features" that attract I<sup>0</sup>; in English C<sup>0</sup> is marked as [+wh] so that it attracts I<sup>0</sup> only in interrogatives. This solution, which depends on invisible elements with different properties, can create significant problems for a learning model.

A basic distributional criterion suggests that root clauses are unmarked relative to marked clauses: typically, a sentence must have a root clause. Greenberg's criterion of conditional distribution also suggests that root clauses are unmarked relative to marked clauses. The presence of a root clause does not imply the presence of an embedded clause, but the presence of an embedded clause usually implies the presence of a root clause.

The asymmetry in the distribution of phenomena permitted in root and embedded clauses was noticed first by Ross (1973), and discussed by Emonds (1976) and Koster (1978), among others. Ross (1973:397) proposed the Penthouse Principle in light of his observation that "more goes on upstairs than downstairs".

9. The Penthouse Principle  
No syntactic process can apply only in subordinate clauses.

The Penthouse Principle claims explicitly that all constructions permitted in embedded clauses are grammatical in root clauses, though not vice versa, and implies that it is universally safe to generalize from [ROOT-] contexts to [ROOT+] contexts: if a construction is licensed in a subordinate clause, it must also be markedness. If the approach to markedness adopted in this thesis shows promise, cross-language research to formalize the notion of syntactic markedness and ascertain examples will be required.

Unfortunately, the Penthouse Principle is too strong.<sup>15</sup> Finite-verb-final word order in German is not grammatical in root clauses: German has V-final word order in embedded clauses, but V-second is obligatory in embedded clauses. English has standard word order in embedded questions but auxiliary inversion is obligatory in root questions.<sup>16</sup>

Emonds (1976) argued that the differences between root and embedded clauses follow from the fact that transformations in embedded clauses are always structure-preserving whereas both structure-preserving and non-structure-preserving transformations are permitted in root clauses. Structure-preserving transformations involve movement of a constituent to a node that is licensed by X-bar theory, and therefore pre-exists at deep structure.

---

<sup>15</sup>The distributional asymmetry reflected in the Penthouse Principle is also the reason that it has been suggested that sometimes syntactic learning is safer in embedded clauses than in root clauses. In the model of Roeper and Weissenborn (1990), the evidence that triggers setting of the so-called Pro-drop or Null Subject Parameter must come from the embedded clause, not the root. The reason is that there is more syntactic variation in the root clause than the embedded clause. For example, even in an overt subject language like English, subjects are sometimes dropped in root clauses, but not in embedded clauses. Thus Roeper and Weissenborn claim (p.155) that "parametric decisions have no local exceptions in subordinate clauses". But the claim that children can only learn to set core parameters based on embedded sentences is puzzling from a developmental perspective: it would be surprising if children had to learn basic structures from embedded clauses, given that their memory and probably their processing abilities are limited.

<sup>16</sup>Andersson and Dahl (1973) also discuss a case in Swedish that runs counter to the Penthouse Principle. The auxiliary *ha* is optionally deletable in embedded clauses, but is not deletable in root clauses.

Structure-preserving transformations, such as passivization and dative shift, are typically found in root and embedded clauses, distributed roughly evenly. Root transformations, on the other hand, are not structure-preserving, and so are licensed only in root clauses. Examples of root transformations include topicalization, right and left dislocation and subject-auxiliary inversion.<sup>17</sup>

However, the distinction between root and structure-preserving transformations made by Emonds is also too strong: for example, topicalization is sometimes grammatical in embedded clauses, as noted in (6). Koster (1978) re-interpreted the distinction between root and matrix clauses in terms of core and peripheral phenomena. He argued that topicalized constituents in embedded clauses were marked relative to topicalized constituents in root clauses; he argued that the marked cases were lexically driven and subject to individual variation. As noted above (p.116),

---

<sup>17</sup>The existence of a special class of root transformations which involve creation of a node by attachment accounts for the asymmetrical distribution of these phenomena in root and embedded clauses. The asymmetry between Emonds' structure-preserving and non-structure-preserving transformations means that the learner must know which constructions ROOT is relevant to. In generalizations that produce structure-preserving constructions like passivization and double objects, the feature ROOT need not be mentioned. Mentioning ROOT in the generalizations about these constructions will produce bulky and redundant grammars -- it doesn't make sense for a grammar to have one syntactic generalization about passivization in [ROOT+] and another one in [ROOT-]. Thus grammars -- and learners -- must know when it is, and is not, appropriate to include the feature [ROOT+/-] in the description of some phenomenon. Fodor (p.c.) suggests that because passivization and double object constructions are lexical, they may never have [ROOT+] or [ROOT-] distinctions.

Koster took the contrast between the general grammaticality of 'root phenomena' in root clauses and their limited distribution in embedded clauses as evidence that embedded clauses are peripheral and therefore marked. If more things happen in some context, the context is unmarked; if fewer things happen in some context, the context is marked. This is an example of neutralization.

However, Koster's claim that all phenomena that do not conform to the Penthouse Principle are lexically driven is also too strong. Although subject-auxiliary inversion is not grammatical in embedded clauses in standard English, certain dialects of English such as Belfast Irish license subject-auxiliary inversion in embedded clauses as well as root clauses.

10. Belfast English (Henry 1992(42))  
I wonder what street *does* he live on.

But the weakening of the strict distinction hypothesized by Ross and Emonds does not nullify the observation that, in general, root clauses license more syntactic variation than embedded ones do. The central claim is that if the distribution of some phenomenon that adds complexity is limited in some language, it is likely to be limited to root clauses rather than to embedded clauses.

It is important to note that the claim that there exist syntactic defaults does not preclude the possibility of deeper explanations for these defaults; in fact the existence of defaults cries out for explanation, which

regrettably cannot always be given. In the case of the default [ROOT+], the deeper explanation might be connected to the fact that a sentence conveys a single concept or predicate, and that embedded clauses and conjunction are UG-provided mechanisms to embellish or enhance the single central concept. However, the truth of this speculation will not be central to the descriptive purposes for which the [ROOT+] default is used in what follows.

#### 4.1.2.2 CONTEXT-FREE DEFAULT FOR MARKING

I assume that context-free default value for MARKING is *null* because *null* is neutral in that it involves no language-particular learning.<sup>18</sup> In HPSG, [MARKING null] is needed in languages like English to prevent the grammar from licensing a potentially infinite series of markers at the beginning of clauses, as in (11).

11.            Schema 4  
           [MARKING [2]]  
           /                    \  
       [MARKING [2]]            XP  
                                   [MARKING null]

---

<sup>18</sup>P&S-94 (Ch.1,fn.51) note that mother nodes of all schemas other than Schema 4 must be specified as [MARKING null]. This is necessary to prevent all schemas except Schema 4 from licensing markers on their head daughters (recall that MARKING is a head feature). Second, verbs need access to the value of MARKING for purposes of subcategorization (as discussed on p.72ff.). Since not all verbs are subcategorized for S[MARKING null], those verbs which are must be labelled accordingly. *think* is subcategorized for S[MARKING *that*] and S[MARKING null], while *quip* is subcategorized only for S[MARKING *that*].

- i.    a.    Leslie thinks S[*that* Kim likes pandas].  
       b.    Leslie thinks S[Kim likes pandas].  
 ii.  a.    Leslie quipped S[*that* Robert likes pandas].  
       b.    \*Amy quipped S[Robert likes pandas.]

(11) correctly does not license sentences such as (12).

12. (=Ch. 3(32.a))  
 \*I think [that that Kim likes pandas].

However, as noted in Ch. 3, the requirement that there only be a single marker per clause is not universal, at least on the assumption that adverbial adjuncts like *but* in English or *bevor* in Swiss Bernese are markers.

13. (=Ch.(33))  
 a. Swiss Bernese (Penner and Bader (3.6d))  
 i gibe dir s Gäüt bevor dass du s Buech nimmsch  
 I give you the money before that you the book take  
 b. Swiss Bernese (Penner and Bader (3.6e))  
 \*i gibe dir s Gäüt bevor du nimmsch s Buech  
 I give you the money before you take the book

(13) shows that the head daughter of Schema 4 is not always [MARKING null]. In the typical case, there is only one complementizer per clause, but in particular languages (or contexts) the default can be over-ridden, as in (13.a). Since the non-head daughter of Schema 4 is obligatorily [MARKING lex], the default value [MARKING null] on the non-head daughter serves anti-recursion mechanism without absolutely precluding sequences of markers.

#### 4.1.2.3 CONTEXT-SENSITIVE DEFAULT FOR ROOT AND MARKING

The FSDs in (2) are of two types, context-sensitive defaults and context-free, as in the markedness conventions of SPE (Chomsky and Halle 1968:405-407) which GKPS cite as a precedent for FSDs. Context-sensitive default statements typically take the form in (14), where some feature (or features) establishes a context that changes the markedness

value of the other feature, and context-free defaults take the form in (15).

14. Context-sensitive markedness specifications

[uF = +] / [G+]  
[uF = -] / [G+]

15. Context-free defaults

[uF = -]  
[mF = +]

Fodor (1992b:34) notes that "context sensitive defaults are particularly useful, because they can extract subtle universal trends in feature distribution than does a single across-the-board default", and she suggests that defaults are "governed by an Elsewhere convention giving priority to more specific principles over less specific ones". Thus, context-free defaults apply whenever there is no relevant context-sensitive default, just as more general rules apply only if more specific ones have not already applied, as proposed in the Elsewhere convention of Kiparsky (1973:94).<sup>19</sup> Since context-sensitive defaults take precedence over or over-ride context-free defaults, defaults

---

<sup>19</sup>Kiparsky's Elsewhere Condition is as follows:

Elsewhere Condition (=Kiparsky (1973(4))

"Two adjacent rules of the form

$$A \rightarrow B / P \underline{Q}$$

$$C \rightarrow D / R \underline{S}$$

are disjunctively ordered if and only if:

- a. the set of strings that fit PAQ is a subset of the set of strings that fit RCS, and
- b. the structural changes of the two rules are either identical or incompatible."

Note that the Elsewhere Condition, where the more specific rule bleeds the more general one, applies only when the first rule is a special case relative to the second one; it does not apply when two rules are unrelated.

are inherently ordered: (14) bleeds (15), so (14) must apply prior to (15).

In this section, I argue that the association of complementizers with embedded clauses is an example of context-sensitive markedness. The fact that complementizers tend to occur in embedded clauses but not in matrix clauses is an example of the type of syntactic asymmetry that characterizes natural languages.<sup>20</sup> For example, the English complementizers *that* and *for* specify [ROOT-] clauses exclusively. Complementizers contrast in this respect with question expressions, which are like complementizers in that they must occur at the periphery of a clause but unlike complementizers in that they are grammatical in both root and non-root clauses. Question expressions and other 'fillers' are also unlike complementizers in that the former are sometimes inflected, and may be phrasal.<sup>21</sup>

The distribution of complementizers relative to the feature ROOT is characteristic of markedness.

---

<sup>20</sup>As noted above, the criteria for determining markedness proposed above are preliminary but I rely on them in order to make a concrete proposal.

<sup>21</sup>This difference does not exclude the possibility that fillers are a kind of marker, or even a kind of complementizer. As pointed out by Fodor (p.c.), question expressions have much stronger semantic content than complementizers so perhaps question expressions show the influence of purely syntactic phenomena less transparently. In Ch.6, I will suggest that fillers resemble markers in that they select syntactic properties of their sister node.

16.

a.

|            |
|------------|
| ROOT+      |
| COMPL null |

b.

|           |
|-----------|
| ROOT+     |
| COMPL lex |

c.

|            |
|------------|
| ROOT-      |
| COMPL null |

d.

|           |
|-----------|
| ROOT-     |
| COMPL lex |

(16a) and (16d) are more typical than (16b) and (16c). To the best of my knowledge, although languages which have only (16a) and (16d), but not (16b) and (16c) are documented, the converse is not true. Still, (16a) and (16d) are not syntactic universals: it is false that complementizers are obligatory in embedded clauses and it is false that complementizers are prohibited in matrix clauses.

For example, English does allow embedded clauses to occur without complementizers, at least with some verbs.

17. a. Robin believes Kim likes pandas.  
 b. \*Robin quipped Kim likes pandas.  
 c. Robin believes that Kim likes pandas.  
 d. Robin quipped that Kim likes pandas.

English does not allow root clauses to have complementizers. But in Latin, although complementizers typically are licensed only in embedded clauses, the complementizer *utinam*, which introduces wishes, is grammatical in both root and embedded clauses.

18. Atticus 3,3 cited in Lane (1903) (=Ch.3(21))  
 Utinam illum diem  
 Would that that-ACC-SING-MASC day-ACC-SING-MASC  
 videam [ROOT+]  
 see-PRESENT-SUBJUNCTIVE-1PERS-SING  
 'I hope that I may see that day.'

Penner and Bader (ms.:46) note that complementizers are found in root exclamatives in Germanic languages.

19. wo dass dä scho überau isch gsi!  
 where that he already everywhere is been  
 'Where he has been!'

In the case of (20), a sense of amazement is conveyed  
 (if this sentence is indeed [ROOT+]).

20. German (=Ch.3(22))  
*Dass er so gemein ist!*  
 'That he is so mean.'  
 '(I wonder) whether he is so mean.'

If markedness specifications are only context-free, the  
 ranking of relative markedness in (21) results in the  
 markedness specifications is as (22).

21. a. [u MARKING = null]            [m MARKING = lex]  
       b. [u ROOT = -]                [m ROOT = +]

|                        |               |               |               |
|------------------------|---------------|---------------|---------------|
|                        | <u>BEST</u>   | <u>MEDIUM</u> | <u>WORST</u>  |
| COMPLEMENTIZER<br>ROOT | u null<br>u + | m lex<br>u -  | u null<br>m - |

But (22) is not empirically correct. In fact, the evidence  
 suggests that (23) is correct:

|                        |               |              |
|------------------------|---------------|--------------|
|                        | <u>BETTER</u> | <u>WORSE</u> |
| COMPLEMENTIZER<br>ROOT | null<br>+     | lex<br>-     |
|                        | lex<br>+      | null<br>-    |

This can be achieved if the default assignment is context-  
 sensitive, as in (24).<sup>22</sup>

24. [u ROOT = -] / [COMPL lex]

---

<sup>22</sup>Alternatively, it could look like (1). I will consider  
 the relative merits of these two formulations below.

1. [u COMPL = lex] / [EMBEDDED +]

The context-sensitive markedness specification flips the markedness rankings for the feature ROOT. The following patterns of markedness ensue.

25. a.

|   |       |      |
|---|-------|------|
| u | ROOT+ |      |
| u | COMPL | null |

b.

|   |       |      |
|---|-------|------|
| m | ROOT- |      |
| u | COMPL | null |

26. a.

|   |       |     |
|---|-------|-----|
| u | ROOT- |     |
| m | COMPL | lex |

b.

|   |       |     |
|---|-------|-----|
| m | ROOT+ |     |
| m | COMPL | lex |

(25b) is marked relative to (25a), (26b) is marked relative to (26a), and (25b) and (26a) are equally marked. This captures exactly the markedness relations that the syntactic evidence supports.

I speculate that the underlying reason that embedded clauses typically are associated with complementizers is that complementizers signal the presence of an embedded clause, as an indication that the default [ROOT+] status is not applicable. This signal need not always be present, for example if there is another signal such as an adverbial adjunct (e.g., *after*), or if the presence of an interrogative filler, makes the complementizer unnecessary, as in (27).

27. I asked S[*if/whether/why* Kim likes pandas]. [ROOT-]

This example shows that context-free defaults alone are too crude to do a satisfactory job of linguistic description. Context-sensitive defaults are needed as well.

However, context-sensitive defaults must be formulated with care. Recall that the feature association that needs to be captured is that (28.b) is marked relative to (28.a) and that (29.b) is marked relative to (29.a).<sup>23</sup>

- |     |    |                     |    |                    |
|-----|----|---------------------|----|--------------------|
| 28. | a. | ROOT+<br>COMPL null | b. | ROOT+<br>COMPL lex |
| 29. | a. | ROOT-<br>COMPL null | b. | ROOT-<br>COMPL lex |

One way to capture this association with context-sensitive defaults like those of SPE is shown in (30).<sup>24</sup>

30. a. [u ROOT = [ROOT+]]  
       [m ROOT = [ROOT-]]
- b. [u MARKING = [[MARKING null]]]  
       [m MARKING = [[MARKING lex]]]
- c. [u MARKING = lex] / [ROOT-]  
       [m MARKING = null] / [ROOT-]

(30.a) is a context-free default that establishes the value of ROOT, (30.b) is a context-free default that establishes the value of MARKING, and (30.c) is a context-sensitive markedness specification that determines the default value

---

<sup>23</sup>From this point on, I assume HPSG representation unless otherwise indicated.

<sup>24</sup>Maxfield (1990) proposes (i) (in GPSG terms):

i. [COMP, +ROOT] --> [COMP [NIL]]

of MARKING in the context of the feature ROOT.<sup>25</sup> In this type of markedness specification, each value of ROOT is associated with a different markedness specification.<sup>26</sup> But there is no a priori reason to assume that ROOT is the context for MARKING rather than vice versa; an alternative to (30) might be (31).

31. a. [u ROOT = +]  
[m ROOT = -]
- b. [u MARKING = null]  
[m MARKING = lex]
- c. [u ROOT = -] / [MARKING lex]  
[m ROOT = +] / [MARKING lex]

(30) and (31) both capture the fact that complementizers are more likely to be associated with embedded clauses than root clauses; either can be interpreted as a syntactic reflection of the fact that complementizers universally mark a relation between a subordinate clause and a main clause.

---

<sup>25</sup>Technically, it is not necessary to list the marked values, since the specification of default values implies that other values are marked, but I list them here for clarity of exposition.

<sup>26</sup>The difference between context-sensitive defaults and elsewhere defaults can be seen from the fact that (30), repeated here as (i), is not equivalent to (?).

- i. a. [u MARKING = null]  
b. [m MARKING = lex]
- ii. a. [ROOT+] --> [u MARKING = null]  
[ROOT+] --> [m MARKING = lex]  
b. [ROOT-] --> [u MARKING = lex]  
[ROOT-] --> [m MARKING = null]

In (i), the unmarked value of MARKING is null everywhere. But in (ii) the feature ROOT must be present in order to determine the default values for MARKING; if there is some context where MARKING is appropriate but ROOT is not, then (30) does not establish a default value for MARKING.

But (30) and (31) capture subtly different syntactic generalizations, so which markedness specification (if either) is correct is an empirical question. (30.c) states that an embedded clause typically has a complementizer. It accounts for the fact that sentential subjects typically require a complementizer.

32. a. That Tom saw pandas surprised Mary.  
 b. \*Tom saw pandas surprised Mary.
33. a. For Tom to see pandas would surprise Mary.  
 b. \*Tom to see pandas would surprise Mary.

I take the *ing* suffix of the gerund as a complementizer for gerundive clauses.<sup>27</sup>

34. a. Seeing pandas surprised Mary.  
 b. \*For/that seeing pandas surprised Mary.
35. a. Tom's seeing pandas surprised Mary.  
 b. \*For/that Tom's seeing pandas surprised Mary.

(31.c) states that if there is a complementizer, the associated clause is typically [ROOT-]. This statement is correct, as far as it goes, but leaves unaccounted for the fact that sentential subjects typically need a

---

<sup>27</sup>Bresnan (1970) and references therein treat Poss-ing as complementizers, as in (i).

i. Kim's liking pandas is well-known.

Note that if infinitival *to* is treated as a (defective) auxiliary verb (as in P&S-94), (ii.a) must be analyzed as marked because the sentential subject has no complementizer.

ii. a. To see pandas would surprise Mary.

b. \*See pandas would surprise Mary.

One alternative is to hypothesize an invisible complementizer in (ii.a), an analysis adopted by P&S-94 for different reasons. This is undesirable because we are striving to avoid elements that are not phonologically realized. Another alternative is to treat infinitival *to* as a complementizer.

complementizer.<sup>28</sup> On this criterion, therefore, (31) is preferable to (30).

(30.c) also makes the incorrect prediction for the head daughter of Schema 4. According to (30.c), the lower S will typically be [MARKING lex].

- 36.
- $$\begin{array}{c} \text{S} \\ / \quad \backslash \\ \text{[MARKING that]} \quad \text{S} \\ \quad \quad \quad \text{[ROOT-, MARKING null]} \end{array}$$

This is not an accurate generalization about the head daughter of a Schema 4 tree such as (36). Normally, a clause has no more than one complementizer.

(30) is also not specific enough because it does not specify default values for enough features. In English, as in many other languages, question expressions and fillers may not occur in the same clause.

37. a. \*I wonder who that you saw.  
 b. \*I wonder that who you saw.  
 c. I wonder S[who you saw \_\_\_].

So (30.c) makes an incorrect generalization for a clause that contains an interrogative expression. If an embedded clause has an interrogative or a relative expression, the default for MARKING should be null. It is only when the embedded clause is neither interrogative nor relative that

---

<sup>28</sup>The GB solution, which is to require government of an empty complementizer, where government requires adjacency, (Aoun and Sportiche 1983) is not readily available in HPSG.

we expect a complementizer.<sup>29</sup> [QUE{}] and [REL{}] must be added to the context for MARKING in (30.c), revised here as (38).<sup>30</sup>

38. [u MARKING = MARKING lex] / [ROOT-,QUE{},REL{}]

In this case the evidence suggests that the correct markedness specifications are those in (31) rather than those in (30).

The subtle distinctions between (30) and (31) reflect the complexity involved in determining context-sensitive defaults. Even assuming that the only possible context is a local tree, within the local tree the relevant contextual feature might be on the mother, the head daughter, or the nonhead daughter. Moreover, other features, not specified in the markedness specification might call for a different context-sensitive default. So I conclude that the evidence supports context-sensitive defaults, but the proper formulation of such defaults can be determined only by careful linguistic research.

#### 4.1.3 OTHER PLAUSIBLE DEFAULTS

In spite of the lack of generally agreed upon criteria for defaults, linguists frequently presuppose them. Furthermore, different observers often assume the same defaults, suggesting that there is an underlying reality

---

<sup>29</sup>Unless interrogatives are a kind of marker, a possibility I will consider at a later point.

<sup>30</sup>What the markedness specifications for MARKING should be if either REL or WH is non-empty will be considered in Ch.6.

that linguistic theory needs to acknowledge. In addition to the proposals for defaults discussed above (especially pp.106-106 and pp.115-117), other candidates for defaults that are plausibly universal are these:

○ A full clause is an island in core grammar; extraction from clauses is marked relative to extraction within a clause (Koster 1978:567, Zaenen 1983:480; Fodor 1992b).

○ Preposition stranding is marked relative to pied piping (Fodor (1992b, fn.57). In other words, the default value for SLASH on a PP host is PP.

○ Word order (Travis (1989), Fodor and Crain (1990); see also Hawkins (1983) for extensive data on word order universals.)

39. a. Head-peripheral word order is unmarked.  
 b. Having direct objects on the opposite side of the head than other arguments is marked.  
 c. Having adjuncts on the opposite side of the head than arguments are marked.

○ Adjacency (of a verb and its direct argument) is the unmarked case (Berwick 1985:284).

○ Control of missing subjects by a higher subject when there is an intervening object is marked relative to control of the missing subject by the intervening object.<sup>31</sup> In other words, the control relation in (40.b) is marked relative to (40.a) (Pinker 1989).

40. a. I told Bill to leave.

---

<sup>31</sup>Pinker chose this example because experiments by C. Chomsky (1969) have shown that sentences like (40.b) are harder for children to interpret than sentences like (40.a).

b. I promised Bill to leave.

○ Defaults for Learnable Phrase Structure Grammar (Fodor 1992b)

41. a. -SLASH

b. -WH

c. X[SLASH X] ~~φ~~THE NEMBYER-A constraint is a default.)

○ The subject acts as the discourse topic by default (i.e. when there is no questioned element or topic or head of relative clauses). (Zaenen 1983:496)

○ The subject of a finite verb is [CASE nominative] by default.

○ Infinitive verbs by default do not have subjects.

So despite some criticism of the notion over years, and although relatively little work has been done on systematically applying the concept in syntax, markedness does seem robustly supported in general by asymmetrical feature associations, and by cross-linguistic tendencies that are not universal. Ultimately, the case for markedness rests upon cumulative evidence that it accounts for the data and captures syntactic generalizations better than accounts without syntactic markedness. Only systematic cross-language investigation can provide evidence for linguistically valid specifications of syntactic markedness. If the approach to markedness adopted in this thesis shows promise, cross-language research to formalize the notion of syntactic markedness and ascertain examples will be required.

#### 4.2 SYNTACTIC DEFAULTS AND LEARNING

Just as the existence of a distinction between root and matrix clauses complicates the grammar, it also complicates learning of the grammar. Learners must acquire and store extra syntactic patterns for phenomena that are only grammatical in root clauses, such as subject-auxiliary inversion. A 'least-effort' learner (discussed in Chapter 1) might make generalizations that don't take the distinction between root and embedded clauses into account at all.

Even if a learner consistently distinguishes [ROOT+] and [ROOT-], the distinctions between constructions permitted in root and embedded clauses show that it is not safe for the learner to generalize from root to embedded clauses, or vice versa. If the learner were to assume that any construction licensed in root clauses is also licensed in non-root clauses, it would then have to 'unlearn' these generalizations, at least for those phenomena which are asymmetrically distributed. As noted above, in the absence of negative evidence, a learner who makes this type of error would receive no input to signal that it had done so.

In addition to the descriptive facts, a markedness specification associating complementizers and embedded clauses is also justified by patterns of language learning. If learners were hypothesizing feature values *de novo*, at least some learners would be likely to hypothesize that complementizers are grammatical in both root and embedded

clauses. This is not the type of mistake that learners make -- I assume that if such a mistake were noticed, it would certainly have attracted attention in the literature.

I showed in the previous chapter that in the absence of negative evidence, a correct grammar can be acquired as long as the learner picks a specific non-disjunctive value as an initial guess. If the default value is incorrect, the input sentence will not be able to be generated; this supplies evidence to the learner that the grammar must be revised. However, a learner free to hypothesize its own default value might invent its own criteria for determining a default; this is very different than UG specified defaults which have a linguistic basis. Since there is no guarantee that personally selected default values will be linguistically sound, such defaults will not necessarily do any work in reducing the redundancy and bulkiness of the learned generalizations that need to be stored, nor can they play any role in accounting for cross-language variation. (See Fodor (1989 and 1992) for a more extensive discussion of this problem.)

I conclude that the evidence examined above supports the hypothesis that the innateness of markedness specification is responsible for the fact that certain feature values and co-occurrences tend to recur across languages. In the

rest of this dissertation, I assume that UG specifies, among many others, the following defaults:<sup>32,33</sup>

42. a. CONTEXT-FREE DEFAULTS

```
[u CAT = S]
[u VFORM = fin]
[u ROOT = +]
[u AUX = -]
[u INV = -]
[u MARKING = null]
[u SLASH = {}]
[u QUE = {}]
[u REL = {}]
```

b. CONTEXT-SENSITIVE DEFAULTS

```
[u ROOT = -] / [MARKING lex]
```

These default values shape both the substance of syntax and syntactic learning in equal measure. In terms of syntax, the defaults represent feature values and associations that typify natural language. But since markedness principles are part of UG, marked values will be allowed by some languages in some contexts. The different distribution of marked and unmarked features reflects the fact that languages may 'choose' from among the feature values made available by UG. The adoption of these default values does not "account for" these linguistic facts in a deep sense. But they do need to be "accounted for" in an everyday sense of their following from some aspect of the grammar. Once we acknowledge the existence of these asymmetries, *real* expla-

---

<sup>32</sup>I assume that all values not specified as unmarked by UG are implicitly marked.

<sup>33</sup>Technically, the value of CAT is the sort *cat*. S is short-hand for a *phrase*, with a HEAD of sort *verb*, whose SUBCAT value is saturated.

nations may follow; until these explanations can be discovered, documenting typical feature values must suffice.

From the point of view of learning, default values are the values that the learner assumes in the absence of evidence that marked values are correct. If these default values were not pre-specified by UG, we would expect that all values would be acquired equally easily. The fact that markedness principles are part of UG explains why marked values, which complicate the grammar, can be successfully learned. Without innate markedness specifications, marked values might never be noticed by learners and therefore would disappear in a generation or two.

The hypothesis that certain values are pre-specified in UG as marked suggests an answer to a question frequently asked about markedness: if marked features complicate grammars, why do languages use marked features at all? Marked features are needed to increase expressive power, so every language will avail itself of some of the marked features available in UG. On the other hand, no language will avail itself of all of the marked features available in UG; such a language would be extremely complex and hence very difficult to learn. Markedness thus reflects a trade-off between expressive power and complexity.<sup>34</sup> A

---

<sup>34</sup>Of course there may be feature values, both unmarked and marked, that are not attested in any extant language. These are actually part of UG but we have no way of knowing that. due to historical and sociological accidents.

language with only unmarked features would be impoverished with respect to its expressive power (Fodor 1992b).

It should also be emphasized that a construction that is marked need not sound awkward, or be difficult to process in a language which contains it. Preposition stranding, which is relatively rare across languages and is often cited as an example of a marked construction (Klein 1990), remains persistently grammatical in English despite stern prescriptivist attempts to abolish it.<sup>35</sup>

#### 4.3 INDETERMINATE DEFAULTS

I showed above that context-sensitive defaults are necessary if defaults are to capture recurring patterns of feature co-occurrence. However, under certain circumstances default values are 'indeterminate', in that context-sensitive defaults do not suffice to establish a unique default value in some contexts. Indeterminacy of default values occurs when context-sensitive defaults conflict, when the context for the context-sensitive default is missing, and when the context-sensitive defaults are circular. If default values are indeterminate, they have no predictive force vis a vis feature co-occurrence, and they make the feature stripping discussed in Chapter 3 dangerous. The

---

<sup>35</sup>Fodor (1992,fn.57) proposes the following defaults (in GPSG terms).

i. In PP[SLASH X], X is PP by default.

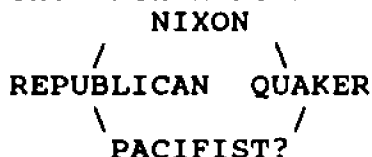
ii. In S --> X S/X, X is by default NP.

Thus preposition stranding satisfies (i), while pied piping satisfies (ii). This might explain why preposition stranding does not seem very marked.

problem of indeterminacy does not arise in a system with only context-free defaults, since each feature has exactly one default value which applies in all contexts.

□ **CONFLICTING DEFAULT VALUES** (Flickinger (1987), Shieber (1986a), Maxfield (1990), Carpenter (1993)). The 'Nixon diamond' is an often-cited example of conflicting default values, a problem well-known in the field of knowledge representation. Quakers are usually pacifists and Republicans are usually non-pacifists. Nixon is a Republican and a Quaker. Is he or is he not a pacifist?

43. The 'Nixon Diamond' Inheritance Hierarchy:

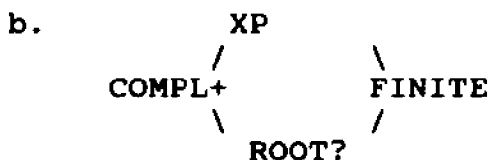


For a syntactic example, assume, for the sake of argument, that in the default case it is unmarked for a finite verb to be in a root clause.

44. a. [u ROOT = -] / [COMPL+]  
 b. [u ROOT = +] / [VFORM fin]

If a clause has a complementizer and is finite, as in the lower clause in (45.a), the value for ROOT is indeterminate.

45. a. I think [that Kim left]. [COMPL+, FINITE]



In this example, it seems fairly obvious that if (44.a) applies then (44.b) does not -- (44.a) bleeds (44.b). The

danger is that in a larger-scale hierarchy, there may well be many conflicting default values, resulting in massive indeterminacy that can be resolved only with a specific list specifying which defaults have priority over which.<sup>36</sup>

This indeterminacy can be prevented if markedness specifications are all listed in UG; however, if a solution based on less extravagant assumptions can be discovered, it would be preferable.

The other two problems, missing context and circular defaults, are associated not with the content of defaults but with the problem of determining what default values are in some context. In particular, they cause problems for the mechanism of feature stripping, as discussed in Maxfield (1990).

□ **MISSING CONTEXT** (Maxfield 1990)

Assume the defaults in (46).

46. [u A = -]            [u A = + / B+]  
       [u B = -]            [u B = + / C +]  
       [u C = -]

If the feature structure is (47.a), [A+] can be stripped off because it is the default in the context of [B+], and [B+] can be stripped off because it is in the context of [C+].

47. a. FEATURE STRUCTURE: [A+,B+,C+]  
       b. STORED VALUE: [C+]

But if the value of A is restored to (47.b) before the value of B, the incorrect default value will be instantiated.

---

<sup>36</sup>Kean (1981) proposes a system of phonological markedness with prioritized markedness specifications.

48. \*[A-,C+]

This problem arises because context-sensitive are ambiguous if the context is unknown.<sup>37</sup> The correct default value can be restored only restoration is carefully controlled. Specifically, the order of the stripping must be reversed. Either the learner would have to remember the order of stripping, or UG would have to specify the stripping order so that the learner would automatically reverse it.

□ CIRCULAR DEFAULTS:

The problem of circular defaults also arises from the ordered nature of feature stripping. Assume the default values in (49) (note that they are slightly different than (46), and the feature structure in (50).

49. [u A = -]            [u A = + / B+]  
       [u B = -]            [u B = + / C +]  
       [u C = -]            [u C = + / A +]

50. ACTUAL FEATURE STRUCTURE:        [A+,B+,C+]

Recall that by convention, any stored value is marked. In this case, which value is stored as the marked value depends upon the order in which features are stripped off. There are three possibilities.

51. a.                            b.                            c.  
       [A+,B+,C+]            [A+,B+,C+]            [A+,B+,C+]  
       [B+,C+]                [A+,B+]                [A+,C+]  
       [C+]                    [B+]                    [A+]

Circularity forces the stripping system into an inconsistency: whichever default value is left to be

---

<sup>37</sup>If the context can't be stripped away, then the problem with *u* and *m* goes away, as will be discussed in the next chapter.

stripped off last can never be stripped off because the default is missing. So [C+], which is an unmarked value in (51.a), will be stored as a result of stripping in the order shown. But the convention that any stored value is marked means that the learner will interpret [C+] as marked. However, if the learner strips off features as in (51.b), it will interpret [B+] as marked. Then the missing context problem discussed above comes into play.

In general, there are two types of solutions to the problem of ambiguity of context-sensitive defaults: 'traffic rule' solutions and 'design' solutions. In a traffic rule solution, the intrinsic organization of the system prevents the ambiguity from arising. For example, (44.a) bleeds (44.b). The danger with this type of solution is that very complex traffic rules will be needed to manage the interaction of these defaults. In the worst case situation, the ordering will be paradoxical, as was the case with ordered transformations in the Standard Theory of Chomsky (1965). Moreover, if UG does not specify an ordered list of defaults, it seems unlikely that the learner could be counted on to remember a complex ordering reliably enough to reverse it accurately; on the other hand, a model of UG which requires that UG include a long list of ordered defaults, though logically and empirically possible, is unattractive from the point of view of theoretical simplicity. A more theoretically attractive traffic

solution would be simple and principled, in that certain kinds of defaults automatically take precedence over others. The Elsewhere Condition of Kiparsky (1973) is an example of a traffic rule of this sort: the requirement that a context-sensitive default take precedence over an elsewhere default follows from the very definition of 'context-sensitive' and 'elsewhere'.

In a design solution, UG simply does not specify conflicting or ambiguous defaults. In any case where linguistics discover apparently conflicting defaults, the linguistic analysis is at fault and at least one of the defaults must be wrong.<sup>38</sup>

#### 4.4 DISCARDING SYNTACTIC DEFAULTS: THE SWITCH FROM GPSG TO HPSG

In addition to the problems with context-sensitive defaults discussed in the previous section, the proposal that an HPSG learning model incorporate defaults runs into another problem: HPSG eschews defaults, in spite of the fact that they were a central component in GPSG. P&S-87 (Ch.8, fn.4) refer to Flickinger's proposal for using defaults for "lexical exceptions and subregularities" without endorsing it. P&S-94 (Ch.1, fn.22) state "Except for expository convenience, HPSG does not employ any notion of defaults".

There are two main reasons for the abandonment of defaults

---

<sup>38</sup>In Ch. 5 I will propose context-sensitive defaults with a linguistically motivated, unambiguous precedence hierarchy; within each level of the precedence hierarchy UG specifies only one default per feature, thus precluding conflicting defaults.

in HPSG: first, defaults are not compatible with a declarative, monotonic system; second, the formulation of the principles adopted in GPSG resulted in infinite search and complex interaction of defaults with syntactic principles. However, although HPSG syntax is in principle default-free, it can be shown that, in practice, lexical defaults continue to play an important, albeit sometimes unacknowledged, role in HPSG theory.

The first problem with defaults is a general computational one, and not particular to syntax or to GPSG. Defaults are theoretically anomalous in the declarative, monotonic framework adopted by GPSG and HPSG (discussed in Chapter 2). Defaults are ordered relative to other principles in that they may be instantiated only after all other principles or rules that might bleed the default have been applied, and non-monotonic in that inheritance from the more general class to the more specific class is not strict.

Because of the absence of order in a declarative system, defaults are also not absolute constraints, and therefore not falsifiable.

Because of the semantics more or less standardly given to nonmonotonic statements, we leave open the possibility that the default may "fail" at any time. We have therefore allowed ourselves the excuse of appealing to the failure of a default, without having to justify reasons for failure. In describing a tendency rather than a condition, a default statement is not predictive enough to constitute a scientific hypothesis. ...this position is untenable if we wish

to be able to show mechanically that certain facts follow from the statements we employ.

Calder (1994:5)

In other words, the fact that a default statement is contained in a grammar does not mean that marked values not specified in the default statement are ungrammatical; it just means that the default statement is violated.

In addition, GKPS's system of defaults proved to be computationally intractable.

"(In GPSG) the real problem was the use of defaults to compare local trees to thousands of other local trees in order to determine acceptability. It was the notion that such decisions weren't entirely local and internal, but sensitive to the range of other possible analyses, that led to huge computational problems."

(Bob Carpenter p.c.)

This problem, which I will refer to as the 'trans-derivational problem', was a product of the particular architecture adopted by GKPS. In order to preserve monotonicity and declarativity, GPSG employ neither pre-existing defaults nor over-writing of defaults. Instead, GKPS define the notion of 'admissible projection': a projection of a rule is admissible only if it meets the defaults. All projections which do not satisfy the defaults are inadmissible and so must be thrown out. One step in determining whether a projection of a rule includes a default value for a feature not mentioned in a rule involves checking all candidate projections of the rule in order to be sure that the default holds applies all projections of the rule. If the non-default value is instantiated on even

one projection of the rule, then any projection of the rule which includes the default value is inadmissible. This means that all possible trees projected by a rule must be checked for grammaticality.

This problem is compounded by the interaction of the Control Agreement Principle or CAP (GKPS:83-94) with the requirement for tree admissibility. The CAP is intended to explain why certain categories in certain languages may manifest agreement, but others never do. Phenomena subsumed under the CAP include subject-verb agreement, verb-direct object agreement, agreement in case, number and gender between adjectives and nouns, control constructions, raising constructions, and termination of unbounded dependencies. In constructions governed by the CAP, a controller constituent determines certain properties of another constituent, the controllee. For example, N' controls DET when there is specifier-head agreement; NP controls the number of V in English subject-verb agreement.

Given the GKPS criteria for tree admissibility, a default cannot be established as false in a control structure, until it is determined whether a controller licenses the non-default value. So in a raising construction, for example, it is necessary to look at projections of rules other than the one which licenses the local tree to see if the marked structure is licensed by some rule.

GKPS utilize the fact that unless every member of the set of local trees composing a non-local tree is an

admissible projection, the non-local tree in question will be ruled illegal. The definition for assessing the validity of default overrides admits certain projections on the assumption that the overrides contained in the projection exist where they do for no other reason than that they are forced to be there, in order for certain local trees to connect with other local trees. At some point, however, the overrides must be introduced into the tree on the basis of necessity--usually because an ID (Immediate Dominance) rule or FCR (Feature Co-occurrence Restriction) forces the marked category into the set of local trees at that point.

Hukari and Levine (1986)

The operation of checking all possible projections of all rules which license one of the set of local trees contained in a local tree potentially involves checking an infinite number of projections, and is the source of the exponential search problem.

A related problem with the FSDs of GPSG, discussed in Shieber (1986b) and Maxfield (1990), is that they interact in complex ways with the HFP, the Control Agreement Principle, and the Foot Feature Principle.

On the GKPS view, there are no stages of derivation. There is just a tree, with both inherited and freely instantiated feature specifications. These two kinds of feature specifications must be differentiated (for the FFP (Foot Feature Principle) and FSDs at least). We could think of them (metaphorically) as written in different colored ink. The FSDs also need to know which feature specifications are required by the HFC and the CAP; and all the principles must be aware which feature specifications are required by the FCRs; so there must be more colors. It is as if each principle has its own colored ink and colors the features that it checks or requires to be present. Then some other principles can be made sensitive to some features but ignore others. ... the definition of local admissibility was complicated by the fact that lexical categories are exempted from the FSDs only if they covary with a sister, but nonlexical categories are

exempted if they co-vary with any other category in the local tree.

Maxfield(1990:85)

The details of how the principles interact is not relevant here; what is important is that the complexity of the principles themselves means that their interaction is very complex. Shieber (1986b:5) observed that the principles of GPSG "engender(ed an) "implicit precedence ordering" and suggested "a much simpler formulation" involving a procedural (ordered) mapping of the principles.

Rather than characterize the well-formed trees directly, we progress in two stages by procedurally characterizing the well-formedness axioms themselves (i.e. the universal principles), which in turn characterize the trees.

Shieber (1986b:1)

Shieber (1986b:6) proposed constraining the "combinatorily explosive" process of compiling the principles and constraints into context-free rules "by carefully choosing a system for stating constraints on local sets of nodes". His system thus retains defaults by allowing limited ordering of principles and constraints within a unification-based system.<sup>39</sup>

However, P&S-87 and P&S-94 chose to discard syntactic defaults rather than to revise their interpretation of them. But in the absence of defaults, HPSG has no mechanism to handle linguistic asymmetries; all feature values are subsumed by more general feature structures and no particular

---

<sup>39</sup>Maxfield (1990) proposes some modifications of Shieber's proposal.

feature value or set of feature values is more highly valued than any other. Typed feature structures capture universal or exceptionless generalizations, but there is no mechanism for representing, for example, that *accusative* is more highly valued than *oblique* as a value for CASE.

In addition, the desire to avoid defaults sometimes forces an awkward analysis of a problem that seems to call for them. Sag et al. (1992) discuss the problem of quirky case in Icelandic. In Icelandic, the 'normal' case for subject NPs of both finite and non-finite verbs is nominative, although quirky verbs may assign other cases such as dative and accusative to their subjects; the 'normal' case for object NPs, including those that have been 'raised' to object, is accusative, although a quirky verb that is the head of an infinitival complement assigns quirky case to its subject, even when the subject has been raised to object.

The insight behind the solution of Sag et al. is that the notion of defaults is needed.

The problem ... is that the notion of default case assignment is not being expressed. The default case for subjects of finite verb forms is nominative; the default case for direct objects is accusative. The non-quirky verbs obey these defaults; the quirky verbs do not. This is the fundamental intuition of the solution we now propose.

Sag et al.:310)

To capture the intuition of defaults without actually using them, Sag et al. adopt an analysis in which there are two features: DCASE (default case) and CASE. They speculate (p.310) that the value of DCASE is "associated with the

grammar rule that introduces subjects", so that all realized subjects (of both finite and non-finite verbs) are specified as [DCASE nom]. Non-quirky verbs bear the SUBCAT value in (52), where the CASE value and the DCASE value of the subject are structure-shared.

52. SUBCAT value of regular (non-quirky) verb in lexicon  
Sag et al. (29)

$$\text{SUBCAT} \left\langle \begin{array}{l} \boxed{\text{CAT np}} \\ \boxed{\text{CASE [1]}} \\ \boxed{\text{DCASE [1]}} \end{array} \right\rangle$$

Note that the verb's SUBCAT value does not specify the value of CASE or DCASE for the subject NP; rather the grammar assigns *nominative* as the value of DCASE. The value of DCASE is structure-shared with that of CASE in the SUBCAT value of a regular verb, so that the subject of a non-quirky verb is always *nominative*, as in (53).

53. SUBCAT value of regular (non-quirky) verb in the syntax

$$\text{SUBCAT} \left\langle \begin{array}{l} \boxed{\text{CAT np}} \\ \boxed{\text{CASE [1]nom}} \\ \boxed{\text{DCASE [1]}} \end{array} \right\rangle$$

But quirky verbs subcategorize for a special CASE value for their subjects.

54. SUBCAT value of dative-licensing quirky verb in lexicon

$$\text{SUBCAT} \left\langle \begin{array}{l} \boxed{\text{CAT np}} \\ \boxed{\text{CASE dat}} \end{array} \right\rangle$$

The grammar imposes [DCASE nom] on the realized subject of quirky verbs, but the value of CASE and DCASE is not identified for quirky verbs.

55. SUBCAT value of dative-licensing quirky verb in the syntax

SUBCAT  $\left( \begin{array}{l} \boxed{\text{CAT np}} \\ \boxed{\text{CASE dat}} \\ \boxed{\text{DCASE nom}} \end{array} \right)$

Therefore, in (55), "the constraint that realized subjects are specified as [DCASE nom] has no particular consequence" ((Sag et al.:311). So the morphologically realized case of a quirky verb like *áskotnaðist* ('luck into') will be dative.

This analysis relies on the fact that the values of CASE and DCASE are structure-shared by regular verbs but not by quirky verbs.<sup>40</sup> By this technicality, Sag et al. avoid a non-monotonic analysis in which a non-default value over-writes, and therefore erases, a default value; they also avoid a transformational analysis which is procedural because lexical case assignment takes place first, and then

---

<sup>40</sup>If this analysis has any psychological plausibility, we must assume that the adult speaker of Icelandic knows the two facts in (i):

i. a. Typically the values of CASE and DCASE are structure-shared.

b. If the values of CASE and DCASE are not identical, the value of CASE is morphologically realized.

Each of these facts about the grammar of Icelandic is potentially either universal, in which case it is arguably innate, or language-particular, in which case it must be learned.

If both (i.a) and (i.b) are universal, then UG must include explicit syntactic information, a central hypothesis of this thesis. But if either (i.a) or (i.b) is not universal, then syntactic learning is required as well as lexical learning.

the Case Filter ensures that structural case is assigned to all NPs that have not yet received Case.

Sag et al.'s solution works because of the dual features for case: DCASE and CASE. The value of DCASE is structurally determined; the value of CASE is lexically determined. This solution explicitly captures the intuition of syntactic or structural defaults while avoiding non-monotonicity and non-declarativity. However, although there are technically no defaults, in fact the reason that the solution works is that the value of CASE takes precedence over the value of DCASE if the two values are different. The cost of this solution is the need for two features to represent the property of case, instead of just one. This clumsiness is not damaging if there are relatively few such situations in natural language. But if there are numerous cases where linguistic data has to be forced into a default-free model, the weight of linguistic evidence suggests that a purely non-monotonic, declarative analysis of this kind is not adequate.

#### 4.5 LEXICAL DEFAULTS IN HPSG

Although syntactic defaults have been eliminated from HPSG, lexical defaults have not: they are an integral component of the lexical model which P&S-87 and P&S-94 presuppose, and they are frequently used in descriptive work. Lexical defaults are tolerated, although sometimes without explicit acknowledgement, because it appears that

they are more likely to be tractable computationally than syntactic defaults are. The primary reason is that lexical defaults are local in that they are not passed through trees by structural recursion; therefore they cannot be trans-derivational. In addition, Calder suggests that lexical defaults might solve the problem of the unfalsifiability of defaults.

(One solution) would be to prescribe a domain of finality. The system of the *Sound Pattern of English* (Chomsky and Halle 1968:402-403) assumes that only lexical specifications may contain values to be filled in by default and that such values are cashed out prior to manipulation of elements by rule.

Calder (1994:5)

So although the computational properties of defaults are as yet not well-understood, there appears to be a consensus that lexical defaults are preferable to defaults.

In fact, defaults play an important role in the lexical model presupposed by P&S-87 and P&S-94. In this model, the lexicon is organized as a multiple inheritance hierarchy of the kind frequently used in the field of knowledge representation.<sup>41</sup>

...the theory of grammar presented here relies crucially on complex lexical information, which determines, in accordance with general principles such as the HFP and the Subcategorization Principle, the essential grammatical properties of phrasal expressions. This does not mean, however, that HPSG relies on complex lexical stipulations, or that the presence of distinct lexical entries with shared properties leads to massive redundancy within the

---

<sup>41</sup>See Shieber (1986a (Ch.5, Section 8)) for an introduction to multiple inheritance hierarchies and default inheritance in the lexicon.

lexicon. As described in P&S-87 and Flickinger 1987 (see also Flickinger et al. 1985; Flickinger and Nerbonne 1992; and Fraser and Hudson 1992), properties of lexical entries and relationships among them are expressed in a concise and principled fashion in terms of classification by a 'multiple inheritance hierarchy' (P&S-87, sec. 8.1) and lexical (redundancy) rules (P&S-87, sec 8.2) respectively.

F&S-94:36

Lexical defaults play an important role in all of the lexical models to which P&S-94 refer in this citation. For example, Fraser and Hudson discuss the importance of defaults in multiple inheritance hierarchies.

Since the scientific study of language first began, a central concern of linguists has been the identification of linguistic generalizations and, where necessary the stating of exceptions to these generalizations. However, it is only within the last few years that linguists have begun to think of this process in terms of the construction of default inheritance hierarchies. This new way of envisaging old problems is attractive for at least three reasons. Firstly, it encourages linguists to be explicit not just about the relations that hold between individuals and classes, but also about the relations that hold between different classes. For example, where the nouns of a language have traditionally been assigned to some number of morphological paradigms, the default inheritance approach encourages the morphologist to pay attention to generalizations that cut across paradigms. If these generalizations are inherited, then there must be some shared super class to inherit from, and the system of word classes must be designed accordingly.

Secondly, whereas these generalizations have traditionally been class-based, in the inheritance approach they are based on typical cases and their features, any of which may be overridden. Thus the shading from core members of a class to peripheral members can be accommodated--indeed, the existence of peripheral members is predicted by the mechanism for overriding defaults. The third and more pragmatic reason why it is useful to recast well-known linguistic problems in terms of default inheritance is that there

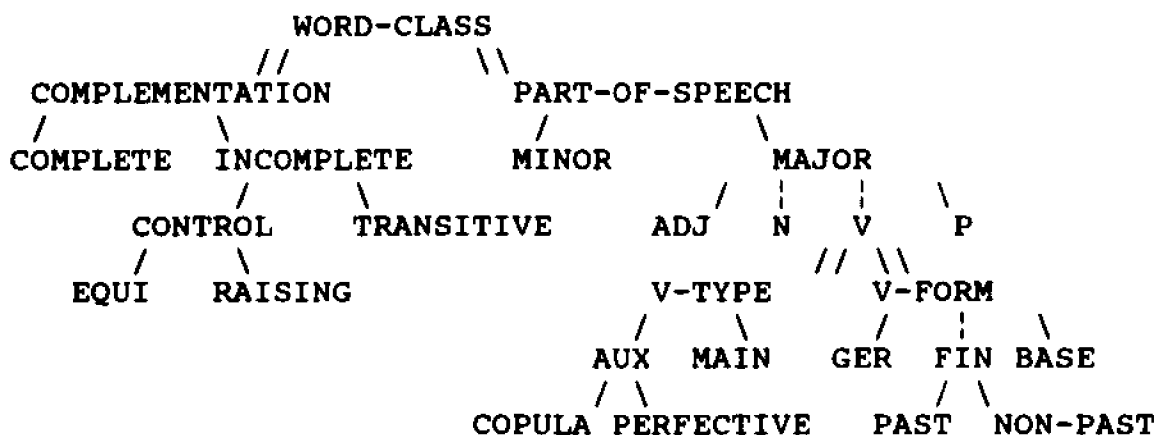
is a fairly well-developed--though by no means conclusive--body of knowledge on the subject in the artificial intelligence field of knowledge representation.

Fraser and Hudson (1992:133)

The inheritance hierarchy mentioned by P&S-94 is indeed a default hierarchy, as Fraser and Hudson describe it.

(56) shows a fragment of the word class hierarchy of Flickinger (1987). (The double lines (// and \\) represent a 'perspective' link to one of the dimensions into which the class is partitioned; the single lines (/ and \) indicate a 'subset' link between the parent class and a proper subset, within a given dimension. A word may inherit from all dimensions which are appropriate, but it can only inherit from nodes on one path within each dimension.)

56. Flickinger 1987(Ch.2(1),(2) and p.37)<sup>42</sup>



The verb *tried*, for example, inherits properties from the class EQUI which is within the dimension COMPLETATION; it inherits MAIN-VERB properties within the dimension V-TYPE, which is itself a dimension of PART-OF-SPEECH; it inherits

<sup>42</sup>See also the sort hierarchy of P&S-94:396.

PAST properties from the dimension VFORM, which is another dimension of PART-OF-SPEECH.

In Flickinger's model, classes are associated with default values, which may be over-ridden by default values of a subclass lower on the hierarchy or by an actual value in an individual lexical entry.<sup>43</sup> For example, Flickinger proposes the following default subcategorizations for main verbs that subcategorize for VP.

57. MAIN VERB [SUBCAT <VFORM inf>]      e.g. *plan* <to go>  
 AUXILIARY [SUBCAT<VFORM base>]      e.g. *will* <go>  
 PERFECTIVE [SUBCAT<VFORM perfect>] e.g. *has* <gone>

The default subcategorization for the PERFECTIVE class overrides the default subcategorization for the superclass AUX.

- 58.
- ```

      V -TYPE
     /      \
    AUX      MAIN
  [SUBCAT<VFORM base>] [SUBCAT <VFORM inf>]
   /      \
  COPULA  PERFECTIVE
         [SUBCAT<VFORM perfect>]
  
```

Since *has* is a member of the perfective class, it subcategorizes for [VFORM perfect].

The modal *ought* is an example of an exceptional lexical item which does not subcategorize for [VFORM base] as do most auxiliaries; instead it subcategorizes for [VFORM inf]. The lexical entry of *ought* includes the specification [SUBCAT <VFORM inf>] which over-rides the default subcategorization for the PERFECTIVE class.

⁴³See Warner (1992) for a more recent analysis of auxiliaries that also relies on defaults.

Without defaults to over-ride defaults of superclasses, the inheritance hierarchy would be flattened out, resulting in a loss of the economy of description provided by the depth of the inheritance hierarchy.

59.

```

                                V -TYPE
          /                       /                       \
PERFECTIVE          AUX          MAIN          COPULA
[SUBCAT<VFORM perf>] [SUBCAT<VFORM bse>] [SUBCAT<VFORM inf>]

```

In summary, defaults are central to the architecture of the lexical model endorsed by HPSG; without defaults, the lexicon would be complex and redundant because it would fail to capture many not-quite-universal patterns.

Moreover, lexical defaults tend to be presupposed in much of the descriptive work carried out in this framework, as if, in practice, they are indispensable. For example, Sag and Fodor (1994) say:

By default, a lexical entry has the empty list as its value for a nonlocal feature such as SLASH. Lexical entries that are produced by the (Complement Extraction Lexical Rule, which removes an element from the list of a word's complements, placing the complement instead onto a value of SLASH), however, contain non-empty SLASH specifications that will percolate up the tree until an appropriate binding environment is found ...

Sag and Fodor (WCCFL 1994:12)

Miller and Sag (1993:11) posit the Complement Cliticization Lexical Rule (CCLR). This rule applies only to constituents that are cliticizable, i.e. to a relatively small number of constituents that must be specified as [CL+]. In order to prevent the CCLR from applying to non-cliticizable constituents, they must be specified as [CL-];

otherwise the CCLR will apply to many constituents to which it should not. In fact, Miller and Sag assume [CL-] as the default state -- they do not explicitly include this feature in every lexical entry other than [CL+] entries. This is exactly parallel to the GPSG example concerning INV discussed in Section 4.1, except that the CCLR is called a lexical rule rather than a syntactic one.⁴⁴

As noted in Chapter 3, P&S-94 posit a part of speech *marker*, with the attribute MARKING, with values of sort *marking*. (Recall that these terms are not the same as those used in markedness theory.) The sort *marking* has two subsorts, *marked* and *unmarked*.

... *unmarked* is the default value, in the sense that it is the value borne by words other than markers.
(P&S-94:45)

The value [MARKING unmarked] is needed in the syntax to prevent phrases that do not have marker daughters from being analyzed as if they do have marker daughters. P&S appear to intend that lexical entries for words other than markers

⁴⁴Acquiring lexical rules that are not universal is like acquiring transformations in that the input to the rule needs to be deduced from the output. The input must be appropriately constrained so that the lexical rule does not over-generate. I will not consider this problem here.

However, it is worth noting that lexical rules are anomalous in HPSG because they are neither structural descriptions nor constraints. Therefore I assume that a desirable goal for HPSG, both from the point of view of HPSG theory and of learnability, is to find a substitute for lexical rules.

Of course, universal lexical rules that are innate pose no learning problem; if the default for each lexical item is that it is not subject to some lexical rule, positive evidence will tell the learner which lexical items are subject to the rule.

will have no overt specifications for MARKING, but will be assigned [MARKING unmarked] by some general process as they leave the lexicon.

In conclusion, lexical defaults serve two important functions in HPSG: they make the lexicon more compact, and they account for the presence of feature values that follow from no other principles. So although it is not yet known exactly what kinds of defaults are safe, lexical defaults are retained, at least in practice, because it appears likely that they are computationally more tractable than syntactic defaults are.

4.6 CONCLUSION

In this chapter, I have made two points which may appear irreconcilable. On the one hand, I have argued that syntactic defaults have a sound linguistic basis:

- Defaults account for syntactic asymmetries;
- Defaults impose predictable neutral values whose presence follows from no other syntactic principle;
- Defaults explain how grammars can represent compact syntactic generalizations which do not over-generalize or under-generalize;
- Defaults preclude the need for the grammar to store redundant values.

On the other hand, I have observed that syntactic defaults are problematic computationally:

- An implementation of defaults that is transderivational is evidently intractable;
- An implementation that involves ordering of rules is (at least partly) procedural;
- An implementation that involves over-writing of feature values is (at least partly) non-monotonic;
- Context-sensitive markedness specifications of the sort assumed in HPSG are potentially ambiguous.

These computational problems make it clear that no matter how great the linguistic merits, defaults are acceptable only if they are constrained in some way. Ideally, a way could be found to retain the linguistic advantages of defaults without the computational disadvantages. Further computational study of defaults to understand how defaults can be safely implemented is needed; in its absence, I take as my starting point the current *de facto* HPSG position, that there are no syntactic defaults and that the only defaults which may be tolerated are lexical.

In Chapter 5, I explore the consequences, for syntax and for learning, of defaults that are restricted to the lexicon. I propose a model for the acquisition of the syntactic properties of complementizers which relies on lexical defaults that do not interact with syntactic principles. I show that if lexical defaults are treated as

strict or falsifiable in the syntax, an interesting and explanatory learning model can be achieved.

CHAPTER 5 LEXICAL DEFAULTS AND HPSG LEARNING

The learning model that I propose in this dissertation is intended as a kind of experiment to test the plausibility of two hypotheses. The first is P&S-94's hypothesis, which is shared by the linguistic community in general, that all language-particular differences are lexical. If this hypothesis is correct, all language-particular learning, except for parameter setting, is lexical. P&S-94 analyze selectional properties of complementizers as lexical rather than syntactic, and I propose a model consistent with the architecture of HPSG to show how these properties can be acquired.

The second hypothesis that I explore is that of Fodor, that markedness relations that account for cross-language variation and that capture the asymmetrical feature associations of natural language also explain how learners can acquire a correct generalizations from positive evidence. In the previous chapter, I showed that defaults are linguistically sound and useful, and indeed virtually indispensable, but that their linguistic merits are apparently irreconcilable with the computational complexities of *syntactic* defaults. In this chapter, I argue (i) that the problems with syntactic defaults discussed in Chapter 4 can be avoided with lexical defaults; and (ii) that many syntactic defaults can be recast as lexical defaults.

To illustrate these points, I propose a model for acquiring the selectional properties of complementizers which involves only lexical learning, and in which defaults are restricted to the lexicon. The richness of the HPSG feature structure, which allows complementizers to specify syntactic properties of their heads, makes natural the treatment of apparently syntactic defaults as lexical.

During the learning process, the fully specified structures of local trees that license grammatical sentences of the language are plucked out of the syntax, *sans* universal syntactic features, and passed into the lexicon. There, defaults are factored out of the lexical feature structures of the input lexical entries, leaving only marked and therefore unpredictable values to be stored. The architecture of this learning process corresponds to the architecture of the grammar except that information flows in opposite directions: in learning, information from the lexical nodes of input trees is passed from the syntax into the lexicon for storage; in parsing and generation, lexical information passes from the lexicon, where it is stored, into the syntax, where it is a constraint on local trees.

The computational problems discussed in Chapter 4 are solved because defaults are strictly local and because allowable contexts for context-sensitive defaults are strictly limited. As I explain the learning process, I show how these problems are solved on a case by case basis, and I

recapitulate the solutions in Section 5.4.3. In Section 5.5, I discuss the problem of collapsing lexical specifications after they have been individually acquired, in order to form broader generalizations. Finally, I show how this model can be adapted to account for the problem of acquiring a correct grammar during the period of time when the learner does not yet recognize all relevant features.

In Chapter 6, I turn to the problem of learning the selectional problems of interrogative fillers such as *who* and *which*, and show that the learning model presented in this chapter accounts for acquisition of selectional properties of fillers from positive evidence. I consider the question of whether the existence of phrasal fillers is compatible with the hypothesis that all learning is lexical. In the final chapter, I explore further the question of whether lexical defaults and lexical learning suffice or whether syntactic learning and syntactic defaults are also needed for at least some phenomena.

5.1 LEXICAL VS. SYNTACTIC: THE DISTINCTION

Before turning to the model for acquiring the syntactic properties of complementizers, I consider the distinction between 'lexical' properties and 'syntactic' properties. As noted on p.68, much recent linguistic work assumes the existence of universal linguistic principles, and attributes all cross-linguistic variation to parameters and to the lexicon. On this view, 'true' syntactic learning consists only of

parameter setting.¹ All other apparently 'syntactic' learning is lexical.² The hypothesis that all learning is lexical is desirable from the point of view of learnability, because it precludes the necessity for the learner to choose between modifying the lexicon and modifying the syntax; only the lexicon can be modified.

A reasonable initial definition is that lexical properties are associated with particular lexical items and therefore stored in the lexicon. Syntactic properties are properties of phrases (XP or X') and therefore cannot be stored in the lexicon (at least according to a standard view of the lexicon).

However, in HPSG, many apparently phrasal properties which are not inherited from lexical categories (X^0) are captured quite neatly by selectional properties and by projection from lexical heads. Subcategorization of heads for their complements is a familiar lexical mechanism which captures information about the 'horizontal' relationship between sister nodes. In HPSG, additional mechanisms of this same general kinds exist for the lexicalization of selectional relationships by certain non-heads; for example,

¹Although P&S-94:31 rely on parameters "in response to the empirical demands of some language or other", they say that they are "inclined to view parameter-based accounts of cross-linguistic variation as highly speculative."

²For example, Chomsky (1993) assumes invariant syntactic principles, and conjectures that language variation is restricted to the lexicon and to phonology. See also Kroch (1994) and Hermon (1992).

as discussed in Chapter 3, complementizers specify properties of their sisters by the feature SPEC, and adjuncts select their heads via the feature MOD.

Many 'up and down' patterns of feature distribution are also captured lexically in HPSG, in the sense that the language-specific facts are recorded in the lexicon, and transmitted through the tree by general (innate) principles. For example, the HFP ensures that the (head) properties of the head daughter are structure-shared with the mother node and thus projected up through the tree (as discussed in Chapter 2). Schema 4 requires that the MARKING value of a marker daughter be identical to (structure-shared with) the MARKING value of the mother (as discussed in Chapter 3). So information from the lexical entry of a head is passed up through the tree where it constrains higher nodes.

Given the broader view of the lexicon as storing horizontal and vertical information, the initial definition of syntactic properties as phrasal is too narrow. A more accurate characterization is that syntactic properties are properties of phrases not inherited from a lexical constituent.

In this chapter, I present a model in which all language-particular differences are stored lexically; therefore all learning involves the learning of lexical entries. Whether there are generalizations across lexical entries that can or must be detected by learners is another

issue³ which falls under a broad line of research on the existence or non-existence of lexical rules, and how they might be acquired.⁴ The most recent extensive discussion of this is in Pinker (1989). The work reported there indicates that theories of lexical learning in this sense are in a somewhat unsettled state. However, there is a growing consensus that lexical learning is very conservative. Each lexical item is potentially idiosyncratic and therefore its properties must be acquired in a way that preserves any possible idiosyncracies. The acquisition of properties of lexical classes must not proceed in a way that is existence-predicting for the properties of particular lexical items. (Fodor 1992a and 1994a). The hypothesis that all learning is lexical, in the sense that only properties of lexical items are learned is consistent with any theory of lexical rules (including their non-existence) as long as conservative learning is somehow guaranteed.

³Generalizations that allow no lexical exceptions in any language would be arguably syntactic, even if they did concern properties of lexical categories. For example, it appears that the order of a verb and its object is never different for different verbs in the same language, though the order of adpositions and their objects does vary across items in a language (e.g., in Dutch). By the characterization I gave above, both of these would be lexical facts, though the former has no lexical exceptions. Some explanation is needed, but I will not pursue it further (See Kroch 1994).

⁴I assume that the process of 'deciding' which schemas are actually instantiated in the grammar that the learner is trying to acquire is part of string-to-structure learning (see Ch.1). Under the working assumption that the learner forms generalizations over fully specific local trees, the acquiring of syntactic generalizations is distinct from the string-to-structure process.

5.2 THE 'SYNTACTIC' PROPERTIES OF LEXICAL DEFAULTS

As noted above, syntactic defaults were central to the architecture of GPSG, but HPSG has only lexical defaults. In this section, I show that the richness of the HPSG feature structure for complementizers makes it linguistically natural to treat apparently syntactic defaults as lexical.

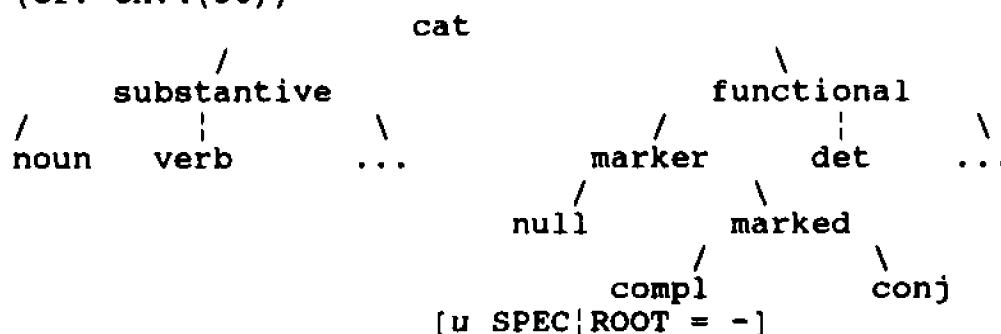
5.2.1 CONTEXT-SENSITIVE LEXICAL DEFAULTS

If facts that are apparently syntactic can be reduced to lexical facts, it follows naturally that context-sensitive defaults are associated with particular lexical entries. Note that the context-sensitive default association between complementizers and ROOT discussed in Chapter 4 is actually the association of a particular value of the feature ROOT with a lexical category that is represented by the feature [MARKING *compl*]. For example, the fact that UG specifies that complementizers are typically associated with embedded clauses (as discussed in Chapter 4) can be represented by (1). This is a lexical default statement in that the default is associated with the lexical class *complementizer*, a subsort of MARKING.

1. [MARKING *compl* [SPEC [u ROOT = -]]

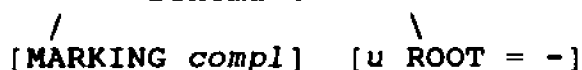
A context-sensitive default such as this would reside in the hierarchy of lexical categories, as proposed by Flickinger (1987).

2. (Cf. Ch.4(56))



Thus, the default statement in (1) is lexical in that it is stored in the lexicon. The selectional feature SPEC, allows the lexical entry of the complementizer to specify syntactic properties of its sister. Since complementizers are a kind of marker, which can only be the non-head daughter of Schema 4 (as discussed in Chapter 3), (1) is equivalent to the configuration in (3), but no other.

3. (cf. Ch.3(10))
Schema 4



Schema 4 is the only schema that licenses markers, so the default value [ROOT-] associated with the SPEC value of the complementizer can be instantiated on the head daughter of Schema 4. So in (1), as in (3), there is no ambiguity as to which node of the Schema 4 type local tree the default value is associated with. If the default were associated with the complementizer node itself (an association that is in fact impossible since ROOT is not appropriate for complementizers), the default would be stated as in (4).

4. [u ROOT = -] / [MARKING compl]

The fact that a lexical default can be represented unambiguously in a local tree reflects the fact that the lexical default in (1) captures a relationship between sister nodes that in earlier versions of phrase structure grammar, might have been represented by a phrase structure rule:

5. XP[COMPL +] --> COMPL, XP[u ROOT = -]

As a result, context-sensitive defaults like (1) and (5) specify exactly which node of a local tree the default applies to.⁵

With lexical defaults, the maximal domain for the default is a lexical entry. The context for a context-sensitive default like (1) is a feature or collection of features in that lexical entry. In the present case, it is the feature MARKING which characterizes a complementizer. The context-sensitive default applies only to a lexical entry for a complementizer which is in the language. If a language has no complementizers or if a word is of a different syntactic category, the context-sensitive default associated with complementizers is irrelevant.

5.2.2 CONTEXT-FREE LEXICAL DEFAULTS

In the previous section, I suggested that a context-sensitive defaults whose context is a lexical entry is quite naturally interpreted as lexical. But it is less obvious

⁵Fodor (1992b) distinguished between the FSDs of GPSG, which apply only within a single category, and Local Tree Defaults (LTDs) which are cross-category defaults. With context-sensitive default statements like (1) and (5), LTDs are not needed.

what it means to say that context-free defaults are lexical. In principle, the context-free default values proposed in Chapter 4 can be associated with a lexical feature in the same way that [ROOT-] was in (1).

6.

MARKING compl, SPEC	[u ROOT = -] [u VFORM = fin] [u INV = -] [u AUX = -] [u MARKING = null] [u SLASH = {}] [u QUE = {}] [u REL = {}]
---------------------	---

However, this solution is syntactically undesirable. Strict lexicalization of defaults would mean that all default values would be associated with particular lexical items. As a consequence, an economical generalization could not be represented. In particular, the fact that context-free defaults typically represent the neutral value would not be captured. The lexicon would contain numerous duplicative default specifications, resulting in the loss of the economy of storage offered by defaults.⁶

⁶Adopting the analysis of Sag et al. for CASE (Section 4.4) to complementizers would make this feature structure twice as bulky; for each feature, there would also have to be another feature holding the default value.

i.

MARKING compl, SPEC	[D-EMB [1] [+]] {EMB [1]} [D-VFORM [2] fin] [VFORM [2]] ...
---------------------	---

Equally important, the clear distributional asymmetries that characterize the association of complementizers and ROOT do not characterize the association of complementizers with other feature values besides [ROOT-]. These neutral defaults apply more generally than to just one (or a few) lexical categories; they crop up over and over again. For example, there is no basis for the association of the lexical category complementizers with a particular value of VFORM. Since more languages have complementizers for [VFORM fin] than for [VFORM inf], there is some evidence that for complementizers, the default value for VFORM might be *fin*. But the distributional evidence suggests that the default for VFORM in matrix clauses is also *fin*; this follows if *fin* is a context-free default. Therefore re-interpretation of context-free defaults as lexical context-sensitive defaults is not satisfactory.

But although they are not associated with particular lexical classes, context-free defaults are nevertheless lexical in the model presented here in that they can be instantiated on lexical elements. Context-free defaults thus satisfy the definition of lexical in Section 5.1. Values associated with lexical elements will be passed to phrasal categories in the syntax. I will suggest below that

defaults are imposed on lexical entries by the operation of default unification.⁷

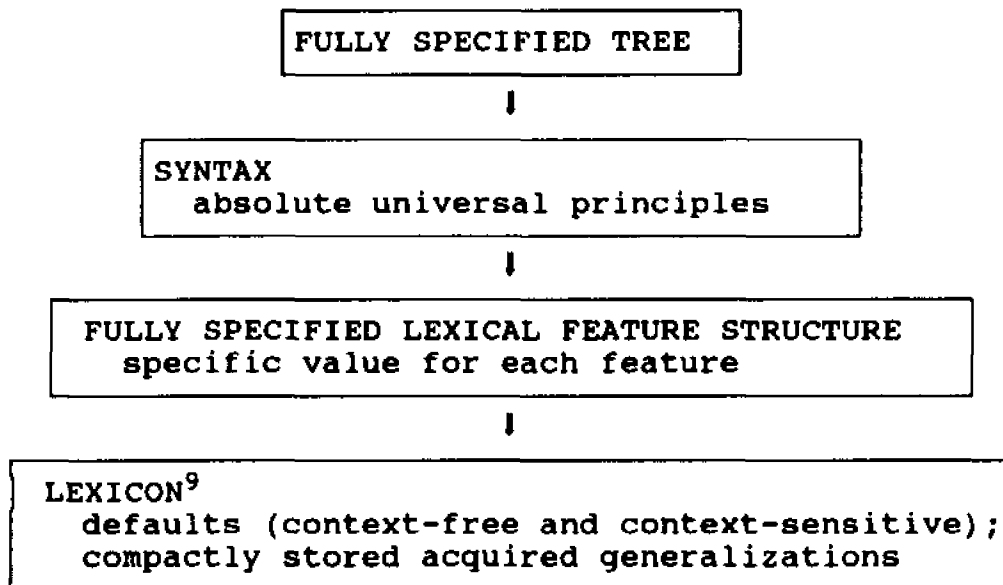
5.3 THE LEARNING PROCESS

I begin with a general picture of the learning process which is based fundamentally on the GPSG model of Fodor and Maxfield, but is consistent with the architecture of HPSG, and the assumption that defaults are only in the lexicon.⁸ Below I give an example of how the learning process works.

⁷Fodor (1994b) suggests that defaults are instantiated on lexical entries by universal lexical rules, which do not have to be learned.

⁸Recall (from Ch.1) that in most of this dissertation, I make the working assumption that the learner makes generalizations from fully specific, correct trees. In Sec. 5.7, I briefly consider the problem of how the learner is able to acquire correct generalizations during the time period when it does not yet recognize relevant syntactic distinctions.

7. HPSG LEARNING MODEL (Preliminary version)



This learning model has the following properties.

- There is a natural boundary between derivation in syntax and storage in the lexicon.
 - The syntax operates on fully specific local trees to ensure that all universal principles and constraints are satisfied. The domain of these principles is the syntax; as noted in Chapter 1, I assume the syntactic principles of P&S-94, Chs. 1-8, modified as necessary for successful learning.
 - The lexicon is a separate sub-system in which acquired lexical information is stored.
- Specific values for all appropriate features must be passed to lexical nodes. Lexical nodes thus carry accurate

⁹I assume that there are universal feature co-occurrence restrictions (in HPSG, they take the form of typed feature structures) in the lexicon and in the syntax.

language-particular information. These nodes are 'plucked' out of the syntax, *sans* universal syntactic properties, for shipment to the lexicon.

□ The fully specified lexical node is a generalization that licenses a type of embedded clause, just as a fully specified phrase structure rule is. In the syntax, the lexical node licenses not only the clause token from which the lexical generalization was derived, but also an infinite number of clauses of the same type.

□ At some moment in learning, the fully specified lexical node passes from the syntax into the lexicon for storage. This is consistent with the time-bound nature of learning; at some point in time, the learner has only universal linguistic knowledge; at some later point, the learner has acquired language-particular knowledge.

□ The lexicon is the exclusive domain of defaults, which enter the syntax only as strict or falsifiable values. Since default values, *qua* defaults, never enter the syntax, they are not passed through syntactic trees, so that the exponential search problem that beset the syntactic defaults of GPSG is averted.

□ Generalization by feature omission takes place in the lexicon, where the fully specified lexical feature structure is 'factored' by default unification into two parts: i) UG-specified defaults and ii) language-particular information

that needs to be learned and stored because it is not predictable.

□ Indeterminate context-sensitive default values are precluded because the class of entities that may serve as the context for context-sensitive defaults is strictly limited by UG. In this chapter, the only context I consider is lexical categories.

5.3.1 ACQUISITION OF SPEC VALUES

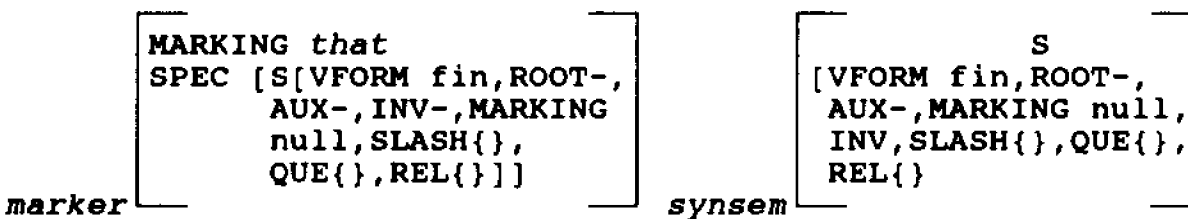
Consider again the learning process from the perspective of a learner who encounters the structure in (8.b) for the embedded clause in (8.a).¹⁰

8. (Cf. Ch.3(40))

a. I think S[that S[Kim likes pandas]]

b.

S
[VFORM fin, ROOT-, AUX-, INV-,
MARKING that, SLASH {}, QUE {}, REL {}]
/ \



All universal syntactic principles and constraints operate on the fully specific local tree. The principles which I consider in this example include the following:

9. a. **Head Feature Principle (HFP):** The head features of the head daughter and the mother node must be identical.

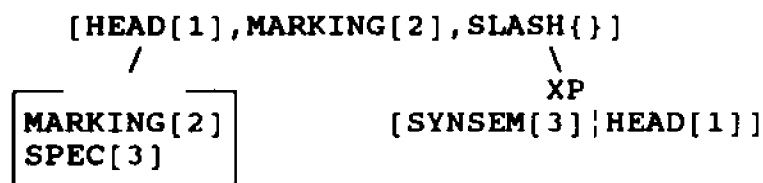
¹⁰As above, I use a simplified representation of the complex feature structures of HPSG; for further details refer to Ch.2.

- b. Immediate Dominance Schema 4: In a local tree, the MARKING value of the non-head daughter is structure-shared with that of the mother.
- c. SPEC Principle: Complementizers specify SYNSEM values of the head daughter.
- d. NONLOCAL Feature Principle: The NONLOCAL values of the mother node are structure-shared with the union of the NONLOCAL values on all its non-head daughters.

These universal principles apply in any order in the syntax; they therefore unify to form the constraint on a Schema 4 type local tree in (10.a). The variables in (10.a) point to the features in (10.b).

10.

- a. UNIFICATION OF HFP, SCHEMA 4, AND MARKING PRINCIPLE:



- b. [1] [VFORM fin, ROOT-, AUX-] HFP
- [2] that Schema 4
- [3] S{HEAD[1], MARKING null, SLASH{}} SPEC Principle
- [4] [SLASH{ }, QUE{ }, REL{ }] NONLOCAL Feature Principle

Recall from Chapter 2 that unification is a single operation that merges consistent information, but that conceptually it can be thought of as consisting of two processes: one process checks that all information that is being unified is consistent, and one process copies information from each of the unified feature structures to the unified structure. So unification can be understood as simultaneously checking for grammaticality and constructing a local tree (or feature structure).

13. FULLY SPECIFIED LEXICAL FEATURE STRUCTURE FOR *that*:

[5] [SPEC S[VFORM fin,ROOT-,AUX-,INV-
MARKING null, SLASH{}], QUE{}], REL{}]]
marker [MARKING *that*]

So in this model the lexical node is in taken out of the syntax and passed into the lexicon, sans universal syntactic information; this is the converse of what happened in the GPSG learning model where universal features were stripped off, leaving a syntactic rule behind. Note that some version of the Lexical Feature Principle is needed in any feature-based system in which the lexicon is distinct from the syntax. An advantage of the Lexical Feature Principle is that it eliminates the need for the Store-all Principle (Chapter 3(47)), which forced the learner to store all features for the purposes of precluding over-generalization at the cost of drastic under-generalization.

The fully specified lexical entry in (13) is a generalization in that it licenses a potentially infinite set of *that*-clauses that are of the same type as (8.b) and therefore equivalent to a syntactic rule from which predictable syntactic information has been stripped off. However, (13) is unnecessarily bulky, in large part because it includes predictable default values. Recall that in the GPSG model, the syntactic rule acquired from the local tree (Chapter 3(5.b)), was similarly bulky and redundant before default values were stripped off. In the GPSG model, generalization by stripping off of default values takes

place in the syntax. In this model, the process of generalization by feature omission takes place in the lexicon; default values are thus strictly lexical in that they are accessible only to information stored in the lexicon.

The defaults which I assume are those in (14); the context-free defaults in (14.b) apply to any feature if a context-sensitive default such as (14.a) does not take precedence.

14. a. CONTEXT-SENSITIVE DEFAULT(=(1), cf.(2))
 [MARKING *compl* [SPEC [u ROOT = -]]
- b. CONTEXT-FREE DEFAULTS (cf. 4(42.a))

u	CAT	=	S
u	VFORM	=	fin
u	ROOT	=	+
u	AUX	=	-
u	INV	=	-
u	MARKING	=	null
u	SLASH	=	{ }
u	QUE	=	{ }
u	REL	=	{ }

The precedence of context-sensitive defaults over context-free ones is definitional, so the hierarchy in (15) is universal. Note that it is simpler than that of the Elsewhere Principle of Kiparsky (1973), which employs a metric of specificity of context to determine which levels take precedence.¹¹ (A \leftarrow B means that A is lower in the precedence hierarchy than B.)

¹¹In principle, there could be any number of levels in the Precedence Hierarchy as long as there is a finite number of pre-specified representations, with a clear order of precedence.

15. **PRECEDENCE HIERARCHY 1**
 CONTEXT-FREE DEFAULT \leftarrow CONTEXT-SENSITIVE DEFAULT

Since context-sensitive defaults take precedence over context-free defaults, the defaults for the SPEC value of complementizers are those in (16.c).

16.

a. (=14.b)
 CONTEXT-FREE
 DEFAULTS

```
u CAT = S
u VFORM = fin
u ROOT = +
u AUX = -
u INV = -
u MARKING = null
u SLASH = {}
u QUE = {}
u REL = {}
```

b. (=14.a)
 CONTEXT-SENSITIVE
 DEFAULTS

```
[MARKING compl,
 u SPEC|ROOT=-]
```

c.
 DEFAULTS FOR SPEC
 OF COMPLEMENTIZERS

```
u CAT = S
u VFORM = fin
u ROOT = -
u AUX = -
u INV = -
u MARKING = null
u SLASH = {}
u QUE = {}
u REL = {}
```

That attested values take precedence over defaults is fundamental to the nature of learning and intrinsic to the distinction between defaults and attested values. Therefore (17) is also universal, where an attested value is a strict value that has been transferred from the syntax into the lexicon.

17. **PRECEDENCE HIERARCHY 2**
 DEFAULT \leftarrow ATTESTED VALUE

(15) and (17) can be combined as (18), a universal constraint on the lexicon.

18. **PRECEDENCE HIERARCHY 3**
 CONTEXT-FREE DEFAULT \leftarrow CONTEXT-SENSITIVE DEFAULT \leftarrow
 ATTESTED VALUE

It is the attested values that are recorded in the lexicon, except where they agree with defaults.

The defaults in (20.b) are 'conservatively subtracted' out of the fully specific lexical entry so that only marked values are recorded in the acquired generalization from which predictable default values are omitted (cf. the *add conservatively* of Shieber 1986a). In order to simplify the somewhat awkward representation of default values with *u*, and to represent visually their low rank in the precedence hierarchy, I represent features associated with default values in italics from now on; thus [CAT S] means that CAT is associated with its default value S. (Recall that by 'input', I mean a correct, fully specific feature structure.)

19. (= (8.a))
I think S[that S[Kim likes pandas]].

20.

a. INPUT

CAT S
ROOT -
VFORM inf
AUX -
INV -
MARKING null
SLASH {}
QUE {}
REL {}

b. UG DEFAULTS

<i>CAT S</i>
<i>ROOT -</i>
<i>VFORM fin</i>
<i>AUX -</i>
<i>INV -</i>
<i>MARKING null</i>
<i>SLASH {}</i>
<i>QUE {}</i>
<i>REL {}</i>

c. STORED

In this elementary example, the default values are identical to the actual values, so no marked features need to be recorded in acquired generalization. [] denotes the maximally succinct generalization, in which the complementizer specifies a sister consisting only of default values, so that no marked values need to be stored.

21. [MARKING *that*, SPEC[]]

However, in more complex cases, some of the features licensed by the complementizer are marked and therefore need to be stored. For example, the learner of English will encounter sentences with subordinate clauses such as the one in (22.a).

22. a. I would like S[for [Kim to support Sandy]].

b.
$$\begin{array}{c} \text{S} \\ \text{[VFORM inf, ROOT-, AUX-, INV-, MARKING for, SLASH{}]} \\ / \qquad \qquad \qquad \backslash \end{array}$$

<i>marker</i>	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> MARKING <i>for</i> SPEC [S[VFORM inf, ROOT-, AUX+, INV-, MARKING null, SLASH{}], QUE{}, REL{}] </div>	<i>synsem</i>	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> $\begin{array}{c} \text{S} \\ \text{[VFORM inf, ROOT-,} \\ \text{AUX-, INV-, MARKING} \\ \text{null, SLASH{}}, \\ \text{QUE{}}, \text{REL{}]} \end{array}$ </div>
---------------	---	---------------	--

The SPEC value of *for* contains two marked values, [VFORM inf] and [AUX+].¹² The difference between the 'input structure', i.e., those values which license the sentence that the learner encountered, and the defaults means that the learner needs to acquire the values of VFORM and AUX.

¹²Recall that I follow P&S-94 and Pullum (1982) in assuming that *to* is a defective auxiliary verb.

23.

a. (=16.c)

INPUT STRUCTURE

```

CAT S
ROOT -
VFORM inf
AUX +
INV -
MARKING null
SLASH{}
QUE{}
REL{}

```

b. (=13)

UG DEFAULTS

```

CAT S
ROOT -
VFORM fin
AUX -
INV -
MARKING null
SLASH {}
QUE {}
REL {}

```

c.

STORED

```

VFORM inf
AUX +

```

The acquired generalization (23.c) is stored in the lexical entry of the complementizer *for*.

24. [MARKING *for*, SPEC[VFORM inf,AUX+]]

The acquired generalization is simply a compact form of the fully specific lexical node from which all predictable information has been removed; the feature values which have been omitted can be interpreted only as default values.

(24) is a linguistically perspicacious feature description in which only marked values are specified; it is appropriately more complex than (21), a maximally succinct entry in which no language-particular information needs to be stored. 'Conservative subtraction' like that in (23) thus allows the learner to acquire a compact generalization from which predictable features are omitted. However, subtraction is inherently procedural, and therefore incompatible with the declarative theory of HPSG. In Section 5.3.2, I propose that in an HPSG model, this subtraction should be understood instead as default unification.

Note that if the learner were free to hypothesize any default value for any feature, this learning process would not guarantee successful learning (see Chapter 3). If the learner were to hypothesize disjunctive defaults (e.g. [VFORM fin \vee inf]), the Subset Principle would be violated, so disjunctive defaults would have to be ruled out. If the learner were to pick defaults that weren't linguistically motivated (e.g., [INV+] or [ROOT+]), the defaults would not produce any economy of storage in most cases.

Two conditions are necessary to ensure that the learner can successfully acquire (24) from positive evidence alone. First, UG must specify a single default value for each feature at each level of precedence; second, lexical defaults must be treated as 'strict' or actual values in the syntax. I will address these points in turn.

The Unique Default Value Principle is a version of the Specific Defaults Principle of Fodor (Ch.3(9)).

25. UNIQUE DEFAULT VALUE PRINCIPLE:

UG must specify exactly one (non-disjunctive) default for every feature of every lexical item, unless the value of the feature is universally fixed, or universally free.

This means that UG will specify exactly one context-free default value for every feature whose value is not universally fixed or universally free; it also means that UG will supply context-sensitive defaults that do not conflict with each other. (Since context-sensitive defaults bleed context-free defaults, it does not matter if context-free

defaults and context-sensitive ones conflict.) (25) ensures that learning is possible in accord with the Subset Principle because a specific default is available to be 'falsified' (if necessary) by positive evidence. (25) also precludes conflicting context-sensitive defaults where the context is a lexical category. For example, UG may not include (?) and (26).

26.¹³ [MARKING *compl* [SPEC|ROOT+]]

Second, lexical defaults must be treated as 'strict' or actual values in the syntax; strict values differ from defaults in that they are falsifiable. UG imposes on the learner the initial hypothesis that the SPEC values of any complementizer are those in (16.c). But values labelled as defaults in the lexicon are treated as strict values in the syntax so that positive evidence 'informs' the learner if a value on a lexical node is incorrect. For example, UG initially specifies the context-free default *fin* for VFORM. When the learner 'discovers' that [VFORM *fin*] does not license the input sentence, it will replace *fin* with a value that does license the sentence, in this case *inf*, in the

¹³Note that *compl* is a sort (or type), as discussed in Ch.2, while [SPEC|ROOT+] represents the default value of SPEC|ROOT. It would be preferable to represent the default value as gray or shadowy relative to the input value, but technical problems prevent this.

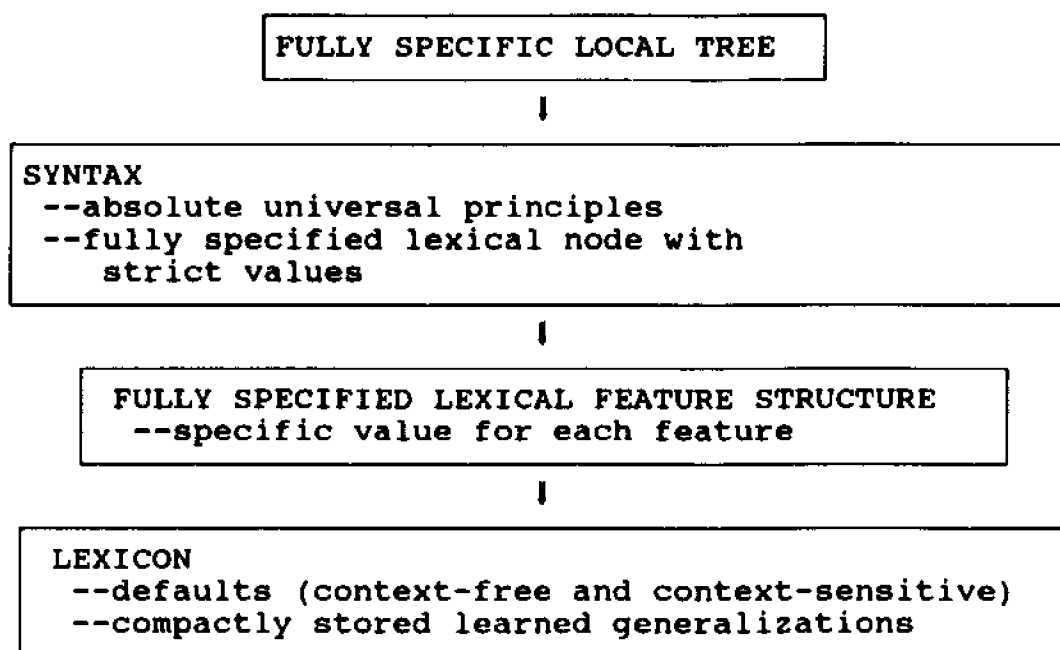
tree.¹⁴ So default values license ungrammatical values until the learner encounters positive evidence that a strict value should be replaced by a value that licenses the input.¹⁵ When the fully specific lexical node enters the lexicon, marked values replace default values in the lexical entry.

(7) should therefore be replaced by the more precise

(27):

¹⁴The evidence that causes the learner to revise its parse need not be purely syntactic; semantic and pragmatic evidence must also play an important role.

¹⁵This analysis predicts that learners will produce grammatical errors in the early stages of learning, before attested values replace strict values. It is also possible that learners will spend some time in an intermediate learning stage where they 'flip' back and forth between possible values until they acquire sufficiently strong evidence that the marked value is correct. This may be responsible for the errors that children are known to make during the acquisition of English subject-auxiliary inversion. (See Weinberg 1990b for a GB analysis of the acquisition of subject-auxiliary inversion in which markedness plays a central role.)

27. HPSG LEARNING MODEL (Final version)¹⁶

(27) depicts a learning process in which acquisition of correct generalizations from positive evidence is virtually guaranteed.

5.3.2 GENERALIZATION BY DEFAULT UNIFICATION

Conservative subtraction like that in (23) is procedural and therefore not compatible with HPSG. For the GPSG model, Fodor (1990:248) proposed the 'vacuum cleaner model' as a feature stripping mechanism which "extracts universal generalizations about the language, and leaves behind everything that is language-specific". In this section, I show that show that the vacuum cleaner model can

¹⁶Presumably universal requirements on feature co-occurrence, which in HPSG are restrictions on typed feature structures, and in GPSG were expressed as Feature Cooccurrence Restrictions could also be factored out, but I leave open the question of whether this information is factored out in the syntax or the lexicon.

be characterized technically in terms of 'default unification' (Bouma (1992); see also the 'add conservatively' of Shieber (1986a), and the 'skeptical' and 'credulous' unification of Carpenter (1993); Copestake (1993) compares these approaches). Default unification allows the learner to factor out default values automatically, without active comparison or reasoning.

The operation of default unification, although not formally equivalent to the conservative subtraction in (23), produces the same language-particular generalization.¹⁷ In (28), as in (23), the acquired generalization records the difference between the input structure and the default values. (% \sqcup represents default unification.)

28.

a. (=23.b)
UG DEFAULTS

CAT S
ROOT -
VFORM fin
AUX -
INV -
MARKING null
SLASH {}
QUE {}
REL {}

% \sqcup

b. (23.c)
STORED

VFORM inf
AUX +

=

c. (=23.a)
INPUT

CAT S
ROOT -
VFORM inf
AUX +
INV -
MARKING null
SLASH {}
QUE {}
REL {}

Since, as noted in Chapter 2, unification is a single operation which merges consistent information by checking

¹⁷ In certain cases of re-entrancy, the result of default unification is mathematically indeterminate. (Carpenter 1993) Since default values are strictly local, and structure-sharing is in the domain of the syntax, lexical defaults should not be involved in re-entrancy.

for consistency and by adding together consistent information, unification can also solve for the third feature structure if either of the other two are known. The information in (28.a) and (28.c) can therefore be used to solve for (28.b). The defaults in (28.a) are available to the learner because they are in UG; the information in (28.c) is available because it has been passed from the lexical node of a syntactic tree into the lexicon. Conceptually, the learner can be thought of as using default unification to record in (28.b) the information in (28.c) that is different from the information in (28.a).

Default unification is compatible with HPSG because it is declarative, procedurally neutral, and never fails.¹⁸ (Bouma 1992). Default unification is also fully monotonic for non-default values, but partially non-monotonic for default values. In other words, default unification preserves all information in non-default values (i.e. unification never over-rides non-default information) but default values are over-ridden by actual values if the actual values conflict with the default values. Default unification thus captures the requisite ordering of the learning process: at some point in time, the learner's experience with sentences in its language leads it to store

¹⁸This is different from standard unification which fails in (i) because the values of PERSON are inconsistent.
 i. [VFORM fin] \cup [VFORM inf] = \perp

marked values which take precedence over UG-specified default values.

Default unification with relevant UG-specified default values is an obligatory constraint on all lexical entries in order to ensure that stored lexical entries are as compact as the UG-specified defaults allow, and, as will be discussed below, it also ensures that a fully specified lexical entry passes out of the syntax into the lexicon. As a result, no special operation of feature stripping is required.

Moreover, given the precedence specifications in (18), the operation of merging context-sensitive defaults with context-free defaults can also be represented by default unification, as in (29).

29.

a. (cf. 16.a)
CONTEXT-FREE
DEFAULTS

```

CAT S
VFORM fin
ROOT +
AUX -
INV -
MARKING null
SLASH {}
QUE {}
REL {}

```

b. (cf. 16.b)
CONTEXT-SENSITIVE
DEFAULTS

$\% \sqcup$ MARKING compl,
SPEC|ROOT -

c. (cf. 16.c)
DEFAULTS FOR SPEC
OF COMPLEMENTIZERS

```

CAT S
VFORM fin
ROOT -
AUX -
INV -
MARKING null
SLASH {}
QUE {}
REL {}

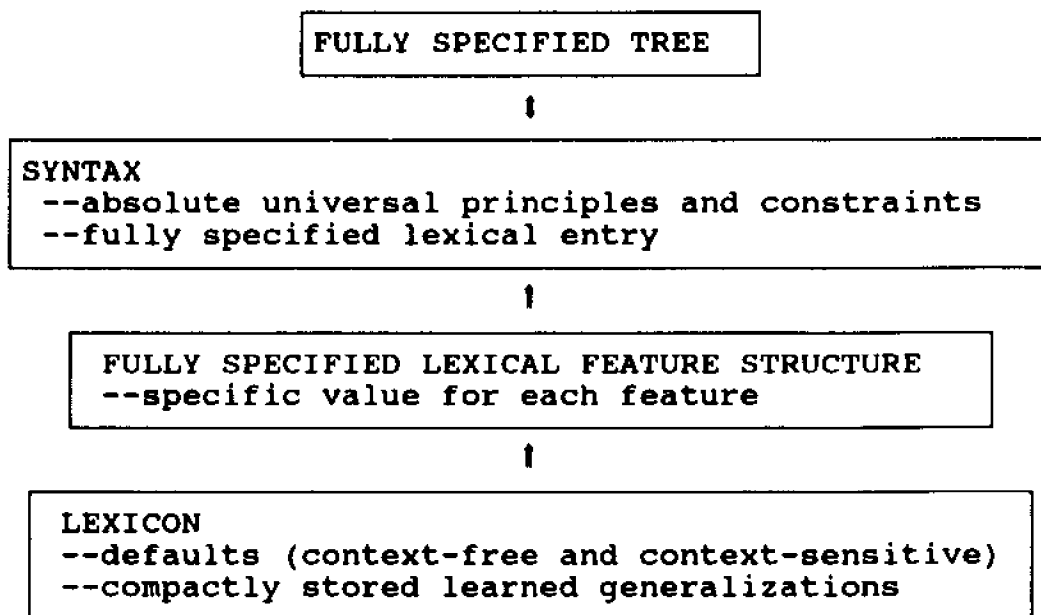
```

So default unification, which is needed for the interaction of actual values with context-sensitive defaults values in the lexicon, also can be used to model the interaction of context-free defaults values with context-sensitive ones.

5.3.3 THE LEARNING MODEL AND GRAMMAR MODEL

An advantage of this learning process is that the architecture of the grammar model and the learning model is the same, but information flows differently in learning than in other grammatical processes.

30. GRAMMAR MODEL (cf.7)



At some moment, in parsing and in generation, the fully specified lexical entry passes from the lexicon, where it is stored in compact form, to the syntax, where it is a lexical node that serves as a (language-particular) constraint on the tree. This is the converse of what happens in learning, where the fully specified lexical node is transferred from the syntax to the lexicon.

Crucially, in parsing and in generation, as in learning, all values, including UG-specified defaults, are treated as strict (falsifiable) values in the syntax. The

fact that the SPEC value of the marker passes out of the lexicon into the syntax is what allows it to do real syntactic work: in the syntax, the value of SPEC is a positive constraint on a local tree imposed by the lexical entry for *that*.

The compact lexical entry resides in the lexicon, along with defaults. Default unification is an obligatory operation in the lexicon; this guarantees that a fully specified feature structure, rather than a compact feature description, will enter the syntax from the lexicon. It also guarantees that the compact feature description will be stored in the lexicon rather than the fully specified feature structure. So the learning model in (7) is fully compatible with the syntactic architecture of P&S-94. In this system, the universal markedness specifications must be accessible to learners and to mature speakers of a language. If the default values were not accessible to the learner, then the learner would have to store fully specified lexical nodes, unwieldy as they are.¹⁹ So learners' grammars would be more complex than adults' grammars; this is logically possible but implausible.

If universal defaults serve as the basis for forming compact generalizations, it follows that these defaults also

¹⁹Recall that if defaults are learner-particular rather than universal (see Section 3.1), generalization is still strictly possible by the mechanisms suggested here, but whether these defaults would lead to any storage efficiency depends upon which values the learner picks as defaults.

play an important role in the grammar of mature speakers. Since the language-particular generalizations that the learner acquires are the result of factoring out universal default values, the default values must remain in the grammar of adult speakers in order to supply the values that the acquired grammar does not explicitly record. Moreover, if it can be proved that learners can function with defaults in learning, it can also be proved that performance mechanisms can function with a reduced grammar, so there is no reason for adults to have a larger grammar.

Recall from Chapter 2 that generalization by omission of features has the potential to lead to over-generalization because a feature structure in which fewer features are specified is more general than one in which more features are specified. In this model, default values are factored out of lexical entries for storage and restored to the lexical entry before it is passed to the syntax for parsing and generation. Over-generalization by virtue of missing features is prevented because feature structures descriptions in which default values are not specified are not permitted to enter the syntax.

5.4 COMPUTATIONAL SAFEGUARDS

I showed above that lexical defaults capture apparently syntactic information in the HPSG framework. In this section, I review the ways in which lexical defaults prevent the problems with syntactic defaults that plagued GPSG. In

general, I suggest that UG is designed so as to preclude problems such as transderivationality and indeterminacy; the seriousness of these problems suggests that the brains of human beings must be 'wired' so as to virtually guarantee that learners do not hypothesize intractable grammars.

5.4.1 LEXICAL DEFAULTS

GPSG's problems with defaults arose because of the unduly complex interaction of defaults with syntactic principles, as discussed in Section 4.4. In this section, I show that the systematic segregation of lexical defaults from syntactic principles prevents those problems.

The first problem, that defaults are unfalsifiable, is averted because only strict values enter the syntax. The Unique Default Values Principle ensures that exactly one default value is assigned to each feature in the lexicon; these defaults enter the syntax as strict values which may be replaced in local trees only if they do not license the input sentence.²⁰ Evidence that default values are ungrammatical in a particular context does not come solely from universal syntactic principles, which by definition must license default values, but from the trees of the language.

²⁰Note that the fact that a local tree does not generate an input structure 'motivates' the learner to replace the incorrect value in the tree. But since the language-particular values are not stored in syntactic rules, syntactic values are not over-written or over-ridden by acquired information. Recall from Ch.1 that I have set aside the problem of how string-to-structure learning works.

The transderivationality of defaults in GPSG was an outgrowth of the process of checking whether trees with default values were admissible (Section 4.4). Confining defaults to the lexicon where they are strictly local and therefore static prevents this problem. Defaults fill in predictable properties of lexical entries but never interact directly with syntactic principles and so are not passed between local trees.

Restriction of defaults to the lexicon also precludes the ordering problems resulting from the complex interaction of GPSG's FSDs with syntactic principles. Since only fully specified lexical entries, with either strict or attested values, enter the syntax, default values are independent of syntactic principles.

I conclude that a model in which defaults are "cached out in the lexicon", as suggested by Calder (1994; see p.159), offers a linguistically plausible solution to the problems with defaults that plagued GPSG. Within the lexicon, the intrinsic properties of different kinds of information establish a simple and principled precedence hierarchy. The restriction of default values to the lexicon means that that only feature values associated with fully specific nodes which can be shown to be incorrect by positive evidence occur in the syntax.

5.4.2 DETERMINATE DEFAULTS

Indeterminate default values are a potentially serious problem in a learning process in which defaults play a central role in accounting for cross-linguistic asymmetries in feature co-occurrence and for successful learning. I assume that UG is designed in such a way that a specific default is available for every (lexical) feature which is not universally fixed or universally free. In this section, I recapitulate the mechanisms that preclude indeterminate defaults.

Before proceeding, I introduce two additional principles, which, although trivial, are crucial to preventing indeterminate defaults. The Lexical Category Principle requires that every lexical node belong to only one lexical category in the syntax (e.g., a lexical node is a noun, and not simultaneously a verb or any other lexical category). Furthermore, the specification of the lexical category is an intrinsic part of the lexical entry that may not be factored out.

31. LEXICAL CATEGORY PRINCIPLE:

- a. Every lexical node in the syntax must be associated with exactly one lexical category.
- b. The lexical category must not be stripped off of the lexical entry by any principle or default value.

Another trivial default is the Hierarchical Context Principle.

32. HIERARCHICAL CONTEXT PRINCIPLE:

A feature lower in the lexical hierarchy may not be the context for a feature higher in the hierarchy.

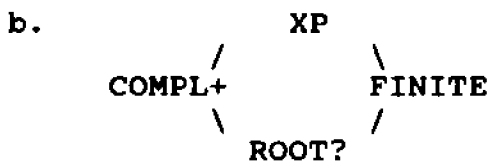
The Lexical Category Principle and the Hierarchical Context Principle together ensure that there will be only one default per feature per context and that the context for a context-sensitive default will be available to establish the appropriate default.

Recall that conflicting default values occur when more than one default assignment statement applies in a single context. For example, assuming the defaults in (29), if a clause has a complementizer and is finite, as in the lower clause in (34.a), the value for ROOT is indeterminate.

33. a. [u ROOT = -] / [COMPL+]
 b. [u ROOT = +] / [VFORM fin]

34. (=Ch.3(45))

a. I think [that Kim left]. [COMPL+,FINITE]



A diamond-shaped inheritance hierarchy like that in (34.b) is outlawed by the Unique Default Value Principle, which requires that UG assign exactly one default per feature per lexical category. In addition, the Lexical Category Principle specifies that the context for a context-sensitive default is a lexical category, which may not be stripped off. The lexical category COMPLEMENTIZER is the context for the default [ROOT-]. As a non-lexical feature, the at-

tribute [VFORM fin] is insufficiently specific and so cannot by itself be the context for a context-sensitive default.²¹

The problems of missing context for context-sensitive defaults and of circularity are prevented in the same way. These problems are caused when the context for a context-sensitive default is not known, so that the correct default value cannot be determined. In this case, the Lexical Category Principle ensures that lexical categories (such as noun or verb) may not be stripped off of lexical entries, so that lexical categories are 'safe' contexts for context-sensitive defaults. What other categories are 'safe' contexts requires further study.

There remains one other potential source of conflicting defaults; if other categories beside lexical ones are permitted to be the context for context-sensitive defaults, there is a possibility that different partitions in a multiple inheritance hierarchy will specify conflicting defaults for the same feature. As an initial hypothesis, I speculate that UG is designed so that it does not specify conflicting defaults. Flickinger (1987:60) speculates as to some possible precedence conventions, but further research is required to discover a solution which is computationally tractable and linguistically motivated. For now, I conclude

²¹A more plausible default is that V[VFORM fin] is sub-categorized by default for an overt nominative subject.

that the lexical category complementizer is a 'safe' context for context-sensitive defaults. Any extension of the class of features which may be the context for context-sensitive defaults must be adopted with care to ensure that indeterminate defaults do not arise.

5.4.3 SUMMARY

I have now demonstrated that learners can successfully acquire lexical entries from positive evidence. Defaults are central to the learning process, just as they were in the GPSG learning model of Fodor and in GPSG itself. Since lexical defaults carry syntactic information, the apparently irreconcilable clash between the evidence that defaults are linguistically sound and the computational problems caused by syntactic defaults is resolved by lexical defaults that store syntactic information. The interaction of defaults with other types of information is constrained by unambiguous innate traffic rules. The design of UG precludes indeterminate defaults without ordering of specific defaults.²² Over-generalization resulting from missing features is averted by the requirement that each lexical entry enter the syntax with a single specific value for every feature. In fact, the system as it stands suffers from not generalizing enough. In the next section, I consider the problem of generalizing over SPEC values by collapsing.

²²This means that if two defaults appear to conflict, at least one of them must be wrong.

5.5 GENERALIZATION BY COLLAPSING OF SPEC VALUES

In this section, I consider how the learner can collapse SPEC values so that they can be stored compactly, without under-generating or over-generating, and without explicit comparison of feature structures by the learner. Recall from Chapter 3 that the learner of English whose task it is to acquire SPEC values from positive evidence alone will encounter (at least) the following SPEC values.²³

35. (=Ch.3(42))
 a. I think [that Kim likes pandas].
 b. that SPEC S[VFORM fin,ROOT-,AUX-,INV-,MARKING null,SLASH{},QUE{},REL{ }]²⁴
36. (=Ch.3(48.b))
 a. What animal do you think [that Kim likes ___]?
 b. that SPEC S[VFORM fin,ROOT-,AUX-,INV-,MARKING null,SLASH{[1]},QUE{},REL{ }]
37. (=Ch.3(49))
 a. I think [that Kim might like pandas].
 b. that SPEC S[fin,ROOT-,INV-,AUX+,MARKING null,SLASH{},QUE{},REL{ }]
38. (=Ch.3(50))
 a. What animal do you think that Kim might like ___?
 b. that SPEC S[VFORM fin,ROOT-,INV-,AUX+, MARKING null,SLASH{[1]},QUE{},REL{ }]

²³Since they are not relevant to the discussion, I do not distinguish between INHERITED|NONLOCAL features and TO-BIND|NONLOCAL features, or between singleton and set values of NONLOCAL features. These distinctions do become important in Ch.6.

²⁴For brevity, I will specify the value of 'MARKING' but not the feature name from now on.

39. (=Ch.3(51.a))²⁵
 a. I demand *that* he throw away that paper.
 b. *that* SPEC S[VFORM base,ROOT-,AUX-,INV-,MARKING null,SLASH{},QUE{},REL{}]
40. (=Ch.3(52))
 a. What do you demand *that* he throw away now?
 b. *that* SPEC S[VFORM base,ROOT-,AUX-,INV-,MARKING null,SLASH{[1]},QUE{},REL{}]
41. (=Ch.3(53.b))
 a. I would like *for* [Kim to support Sandy].
 b. *for* SPEC S[VFORM inf,ROOT-,AUX-,INV-,MARKING null,SLASH{},QUE{},REL{}]
42. (=Ch.3(54.b))
 a. Who would you like *for* S[Kim to support ___]?
 b. *for* SPEC S[VFORM inf,ROOT-,AUX-,INV-,MARKING null,SLASH{[1]},QUE{},REL{}]

Factoring out defaults by default unification eliminates most of the redundancy in the SPEC values, leaving the marked values in (43) and (44) to be stored.

43. a. *that* SPEC []
 b. *that* SPEC [SLASH{[1]}]
 c. *that* SPEC [AUX+]
 d. *that* SPEC [AUX+,SLASH{[1]}]
 e. *that* SPEC [VFORM base]
 f. *that* SPEC [VFORM base,SLASH{[1]}]
44. a. *for* SPEC [VFORM inf, AUX+]
 b. *for* SPEC [VFORM inf, AUX+, SLASH{[1]}]

I showed above (Sec. 3.6.2) that conservative disjunction (where two feature structures differ only by a single feature value) could be allowed only in fully specified feature structures. However, collapsing feature descriptions by disjunction, after defaults have been factored out of the fully specific values of the local tree,

²⁵Verbs like *demand* and *request* might also be analyzed as being subcategorized for [VFORM subjunctive].

leads to a problem: how can the learner distinguish between default values that don't actually occur in its language and default values that do? Suppose that there were a language English', identical to English except that *for* selects only the marked value of SLASH, {[1]}, but not the unmarked value, {}.²⁶ If the learner of English' *assumes* that any default value is licensed in its language, it will license an overly powerful generalization like (45), in violation of the Subset Principle.

45. *for* SPEC [VFORM inf, AUX+, SLASH{}V{[1]}]

In order for collapsing to be done safely on reduced representations with features omitted, it is sufficient to add another variation of conservative disjunction: if there are stored in the lexicon two feature descriptions that are identical except that one has no value specified for some feature (so that it is the default), and the other has the marked value for that feature, then collapsing via disjunction is allowed. In English, (44.a) and (44.b) can be collapsed via conservative disjunction into (45). Since (44.a) will never occur on a fully specific tree in English', (45) will not be stored in the English' lexicon.

For this version of conservative disjunction to work, it is crucial that the learner store SPEC values such as (43.a), the maximally succinct SPEC value; otherwise it will

²⁶ This is the case with the Irish complementizer *aL* which is specified only for clauses with a gap. (See Ch.3(35.b)).

have no evidence that its language licenses a SPEC value with all default values. But as long as the values in (43.a) are treated by the learner as attested values, any pair of them that meets the condition for conservative disjunction may be collapsed. So while (43.a) can be collapsed with (43.b), (43.c), or (43.f), it cannot be collapsed with the other SPEC values in (43). If (43.a) and (43.b), (43.c) and (43.d), and (43.e) and (43.f) are collapsed by conservative disjunction, (46) results.

46. a. that SPEC [SLASH {} \vee {[1]}]
 b. that SPEC [AUX+, SLASH {} \vee {[1]}]
 c. that SPEC [VFORM base, SLASH {} \vee {[1]}]

(Of course, this is not the only order in which such collapsing could occur.) (46.a) and (46.c) could then be collapsed by conservative disjunction into (47).

47. that SPEC [VFORM fin \vee base, SLASH {} \vee {[1]}]

However, from the point of view of 'mechanical' language acquisition, an unfortunate property of generalization by any kind of disjunction is that the learner must notice that two feature structures or feature descriptions are sufficiently similar that they may be collapsed (Fodor 1990). So collapsing by conservative disjunction, while imaginable, is a quite different operation than generalization via unification which is part of the fundamental working of the grammar, and a model that could do without it would be welcome.

5.6 CONCLUSION

In conclusion, the learning model proposed in this chapter satisfies the learnability requirements imposed by the logic of language learning and is linguistically plausible.

LEARNING:

- A least-effort learner acquires the grammar of its language, automatically, from positive evidence, in accordance with the Subset Principle.
- Acquired knowledge is stored compactly in the lexicon in the form of feature descriptions consisting only of marked values, without unnecessary redundancy.
- The mechanism of default unification provides a natural treatment of defaults that is psychologically plausible, results in learning of the desired generalizations, and demands only consistent application of principled precedence specifications.
- Under-generalization is prevented by default unification and conservative disjunction. Over-generalization is prevented because default values are imposed in lexical entries in such a way that they can always be 'corrected' by attested syntactic values; the mechanisms that are responsible for this include the Unique Default Value Principle, the Uniqueness Principle, and the distinction between lexical defaults and strict syntactic values.

SYNTAX:

- Defaults account for predictable feature values and for asymmetrical feature distribution across languages.
- The principles of P&S-94 remain intact, except for modifications necessary for learnability reasons.
- Language-particular information is stored in lexicon, as in P&S-94.
- Defaults are confined to the lexicon, as in Flickinger (1987), and in current descriptive work carried out in the HPSG framework.
- The richness of HPSG feature structures allows information that linguists usually consider 'syntactic' to be stored in the lexicon.
- The context of context-sensitive defaults is sufficiently specific to preclude ambiguous contexts.

COMPUTATIONAL

- The segregation of lexical defaults from syntactic principles means that defaults are strictly local and static.
- The context of context-sensitive defaults is constrained so as to preclude indeterminacy.

In the next chapter, I show that the selectional properties of interrogative fillers can be acquired in the same way that the selectional properties of complementizers are, for both lexical and phrasal fillers.

5.7 APPENDIX: ACQUIRING SYNTACTIC GENERALIZATIONS BEFORE THE STRING-TO-STRUCTURE PROCESS IS COMPLETE

In order to focus on the acquisition of syntactic generalizations, I have made the working assumption throughout most of this dissertation that the learner forms generalizations about fully specific local trees. In this section, I digress to consider the problem of how a learner can safely acquire and store generalizations at the stage when it has not yet acquired values for every feature present in the target feature structure?²⁷ This problem arises when the learner recognizes some, but not all of the features in the target structure. If the learner stores a generalization which does not mention features which it does not yet recognize, it will have acquired an over-generalization that cannot be falsified by positive evidence.²⁸

On the idealized assumption that learning is instantaneous, an attested marked value is always available to override a default value, as in (28), repeated here as (48):

²⁷Many learning models do not mention this problem at all. Fodor (p.c.) notes that the idealization to instantaneous learning of Chomsky (1965) keeps it hidden.

²⁸Reasons for the failure to recognize a feature (i.e. to recognize a linguistic distinction that the feature represents) might be lack of maturity or lack of capacity. Lack of maturity means that the child has not yet attained the cognitive ability to recognize a feature; lack of capacity means that the child cannot yet manage such a large number of features at one time, even though it presumably has the cognitive ability to recognize all relevant linguistic distinctions.

48.

a. UG DEFAULTS

VFORM	<i>fin</i>
ROOT	-
AUX	-
INV	-
MARKING	<i>null</i>
SLASH	{}
QUE	{}
REL	{}

%

b. STORED

VFORM	<i>inf</i>
AUX	+

=

c. INPUT

VFORM	<i>inf</i>
ROOT	-
AUX	+
INV	-
MARKING	<i>null</i>
SLASH	{}
QUE	{}
REL	{}

However, suppose that a learner who has not yet learned to recognize the distinction that SLASH represents, has stored the SPEC value in (49) for an imaginary complementizer which I will call *xcompl*, which specifies infinitive clauses with gaps, but not without.

49. *xcompl* SPEC[VFORM *inf*]

(49) is correct in so far as the learner's knowledge extends, but incorrect in that [SLASH{[1]}] is missing. At some later point, after it has learned to recognize SLASH, the learner will acquire (50).

50. *xcompl* SPEC[VFORM *inf*, SLASH{[1]}]

Nothing in (50) tells the learner that the previously acquired (49) is incorrect. The reason that the lexical entry for *xcompl* in (?) licenses a default which is not actually attested is that the lexical entry for *xcompl* is stored before the learner recognizes SLASH, so that SLASH is unspecified in the feature descriptions. What has gone wrong is that 'accidental' omission of a feature is being confused with the use of omission as a notational convention to signal the default value. 'Default' in the sense of an

innately pre-set value, and 'default' in the sense of a value that need not be overtly specified, are not the same. What is needed is a distinction between unattested defaults and attested defaults. This can be done by selecting any symbol distinct from [-], [+], and [+V-] to represent unattested values. I adopt ?, used by Pinker (1984:103ff.) for what he calls 'preemptability'.

Before any learning has taken place, the 'learned' values can be represented by a feature structure consisting entirely of ?s, one for each appropriate feature, as in (51).

51. INITIAL STATE

a. UG

DEFAULTS

ROOT -
VFORM <i>fin</i>
AUX -
INV -
MARKING <i>null</i>
SLASH {}
REL {}
QUE {}

b. STORED

VALUES

?
?
?
?
?
?
?
?

If the learner encounters a feature structure in which it does not yet recognize a feature, ? is retained as a place holder in the local tree, as in (52).

52. INTERMEDIATE STATE

a. UG DEFAULTS

ROOT -
VFORM <i>fin</i>
AUX -
INV -
MARKING <i>null</i>
SLASH {}
REL {}
QUE {}

%□

b. STORED

inf
?

c. INPUT

ROOT -
VFORM <i>inf</i>
AUX -
INV -
MARKING <i>null</i>
SLASH ?
REL {}
QUE {}

As a result, the lexical entry acquired before the learner recognizes SLASH like (53) instead of (49); [SLASH?] represents a UG-specified place-holder).

53. *xcompl* SPEC [VFORM *inf*,SLASH?]

This prevents the SPEC value in the lexical entry of *xcompl* from licensing [SLASH{}]. Once the learner comes to recognize the distinction embodied by SLASH, the learner must replace the ? with an attested value in all learned feature structures. To force this, I assume that the existence of an actual value forces the learner to over-ride the ?. Operationally, this means that in default unification, the learner stores the ? in the learned value if the value for a feature is unattested because it is not known; it stores the marked value if that is attested; otherwise it stores no value in the learned generalization because the default is the correct one. (Recall that [] is the maximally succinct attested feature structure.)

54. LOCAL TREE

[A-] (input = default)
[A+] (input = marked value)
[A-] (no input yet)

DEFAULT STORED

[A-]	[]
[A-]	[A+]
[A-]	[?]

The differentiation between default values that occur in the input and those that do not makes it possible for the learner to factor out default values safely before it recognizes all relevant features. This is important because, as noted in Chapter 1, it would be very inefficient for the learner to have to wait until some relatively late developmental stage to store feature descriptions from which default values have been factored out.

The symbol ? also solves another problem: what happens in a language where a value for some feature is *never* attested. This problem arises as a consequence of the Universal Feature Appropriateness Principle (Ch.2(6)), which prevents the learner from acquiring overly powerful generalizations that do not mention certain features. Any feature that is licensed on any node in any language, must be instantiated on that type of node in every language, even if the feature is not actually used in the language. This leads to the danger that default values for features that represent distinctions that a language does not actually employ will have to be stored even after the end of the language acquisition period. This is not logically impossible, but seems psychologically implausible.

To see how the distinction between attested and unattested features solves this problem, consider the example of honorifics. If languages like Japanese and French have a feature HON to represent the distinction

between honorifics and informal forms, then the Universal Feature Appropriateness Principle requires that a language like English also have the feature HON, even though most speakers of standard English will never encounter [HON+]. Under the idealization of fully attested values, default unification allows the learner to instantiate the default [HON-], which is neutral and therefore harmless. But when incomplete knowledge is taken into consideration, there is a learning problem: the learner does not know if it is learning a language like Japanese, so that it needs a placeholder for HON, or a language like English, where it does not need HON at all.

Technically, the problem can be solved by requiring that ? be maintained as a value for HON through maturity. But this is undesirable because the adult learner would have to store features that its language never uses. A more plausible conjecture is that features that the learner does not recognize and therefore are labelled with ? in all nodes where they are appropriate, are lost after the end of the language acquisition period. So initially, the learner of English will include [HON ?] in its SPEC values.

I conclude from this brief discussion that learning during the string-to-structure process can be represented in a way that is motivated by the actual learning problems. Having shown that this model has the potential to offer an interesting account of grammatical learning at the point

when the learner does not yet recognize all relevant linguistic distinctions, I set this issue aside for the rest of this dissertation.²⁹

²⁹As noted in Chapter 1, the problem of acquiring grammatical generalizations will doubtlessly need to be reworked once the string-to-structure problem has been solved.

CHAPTER 6
FILLER-GAP CONSTRUCTIONS: SYNTAX AND LEARNABILITY

In this chapter, I consider syntactic and learning problems involved in acquiring the properties of interrogative fillers such as *who* and *what*. First I distinguish the universal properties of fillers from those that vary by language. As previously, I seek to identify those principles that are plausibly universal, so that I can address the question of how language-particular selectional properties can be learned. I show that the same mechanisms that make it possible for children to successfully acquire the selectional properties of complementizers also allow them to acquire the selectional properties of fillers.

I first demonstrate that different fillers have quite different selectional properties, even in closely related languages. Therefore these selectional properties must be learned in a way that guarantees that learners of each language will acquire only the properties of fillers that are licensed in their own language, and not the superset of all such properties that may be licensed in any natural language. I argue that the feature SPEC that allows complementizers to mark syntactic properties of their sister clauses can be adapted for fillers, and I call this feature of fillers SELECT. I assume that SELECT is universally specified as appropriate for interrogative lexical items (which are [QUE{[1]}], and that the value of SELECT is acquired in the same way that the value of SPEC is. Fully

specified lexical nodes are passed from the syntax into the lexicon where the operation of default unification allows the learner to store a compact generalization containing only unpredictable, marked feature values. Positively learned SELECT values allow the elimination of negative constraints such as the Clausal REL Prohibition and the Singleton REL Constraint (which will be discussed below) that otherwise would render HPSG unlearnable.

I then consider the implications of the fact that interrogative fillers can be phrasal, and I present a model in which phrasal selectional properties are stored lexically. In Chapter 7, I consider further challenges to the theory that all learning is lexical, and I examine some cases where phrasal selection is not associated with an overt lexical item.

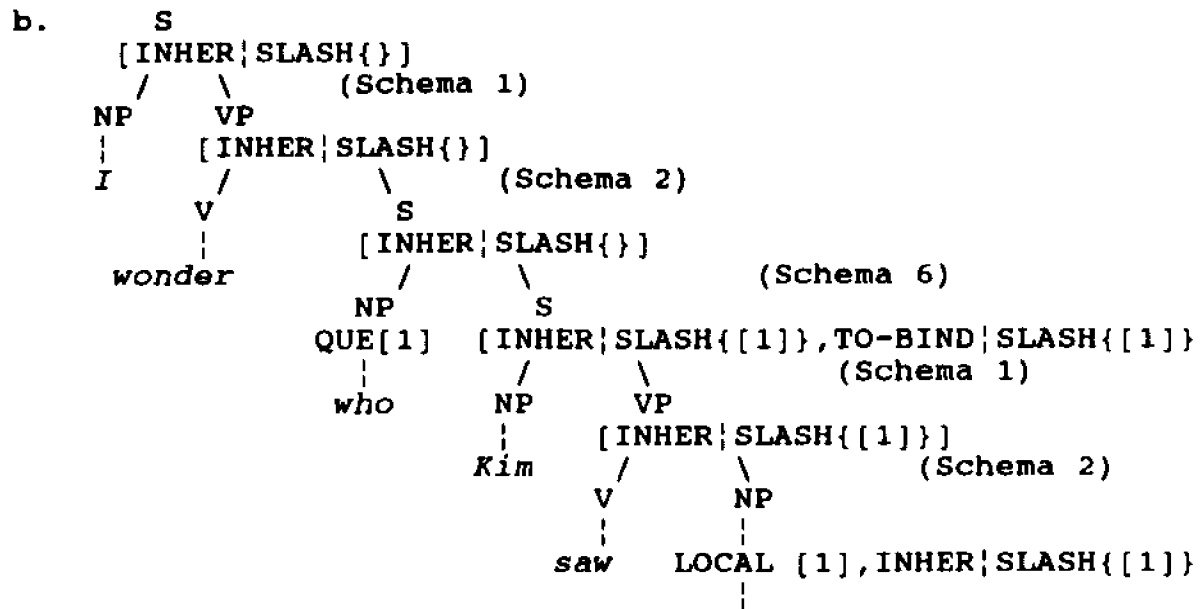
6.1 TOWARD A UNIVERSAL VERSION OF SCHEMA 6

On the assumption that schemas are universal and innate, and that particular grammars select from among the schemas and further specify them (as discussed in Chapter 2) the question for learnability is how children are able to acquire the language-particular 'further specifications' of the universal schemas. To do this for Schema 6, the schema that establishes relations between fillers and gaps, the learner must know which properties of the filler, and of its sister and mother, are universal, so that these properties can be factored out of input trees. Any further specifica-

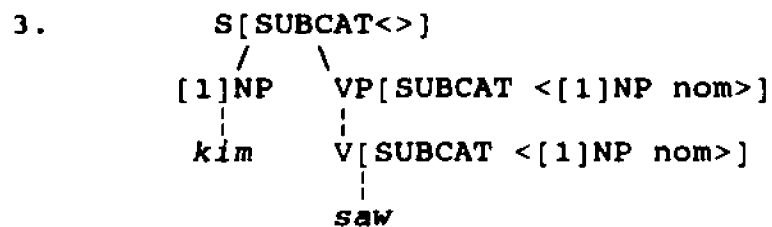
binding off is imposed by Schema 6, which imposes an obligatorily non-empty value on [TO-BIND|SLASH]

2. (cf. Ch. 2 (26))

a. I wonder S[who Kim saw ___]?¹



This is exactly comparable to Schema 1, where one of the subcategorization requirements of the head is unsaturated, and the head specifies a complement to saturate itself.



Note that subcategorization of the VP for its subject is actually stored in the lexical entry for the verb, but is structure-shared with the VP, at which level it can be satisfied.

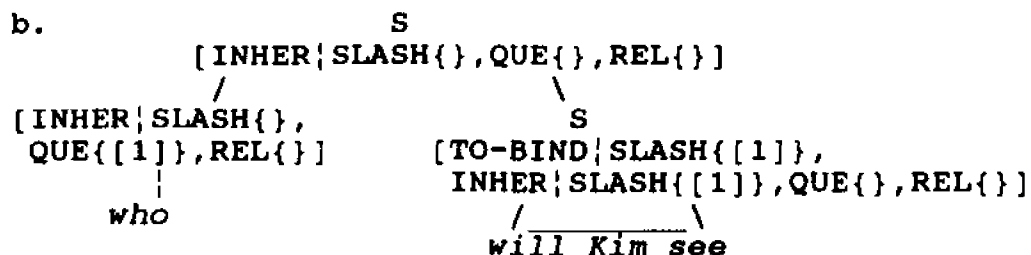
¹In some dialects, *whom* is the accusative form.

The NONLOCAL Feature Principle guarantees that the SLASH values will be properly bound off, so that they are not carried indefinitely up the tree.

4. The NONLOCAL FEATURE PRINCIPLE (=Ch.2(29), P&S-94,4(16))
 For each nonlocal feature, the INHERITED value on the mother is the union of the INHERITED values on the daughters minus the TO-BIND value on the head daughter.

Thus, if the head daughter has only a single gap, the NONLOCAL Feature Principle requires that the value of INHER|SLASH on the mother node be {}. (5) satisfies both Schema 6 and the NONLOCAL Feature Principle.

5. a. Who will Kim see ___?



This machinery handles many examples in English and other languages.

However, the formulation of Schema 6 in (1) is simultaneously too specific and too powerful to be universal. The specification S[fin] on the head daughter is too narrow, even for English. In (6), the VP which is the head daughter of Schema 6 is neither an S nor finite.

6. I wonder what VP[to eat ___].

At the same time, Schema 6 is too powerful. In languages which license sequences of interrogatives, Schema 6 will apply recursively. But the recursive application of

Schema 6, if unconstrained, would license a series of fillers and corresponding gaps in all languages, including English, as in the ungrammatical (7).

7. *What to who where did Kim give _ _ _?

However, sentences like (7) are grammatical in languages like Bulgarian and Polish, as illustrated in (8), and this justifies the need for the set value syntactically.

8. Bulgarian (Rudin (1988(29a)))²
 Koj kakvo na kogo e dal?
 who what to whom has given
 'Who gave what to whom?'

That the difference between (7) and (8) is syntactic, not semantic, is shown by the fact that the English translation, with all *wh*-phrases in situ except one, is well-formed. The grammar of English must not license sequences of interrogative fillers, while the grammar of Bulgarian must specify that its language does license them.

P&S-94 do not propose a constraint to restrict the repeated application of Schema 6 in English interrogatives, but they do discuss the contrast between Marathi, which

²Rudin (1988:447) distinguishes between Bulgarian and Roumanian, where all *wh*-phrases move into [SPEC,CP], and Polish, Czech and Serbo-Croatian, in which only one *wh*-phrase moves into [SPEC,CP]; the others adjoin to IP. The evidence for the different structures comes from distinctions as to whether one or more than one *wh*-words may be fronted, whether elements such as clitics can intervene between *wh*-words, and whether the word order of the *wh*-words is fixed or free. The question of how these distinctions are acquired is a very interesting problem which I cannot address here. A hypothesis which seems plausible is that the set value of a NONLOCAL feature is licensed only in contexts where the singleton value of that feature is also licensed, as will be mentioned briefly below (p.42).

permits a relative clause containing multiple relative phrases (in relative-correlative constructions), and English, which does not permit multiple relative phrases in a single clause.³

9. Marathi (P&S-94,5(51), attributed to Andrews 1975 and Dalrymple and Joshi 1986)
 [jaa mulaa-nii jaa mulii-laa paahile] [tyaa-nii
 which boy-ERG which girl-ACC saw he-ERG
 ti-laa maagNii ghaatlii]
 her-DAT proposed
 'The boy who saw the girl_i proposed to her_i.'
 ('Which boy_j saw which girl_i, he_j proposed to her_i.')

To account for the fact that English does not license sequences of relatives, P&S-94 propose the language-particular Singleton REL Constraint.

10. Singleton REL Constraint (English) (P&S-87,Ch.5(10))
 The cardinality of the value of INHER|REL is at most 1.

Given the ungrammaticality of sentences like (7), a natural move would be to expand the Singleton REL Constraint, so that the cardinality of the value of SLASH and QUE is limited to at most one.⁴ Thus (10) may be reformulated as the Singleton NONLOCAL Feature Constraint.

³I confine virtually all of my examples to interrogative fillers, except when it is helpful to distinguish language-particular properties from universal properties by reference to non-interrogative fillers. I assume that relativizers select properties of their sister nodes in essentially the same way that interrogatives do; details will be discussed in Ch.7.

⁴It is not clear whether the constraint applies within a language to both interrogatives and relatives, or to neither, or whether interrogatives and relatives may differ in this respect. I assume that they may differ.

However, the very fact that P&S-94 assume that English needs a constraint that restricts the value of a feature to the cardinality of one implies that they are taking the set value to be the norm, a kind of default. Unfortunately, the Singleton NONLOCAL Feature Constraint for English is learnable only from negative evidence. Such constraints threaten to render HPSG grammars unlearnable, and need to be eliminated. Even if the Singleton Constraint is descriptively accurate for English, it affords no account of how the grammar of English could be acquired, and so is not explanatorily adequate (Chomsky 1965). The solution is to treat English as the unmarked case, and assume that Marathi learners have to learn multiple relativization from positive evidence, as discussed below in Section 6.3.2.

Once the language-particular aspects of Schema 6 are removed, what remains as presumably universal is that Schema 6 captures the fact that fillers universally require a non-empty SLASH value on their sister clause. (11) amends (1) by removing *S* and *fin*, and it permits the head daughter to be any phrasal projection of *V*.

11. Schema 6 (Intermediate version; see Ch.7 for final revision)
- ```

X --> LOC[1] phrase[LOC|CAT|HEADverb,
 NONLOC|TO-BIND|SLASH{[1]},
 NONLOC|INHER|SLASH{[1]...}]
 (filler) (head)

```

In the next section, I argue that lexical fillers are like markers in that they specify certain properties of their sister nodes. I then argue that this information is

stored in a lexical feature SELECT which is much like the feature SPEC, which is appropriate for markers; in Chapter 7, I will also show that SELECT is much like the feature MOD, which P&S-94 argue is appropriate for adjuncts.<sup>5</sup>

Learners acquire the values of SELECT in the same way that they acquire the value of SPEC for complementizers, so that only attested marked values need to be learned and stored.

## 6.2 SELECTIONAL PROPERTIES OF INTERROGATIVE FILLERS: THE DATA

When an interrogative filler is used, its sister node must have certain syntactic properties. In this section, I give examples of language-particular variation and item-particular variation in these selectional restrictions. If any item in any language exhibits a certain distinction, learners are evidently capable of acquiring that distinction.

One well-known property of fillers in English is that interrogatives require subject-auxiliary inversion in root clauses but *wh*-exclamatives and topicalization require standard subject-verb order.

- |     |    |                                |                       |
|-----|----|--------------------------------|-----------------------|
| 12. | a. | How tall is that building __?  | Question [INV+]       |
|     | b. | *How tall that building is __? | Question [INV-]       |
| 13. | a. | How tall that building is __!  | Exclamative [INV-]    |
|     | b. | *How tall is that building __! | Exclamative [INV+]    |
| 14. | a. | Susan, I can talk to __.       | Topicalization [INV-] |
|     | b. | *Susan, can I talk to __.      | Topicalization [INV+] |

---

<sup>5</sup>In Ch.7, I will compare SPEC, SELECT, MOD, and SUBCAT.

In the absence of a filler, inversion is optional in English questions. (Marked prosodic contours are needed to establish illocutionary force in (15.b)).

15. a. Is that building six stories tall? Question [INV+]  
 b. That building is six stories tall? Question [INV-]

For exclamatives without a filler, inversion is also optional (though marked prosodic contours are also needed to establish illocutionary force in (16.a) and (16.b)).

16. a. Is that building tall! Exclamative [INV+]  
 b. That building is tall! Exclamative [INV-]

So whether or not there is subject-auxiliary inversion depends only in part upon sentence modality: for interrogatives and exclamatives, both modality and the presence or absence of the filler are relevant for establishing syntactic form.

Other properties of the sister clause also vary. Although inversion is obligatory in root clauses with (most) interrogative fillers, non-inversion is obligatory in interrogative embedded clauses, at least in standard varieties of English.

17. Standard English: Obligatory inversion in [ROOT+]  
 a. What can Kim take? [AUX+, INV+]  
 b. \*What Kim can take? [AUX+, INV-]
18. Standard English: obligatory non-inversion in [ROOT-]  
 a. I wonder S[what Kim can take \_\_\_]. [QUE, INV-, AUX+]  
 b. \*I wonder S[what can Kim take \_\_\_].<sup>6</sup> [QUE, INV+, AUX+]  
 c. I wonder S[what Kim took \_\_\_]. [QUE, INV-, AUX-]

---

<sup>6</sup>This example contrasts with (i), where the higher S is like a parenthetical.

i. I wonder, what can Kim do \_\_\_?

However, inversion is optional in *wh*-questions in Belfast English.

19. Belfast English (Henry 1992(42)) (=Ch.4(10))  
I wonder what street does he live on.
20. Black English (William Stewart (p.c.))  
I don't know would he like it.

Bulgarian and Roumanian typically show SVO word order in declarative clauses (although word order is free to a certain extent, so other orders are possible). But in *wh*-questions, inversion is obligatory in both root and embedded clauses; specifically, an NP may not intervene between the *wh*-expression and the inflected verb (Tina Kraskow, p.c.). Note that inversion in Bulgarian and Roumanian is broader in scope than in English, where only auxiliaries undergo inversion.

21. Bulgarian (Tina Kraskow p.c.)
  - a. Ne znam kakvo kupuva Ivan  
not I-know what buys John  
'I don't know what John is buying.'
  - b. \*Ne znam kakvo Ivan kupuva  
not I-know what John buys

Thus, learners must learn (at least) whether or not any particular filler type selects an inverted clause, whether it does so only in root clauses or also in embedded clauses, whether inversion is obligatory and whether the clause may contain another filler. But these generalizations cannot be made about all interrogative pronouns as a class or about all relative pronouns. Selectional properties differ even among particular interrogative pronouns. In what follows, I

consider interrogatives in the most depth. Particular interrogatives occur obligatorily with sisters that have different properties than their general class. Most interrogatives in English selects VP[VFORM inf] but *why* and *how come* do not.

22. a. I wonder *who/what* to avoid.  
 b. I wonder *how/when* VP[to fix it].VP[VFORM inf,AUX+]  
 c. \*I wonder *why/how come* VP[to fix]VP[VFORM inf,AUX+]

However, the other selectional properties of *why* and *how come* are not identical. The idiomatic interrogative *how come* is ungrammatical if its sister clause has subject-auxiliary inversion, while its near synonym *why* is like other interrogatives in that subject-auxiliary inversion is obligatory for finite clauses.<sup>7</sup>

23. a. \**How come* are you going \_\_\_? [QUE, INV+]  
 b. *How come* you are going \_\_\_. [QUE, INV-]
24. a. *Why* are you going \_\_\_? [QUE, INV+]  
 b. \**Why* you are going \_\_\_? [QUE, INV-]

In addition, *why* licenses VP[base] but *how come* (like most English interrogatives) does not.

25. a. *Why* go home? [QUE, BASE]  
 b. \**How come* go home? [QUE, BASE]  
 c. \**When* go home? [QUE, BASE]

---

<sup>7</sup>I treat *how come* as a single lexical item, spelled in two words for historical reasons. But even if *how come* is phrasal, it is exceptional, as (i.a) shows, so the question of how learners acquire its idiosyncratic selectional properties remains.

i. a. *For what reason* are you going? [INV+]  
 b. \**For what reason* you are going? [INV-]

The selectional properties of phrasal fillers will be discussed in Section 6.5.





assumptions, it would be impossible for learners to acquire these selectional restrictions which limit the language over and above the restrictions imposed by UG. In this important sense, the theory of P&S-94, which requires language-particular selectional constraints, is not yet explanatorily adequate.

### 6.3 LEXICAL FILLERS

In this section, I argue that the required relationship between lexical fillers and their sisters can be imposed by means of a feature which I call SELECT, much like the feature SPEC which is appropriate for MARKERS. Since the clause also selects properties of the filler, through the value of SLASH, the filler daughter of Schema 6 and the head daughter of Schema 6 impose requirements on each other in a kind of mutual co-selection.

The evidence considered above suggests that the features to which lexical fillers are sensitive are the same as those to which markers are sensitive. The examples above showed that fillers are sensitive to the value of CAT, VFORM, AUX, INV, ROOT, MARKING, SLASH, and QUE. These are all features whose values complementizers are also sensitive to, as documented in Chapter 5, examples (19)-(38).

The similarity of the selectional properties of fillers and of complementizers suggests that interrogative fillers select their sister nodes in the same way that complementizers do. (The alternative, which is that clauses

select fillers --e.g., uninverted clauses select *how come*, but inverted clauses select *why* -- seems less plausible.) I therefore posit a feature SELECT which is appropriate for interrogative fillers. Although this feature is not part of standard HPSG, as presented in P&S-94, there is no reason that such a feature should not be added. In Chapter 7, I will consider whether SELECT is distinct from P&S-94's SPEC feature.<sup>11</sup>

---

<sup>11</sup>Interrogatives are also like complementizers in that some information about whether or not the selector is an interrogative must be available to verbs like *wonder* which are obligatorily subcategorized for interrogatives, and to verbs like *think* which are obligatorily subcategorized for non-interrogatives.

i. a. I think that/\*whether/\*where you went.

b. I wonder \*that/whether/where you went.

Schema 4 accomplishes this for markers by requiring that the marking value on the marker daughter be structure-shared with the marking value on the mother.

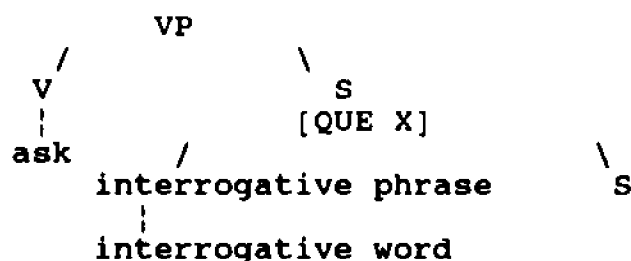
For interrogatives, wh-percolation does part of the work.

ii. a. I know to whom he complained.

b. I asked how kind he was.

But this only insures that the wh-status of the interrogative word is accessible to the node by which it is dominated. It is also necessary for this information to travel further up, to the clausal node that dominates the whole filler-gap construction.

iii.



The information that the filler is an interrogative (i.e., that it has a non-empty value for QUE) must be instantiated on the higher S in (iii) so that the verb's subcategorization information can be satisfied within the local tree.

Whether only information about the presence of interrogatives and/or complementizers needs to travel up, or whether the full phonological value is needed is a question that I must set aside here.

To control the distribution of SELECT, I assume the existence of a universal principle called the SELECT Principle, which is modelled upon the SPEC Principle of P&S-94.<sup>12</sup>

31. SPEC Principle (P&S-94:1(50)); (=Ch.3(14.a))  
     (for markers and determiners)  
 If a non-head daughter in a headed structure bears a SPEC value, it is token-identical to the SYNSEM value of the head daughter.
32. SELECT Principle  
 If a nonhead daughter in a headed structure bears a SELECT value, it is token-identical to the SYNSEM value of the head daughter.

Since the only configuration in which fillers may combine with heads is Schema 6, SELECT values for interrogatives amount to language-particular further specifications of Schema 6 that are stored in lexical entries of fillers, just as SPEC values are language-particular further specifications of Schema 4 that are stored in the lexical entries of complementizers.

Since SELECT is appropriate for the filler daughter of Schema 6 and a non-empty value of SLASH is appropriate for the head daughter of Schema 6, clauses and fillers impose requirements on each other, in a kind of co-selection. The value of SELECT and the value of SLASH thus license particular combinations of fillers and head-daughters.

---

<sup>12</sup>SELECT is not a head feature. Recall from Ch.3, fn.10, that P&S-94 treat SPEC as a head feature because it is appropriate for determiners, which are the heads of determiner phrases.

Since markers are sensitive to value of features within the SYNSEM value of their sister node, I assume that fillers are too. I conjecture that fillers are like markers in a further respect also, i.e., a filler may not be sensitive to any feature values not associated with the SYNSEM value of the head daughter of Schema 6. Assuming that this constraint is universal, a filler, like a marker, cannot select features of any complements within its sister S, including its subject. The head features of such complements never travel up to the head node of Schema 6.

To sum up, markers and interrogative and relative fillers share the property of being non-heads that act like heads in the one respect that they are sensitive to certain features of their sisters. Markers and fillers thus share with heads the ability to impose horizontal selection properties via lexical features, which can be specified in lexical entries. This contrasts with complement (or argument) phrases which impose no selectional requirements at all on the head by which they are subcategorized. However, because Schema 6 has a non-empty value for SLASH on the head daughter, which restricts the properties of the filler daughter, the filler and head-daughter participate in a kind of co-selection. This is unlike Schema 4, where the complementizer specifies syntactic properties of the head daughter, but it appears that the head daughter does not specify properties of the complementizer.

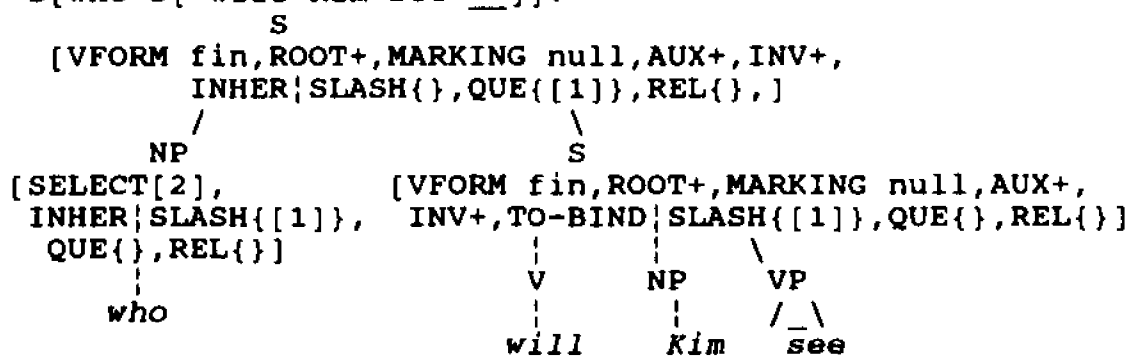
## 6.3.1 ACQUIRING THE VALUE OF SELECT

From this point on, I treat SELECT as a universal syntactic feature, appropriate for interrogative fillers.<sup>13</sup> The SELECT value can be learned from a local tree in the same way that the SPEC value of a complementizer can. Consider (33.a), with the feature structure in 33.b.<sup>14</sup>

33.

a. S[who S[ will Kim see \_\_\_]]?

b.



c. *who* SELECT [CAT S, VFORM fin, ROOT+, AUX+, INV+,  
MARKING null, TO-BIND|SLASH{[1]},  
INHER|SLASH{[1]}, QUE{ }, REL{ }]

The SELECT value in (33.c) constitutes a description of one of the sets of syntactic features that *who* can co-occur with: *who* selects S[fin] but not VP[fin], [AUX+] but not [AUX-], [INV+] but not [INV-], and [MARKING null] but not [MARKING lex]. (33.c) licenses [INV+] only if it co-occurs with [ROOT+]. Because of [INV+], an auxiliary is required,

<sup>13</sup>I assume that lexical exclamatives are also a kind of interrogative filler, as in (i).

i. *How she cried!*

<sup>14</sup>Recall that Schema 6 stipulates that the head daughter in universally NONLOCAL|TO-BIND|SLASH{X}.

necessitating *do*-support if no other auxiliary is appropriate for semantic or illocutionary reasons.<sup>15</sup>

The fully specified feature structure of *who* is transferred from the local tree to the lexicon, just as was the lexical node of complementizers. Given (33.b), (34) will enter the lexicon.

---

<sup>15</sup>On this analysis, *do*-support in subject-auxiliary inversion is required because the learner has only encountered for [ROOT+] a SELECT value which licenses [AUX+, INV+]; *do*, which is semantically empty, is used whenever an auxiliary with semantic content is not appropriate.

The fact that unstressed *do* is normally ungrammatical in uninverted position is another example of the need for conservative learning of the behavior of individual lexical items; in this case phonological properties interact with syntactic properties. The learner of English will never encounter the ungrammatical sentences in (i).

- i. a. \*I wonder what Kim did feed the pandas?
- b. \*I think that Kim did feed the pandas bamboo.

However, the learner will encounter uninverted *do* when it is stressed.

- ii. a. I wonder what Kim DID feed the pandas?
- b. I think that Kim DID feed the pandas bamboo.

So the learner acquires the information that *do* is [AUX+, INV+, STRESS+] from positive evidence. (I assume that the default value for STRESS is [-].) I will not consider in further detail the problem of how phonology interacts with syntax, but it is clear that the Universal Feature Appropriateness Principle (UFAP; see Ch.2) must include such properties as stress, to force learners to include them in their characterization of the input which feeds grammatical acquisition,

34. Lexical entry for the filler *who*

|            |        |       |                                                                                                                  |  |  |  |  |
|------------|--------|-------|------------------------------------------------------------------------------------------------------------------|--|--|--|--|
| PHON <who> |        |       |                                                                                                                  |  |  |  |  |
| SYNSEM     | LOC    | CAT   | HEAD noun[acc],<br>SELECT [S,VFORM fin,<br>ROOT+,AUX+,<br>INV+,MARKING null,<br>INHER SLASH [1],<br>QUE{},REL{}] |  |  |  |  |
|            | NONLOC | INHER | QUE {[2]}<br>REL {}<br>SLASH {}                                                                                  |  |  |  |  |
|            |        |       | SUBCAT <>                                                                                                        |  |  |  |  |

The default values of SELECT must then be factored out in the lexicon so that only marked, predictable values need be stored. To see what the outcome will be, the markedness specifications for the value of SELECT must be determined.

## 6.3.2 MARKEDNESS SPECIFICATIONS

In this section, I propose markedness values for the value of SELECT for interrogative fillers, and I show that reasonable assumptions about markedness, consistent with earlier observations about markedness specifications for complementizers, allow SELECT values to be stored compactly. As previously, I emphasize that the postulation of innate syntactic default values does not preclude the possibility of finding deeper explanations for these defaults. Also as before, I propose markedness specifications here based on skimpy evidence for the sake of making a concrete proposal. Where the distributional evidence for defaults is not clear on other grounds, I assume that the context-free default for

asymmetrical features is always the neutral value that represents the absence of some complicating syntactic property.

I begin by repeating the defaults proposed in Ch. 5.

(Here I use INHER|SLASH rather than SLASH.)

- |                                                                                         |                                  |
|-----------------------------------------------------------------------------------------|----------------------------------|
| 35. a. (cf. Ch. 5(16.a))                                                                | b. (=Ch. 5(16.b))                |
| <b>CONTEXT-FREE DEFAULTS</b>                                                            | <b>CONTEXT-SENSITIVE DEFAULT</b> |
| <pre>CAT S VFORM fin ROOT + AUX - INV - MARKING null INHER SLASH {} QUE {} REL {}</pre> | [MARKING compl [SPEC ROOT-]]     |

Since interrogative fillers license gaps, it might be expected that they specify context-sensitive defaults for NONLOCAL features, and, in fact, the value of INHER|SLASH can never be the empty set on the sister of interrogative filler: in some languages its value can be {[1]} and in others it can be {[1], ..., [n]}.<sup>16</sup> The fact that sentences like (36) are ungrammatical in even one natural language means that the default value for INHER|SLASH associated with the sister of fillers cannot be the unconstrained set value because of the Subset Principle (see discussion above).

36. (= (7))  
\*What to who where did Kim give \_ \_ \_?

---

<sup>16</sup>Recall that Schema 6 specifies that the value of [TO-BIND|SLASH] is obligatorily the singleton value {[1]}, so defaults are not relevant to it. A default can apply only where UG does not absolutely constrain a feature value.

The formulation of Schema 6 with the unconstrained value incorrectly licenses sentences like (36), and the value with cardinality one is not learnable from positive evidence. So although the elsewhere default for INHER|SLASH is the empty set, the default value imposed by an interrogative filler on the INHER|SLASH of its sister node (the head daughter of Schema 6) must be the singleton value. This is a classic context-sensitive default.<sup>17</sup> On this analysis, the ungrammaticality of sentences like (36) in English represents the default case, and requires no special learning. If set values of INHER|SLASH are licensed in a particular context, as in (8), the learner can, and must, acquire the unconstrained set value {[1],...} from positive evidence.

I now briefly consider two other NONLOCAL features, REL and QUE. In English, interrogative pronouns are usually ungrammatical in situ, at least in informational (non-echo) questions.

37. a. *Who did you see \_\_\_?*  
 b. *\*You saw who?*

---

<sup>17</sup>Note that it would be more economical if Schema 6 included the specification that its head daughter is by default [INHER|SLASH({1})]; then it would not be necessary to specify independently that interrogative fillers, relative fillers, and topicalized constituents each select the singleton value of INHER|SLASH by default. However, this type of 'schema-particular' default is syntactic in that it is not associated with a particular lexical item or stored in the lexicon, and is therefore, by hypothesis, not allowable.

However, an absolute prohibition against interrogative pronouns in situ is too strong: interrogatives may appear in situ in multiple questions.<sup>18</sup>

38. a. Who saw *who*? (cf.(37.b))  
 b. Who did you see *where*?

Relatives pronouns are always ungrammatical in situ in English.

39. P&S-94(Ch.5(18a,c))  
 a. \*Here's the student [Kim likes *who*].  
 b. \*Here's the student [Dana met *whose* sister].

However, some language do permit relative pronouns in situ. (P&S-94:237 cite Marathi, Navajo, Lakhota and Bambara.) P&S-94 therefore propose, for English, the Clausal REL Prohibition, which prevents a relativizer from being in situ.

40. Clausal REL Prohibition (P&S-94,Ch.5(31))  
 The INHER|REL value of S must be empty.

As above, the necessity for a language-particular prohibition implies that the default for INHER|REL is the non-empty (singleton or set) value. But the Clausal REL Prohibition is unlearnable in the absence of negative evidence, just as the singleton value of SLASH in Schema 6 and the Singleton REL Constraint shown in (10) are.

However, if the context-free defaults [QUE{}] and [REL{}]

---

<sup>18</sup>The learner of English must also learn that in situ interrogatives in informational questions are licensed only by a fronted interrogative filler. This suggests that there must be two kinds of interrogative words with very different distribution: filler-interrogatives and local-interrogatives. (Fodor 1992a)

apply in Schema 6, as elsewhere, the need for the unlearnable Clausal REL Prohibition and the unlearnable Singleton REL Constraint is eliminated. Since in Marathi (P&S-94:228) a relative clause may contain "multiple relative words, each of which must be associated with a correlative pronoun in the correlative clause", the Marathi learner must learn that its language licenses both in situ relative pronouns and sequences of relative phrases. On this analysis, the properties of Marathi relative clauses require more learning than the properties of English relative clauses do.

An interesting side-effect of the context-free empty set defaults for NONLOCAL features is that they serve as an anti-recursion mechanism. Recall that one of the problems with the formulation of Schema 4 in P&S-94 was that it licensed repeated complementizers.

41. (=Ch.3(32))

\*I think that that Kim likes pandas.

This problem was solved for complementizers by a context-free default that imposes an empty value for MARKING on the sister node of the marker. Markedness specifications serve the same function for fillers. The default values for REL and QUE on the filler daughter of Schema 6, as everywhere else, will be the empty set. If the learner encounters evidence that a non-empty value is licensed for one of these features, the learner must change its grammar so that it licenses a cardinality one. Only after encountering clear

positive evidence can the learner change its grammar to license set values.

#### 42. MARKEDNESS RANKING FOR NONLOCAL FEATURES<sup>19</sup>

| Default       | <-- | Marked           | <-- | Most marked         |
|---------------|-----|------------------|-----|---------------------|
| INHER SLASH{} |     | INHER SLASH{[1]} |     | INHER SLASH{[1]...} |
| QUE{}         |     | QUE{[1]}         |     | QUE{[1]...}         |

However, interrogative fillers select a singleton value of INHER|SLASH by default. The learner may only replace the singleton value with the unconstrained set value; the learner will never replace the singleton value of a INHER|SLASH with the empty value. But since Bulgarian, for example, allows a marked value for SLASH, learners of Bulgarian will have to acquire the marked value that overrides the default, so that in this particular respect, the grammar of Bulgarian is more complex than English and therefore requires extra learning.

If the markedness specifications in (42) are correct, the values of NONLOCAL features show the tidemark effect of Fodor (1992a), where the grammaticality of a marked value in some context implies the grammaticality of an unmarked value in that context: if some language licenses a set value of some NONLOCAL feature in some context, it also licenses the singleton value, but the grammaticality of the singleton value does not entail the grammaticality of the set value. In this case, the tidemark effect plus the requirement

---

<sup>19</sup>The markedness ranking in (42) also applies to relative pronouns, which will be discussed briefly in the next chapter.

|    |       |     |          |     |             |
|----|-------|-----|----------|-----|-------------|
| i. | REL{} | --> | REL{[1]} | --> | REL{[1]...} |
|----|-------|-----|----------|-----|-------------|

imposed by the Subset Principle that the learner assume the narrower value rather than the broader one, work together to ensure that the learner acquires the grammar of the target language.

I now consider whether LOCAL features also have context sensitive default values in the context of interrogative words. Since interrogatives occur freely in both root and embedded contexts, interrogative fillers do not appear to be characterized by a special association with [ROOT+] or [ROOT-]. So for interrogative fillers, I assume simply that the context-free default [ROOT+] is applicable.<sup>20</sup>

What about defaults for CAT and VFORM? As noted above (p.254), only a subset of English fillers that select S[VFORM fin] select VP[VFORM inf]. Furthermore, German fillers select only S[VFORM fin]. Sentences like (44) are ungrammatical in German (and Danish).

43. Ich weiss was Hans zu tun versucht.  
I know what John is trying to do.

44. Northern German McDaniel (1989(19))<sup>21</sup>  
a. \*Ich weiss nicht was zu machen.  
I don't know what to do.

---

<sup>20</sup>The context-free default [ROOT+] is clearly not appropriate for relative pronouns, which typically select [ROOT-]. I assume that selection of [ROOT-] by relative pronouns is absolutely universal.

<sup>21</sup>The equivalent sentences without zu are grammatical (Inge Lasser p.c.), although they are in an unusual register, and an adverb may not precede the infinitive.

i) Ich weiss nicht was machen.  
ii) Ich weiss nicht wen anrufen.

- b. \*Ich weiss nicht wen anzurufen.  
I know not who to-call  
'I don't know who to call.'
45. Northern German McDaniel (1989(15))
- a. \*Hans versucht wen zu bestechen.  
Who is Hans trying to bribe?
- b. \*Wen zu bestechen ist die Frage.  
Who to bribe \_\_ is the question.

This evidence suggests that the unmarked value for the sister of an interrogative filler is [VFORM *fin*], which is the same as the context-free default value.<sup>22</sup> The same evidence that suggests that *fin* is the default for VFORM also suggests that S is the default for CAT, which is the context-free default for CAT.

What about the defaults for INV and AUX? The fact that English interrogative fillers typically require auxiliary inversion in matrix clauses suggests, at least weakly, that [INV+] might be an appropriate defaults for interrogative fillers. But since inversion is ungrammatical in embedded clauses, I continue to assume that the context-free default for INV is [-]. Since auxiliaries are a marked kind of verb, in that their subject receives its semantic role from the head of the VP for which they are subcategorized, I also assume that the default for AUX is [-].<sup>23</sup>

---

<sup>22</sup>This limited evidence suggests that with interrogative fillers, the tidemark model of Fodor (1992a) discussed in fn.2 applies to VFORM: the grammaticality of *inf* implies the grammaticality of *fin*.

<sup>23</sup>As noted in Ch.2, fn.6, I adopt AUX and INV because they are used by P&S-94. They are responsible for some of the complexity of the generalizations that I show that a learner of

In summary, the only context-sensitive lexical default that I have proposed for interrogative fillers is represented in (46), where [1] is structure-shared with the value of QUE on the filler, which obligatorily is non-empty.

46. [QUE{[1]}, SELECT [INHER|SLASH {[1]}]]

In other words, (46) specifies a lexical interrogative is associated by default with a singleton value of INHER|SLASH, rather than a set value. The context-sensitive default, which is more specific, necessarily takes precedence over the context-free default for INHER|SLASH specified in (47.a).

47. DEFAULTS FOR SELECT VALUE OF INTERROGATIVE FILLERS

a. (= (35.a))

**CONTEXT-FREE  
DEFAULTS**

|                |
|----------------|
| CAT S          |
| VFORM fin      |
| ROOT -         |
| INV -          |
| AUX -          |
| MARKING null   |
| INHER SLASH {} |
| QUE {}         |
| REL {}         |

%□

b. (= (35.b))

**CONTEXT-SENSITIVE  
DEFAULTS**

|                   |
|-------------------|
| INHER SLASH {[1]} |
|-------------------|

=

c.

**APPLICABLE  
DEFAULTS**

|                   |
|-------------------|
| CAT S             |
| VFORM fin         |
| ROOT -            |
| INV -             |
| AUX -             |
| MARKING null      |
| INHER SLASH {[1]} |
| QUE {}            |
| REL {}            |

Since interrogative fillers can only be (non-head) daughters of Schema 6, the defaults in (47.c) apply in a configuration licensed by Schema 6. However, none of these defaults are absolute constraints on Schema 6; default values by defi-

---

English might acquire; it is possible that a different analysis of the properties of auxiliaries and inversion might simplify the learning process.

dition are over-ridden in particular lexical entries.<sup>24</sup>

(48.b) is part of the unreduced lexical entry for *who*. The defaults in (47.c) can be factored out of the SELECT value in (48.b) by default unification.

48. a. *Who* will Kim see \_\_\_?  
 b. *who* SELECT [CAT S, VFORM *fin*, ROOT+, AUX+, INV-, MARKING null, INHER|SLASH{[1]}, QUE{}, REL{}]

Like the generalizations about the SPEC value of the complementizer that in Ch. 5, (49.b) is a generalization in that it licenses a particular type of clause as the sister of the interrogative filler *who*; however, it is redundant in that it includes many default values.

49.

a. (=47.c)  
**DEFAULTS**

b.  
**STORED  
 GENERALIZATION**

c. (=48.b))  
**INPUT STRUCTURE**

|                                                                                                                |    |              |   |                                                                                                                |
|----------------------------------------------------------------------------------------------------------------|----|--------------|---|----------------------------------------------------------------------------------------------------------------|
| CAT S<br>VFORM <i>fin</i><br>ROOT +<br>MARKING null<br>AUX -<br>INV -<br>INHER SLASH {[1]}<br>REL {}<br>QUE {} | %□ | AUX+<br>INV+ | = | CAT S<br>VFORM <i>fin</i><br>ROOT +<br>MARKING null<br>AUX +<br>INV +<br>INHER SLASH {[1]}<br>REL {}<br>QUE {} |
|----------------------------------------------------------------------------------------------------------------|----|--------------|---|----------------------------------------------------------------------------------------------------------------|

---

<sup>24</sup>Note that there are relatively few context-sensitive defaults for interrogative fillers: most of the applicable defaults turn out to be just the general context-free defaults. This was also the case for complementizers, as seen in Ch.5(29)). Recall that I noted in Section 5.4 that lexical categories are a safe context for context sensitive defaults and adopted the conservative hypothesis that only lexical categories may be the context for context-sensitive defaults. But should it turn out that hypothesis is too strong, it remains more likely that a set of non-conflicting context-sensitive defaults can be discovered if there are relatively few context-sensitive defaults relative to the number of context-free defaults.

Default unification allows generalization by feature omission. (49.b), which is all that the learner needs to store, is a compact and linguistically sensible generalization which licenses exactly those SELECT values in (49.c). Every feature whose value is not specified in (49.b) is assigned its default value by default unification. Thus, for example, [ROOT+] is instantiated. The acquired generalization includes only item-specific values that must be stored in the SELECT values of the particular interrogative word. The SELECT value for particular interrogatives can thus be acquired in exactly the same way that the SPEC value for markers can.

### 6.3.3 CORRECT GENERALIZATION OVER SELECT VALUES

In addition to the problem of factoring predictable universal features out of lexical entries, the problem of collapsing lexical entries, which arises with complementizers, also occurs with interrogative fillers. Even a single interrogative filler may have several SELECT values. And two or more fillers may have identical SELECT values. If these are not somehow combined and collapsed together, the grammar would be redundant and cumbersome. Consider just the lexical item *who*, which has (at least) the SELECT values in (50)-(53).

50. a. *Who* will Kim like?  
 b. SELECT [CAT S,VFORM fin,ROOT+,AUX+,INV+, MARKING null,INHER|SLASH{[1]},REL{},QUE{}]

51. a. I wonder *who* Kim will like.  
 b. SELECT [CAT S,VFORM fin,ROOT-,AUX+,INV-, MARKING null,INHER|SLASH{[1]},REL{},QUE{}]
52. a. I wonder *who* Kim likes.  
 b. SELECT [CAT S,VFORM fin,ROOT-,AUX-,INV-,MARKING null,REL{},QUE{},INHER|SLASH{[1]}]
53. a. Kim asked *who* to choose for the team.<sup>25</sup>  
 b. SELECT [CAT VP,VFORM inf,ROOT-,AUX-,INV-,MARKING null,REL{},QUE{},INHER|SLASH{[1]}]

When the default values in (49.a) are factored out of the SELECT values in (50)-(53) by default unification, the values that must be stored are the (b) values in (54)-(57). (Recall from (47.c) that the default in the context of an interrogative filler is [INHER|SLASH{[1]}].)

54. a. *Who* will Kim like?  
 b. SELECT [AUX+,INV+]
55. a. I wonder *who* Kim will like.  
 b. SELECT [ROOT-,AUX+]
56. a. I wonder *who* Kim likes.  
 b. SELECT [ROOT-]
57. a. Kim asked *who* to choose for the team.  
 b. SELECT [CAT VP,VFORM inf,ROOT-,AUX+]

Note that the values in (54.b)-(57.b) are much more compact than the SELECT values in which all defaults are specified. Also, the SELECT values are linguistically appealing, and they can be acquired from positive evidence. However, they are still not maximally compact in (54.b)-(57.b). They can be collapsed together by conservative disjunction, as

---

<sup>25</sup>I suggested in Ch.4 that V[VFORM inf] by default does not have a subject. It is desirable that this default be formulated in such a way that acquiring the information that a filler selects [VFORM inf] makes it unnecessary to record that the filler also selects VP.

discussed in Chapter 3, into the three SELECT values in (58).

58. a. SELECT [AUX+, INV+]  
 b. SELECT [ROOT-, AUX +V-]  
 c. SELECT [CAT VP, VFORM inf, ROOT-, AUX+]

(58) is safe and conservative, and although not ideally economical, it is more compact than (54.b)-(57.b).<sup>26</sup> The lexical entries in (58) cannot be collapsed further by single-valued conservative disjunction (Chapter 3). In Section 6.5, I will propose a system with pointers rather than disjunction is psychologically more plausible.

#### 6.4 PHRASAL FILLERS

Interrogative fillers differ from markers in that they can be phrasal as well as lexical.

59. Interrogative head of filler phrase<sup>27</sup>  
 a. DP[*which* panda]  
 b. DP[*whose* (mother's) panda]  
 c. DP[*how many* pandas]

Moreover, the interrogative word is often not the head of the filler, as shown in (59).

60. Non-interrogative head of filler phrase  
 a. PP[*to which* panda]  
 b. PP[*from who*]  
 c. NP[*pictures of who*]

---

<sup>26</sup>Note that without the feature AUX, (58) would be much more compact. [AUX+/-] is correlated with INV in an intriguing way, which suggests that possibly they are not orthogonal, and that revising their descriptive load in some way might offer greater economy of storage.

<sup>27</sup>I assume that the interrogative words *which*, *whose*, and *how* are determiners in (59); recall that P&S-94 analyze determiners as heads of determiner phrases. If such interrogative words are not determiners, it might be the case that interrogative words are never the heads of filler phrases.

## d. NP[pictures of which panda]

In this section, I show that interrogative phrases are sensitive to the properties of the head daughter of Schema 6, and therefore, like lexical fillers, must be associated with the feature which I call SELECT. The issue to be addressed is where the value of SELECT is established in the grammar. Ideally, for uniformity with lexical fillers, it would be stored in the lexical entry of interrogative words, even when the interrogative word is not the head of the interrogative phrase.

## 6.4.1 LEXICAL STORAGE OF PHRASAL SELECT

Interrogative phrases have different selectional properties. For example, prepositional phrases are grammatical with finite clauses, as in (61.a), but not with infinitives, as in (61.b), at least in some dialects of English.<sup>28</sup>

61. a. I wonder [to which panda S[Kim gave the bamboo  
 \_\_\_]]? S[VFORM fin]
- b. \*I wonder [to which panda VP[to give the bamboo  
 \_\_\_]]. VP[AUX+,VFORM inf]

The ungrammaticality of the sentences in (61) is due to syntactic or selectional factors rather than semantic ones,

---

<sup>28</sup>An advantage of this learning model is that it can easily accommodate dialect variation: learners acquire exactly the constructions that they hear, and no others, within the range allowed by UG. Learners of different dialects encounter somewhat different sets of local trees as their input, so they will have different grammars.

as can be seen from the fact that both PP occurs with S[VFORM fin] where the meaning is roughly equivalent.

62. I wonder [to who S[I should give the bamboo     ]].  
S[VFORM fin]

So SELECT values of phrasal fillers are defined, at least in part, on the basis of the syntactic category of the filler.

Interrogative phrases sometimes are distinguished on the basis of other features. For example, Henry (1992) reports that some speakers of Belfast English distinguish between the SELECT value of interrogative phrases and the SELECT value of single words.

63. Belfast English (Henry 1992, fn.3)  
 a. It depends which story that <sub>s</sub>[you believe     ].  
 b. \*It depends which that <sub>s</sub>[you believe     ].

These speakers allow that only after a wh-phrase, not after a single wh-word (and not after *whether*). So learners of Belfast English must learn that single words select [MARKING null] but that phrases select [MARKING that].

(Alternatively, perhaps, *which* selects [MARKING that] when it is a determiner, but selects [MARKING null] when it is a pronominal.)

The opposite distinction is made in Swiss Bernese. *dass* is "generally preferred but not obligatory" with simple interrogative expressions (Penner and Bader (ms.:51)).

64. (Penner and Bader (ms., Ch.4(13a))  
 i ha ne gfragt wo dass er wohnt  
 I have him asked where that he lives  
 'I asked him where he lives'

*dass* is ruled out with restrictive and free relatives; however, complex *wh*-expressions generally require *dass*.<sup>29</sup>

65. Indirect question with complex *wh*-expressions (Penner and Bader (ms., Ch.4, 13b))  
 i ha ne gfragt i welem Huus \*(*dass*) er wohnt  
 I have him asked in which house that he lives  
 'I asked him in which house he lives.'

These examples show that learners must acquire correlations between the features that distinguish interrogative fillers, and their inherent SELECT values. The grammars of speakers of Swiss Bernese and Belfast English must include selectional information about the class of interrogative fillers which are multi-word as well as those which are single words.

On the other hand, classes of interrogative phrases are sometimes blind to the distinction between multi-word and one-word fillers, but sensitive to other distinctions. For example, in Spanish, all interrogative expressions which are arguments select obligatory inversion, as in (66), but inversion is optional for interrogative expressions that are adjuncts, as in (67). (I temporarily adopt the feature LEX (not used in this way in P&S-94) to represent the distinction between single word fillers, which are [LEX+]

---

<sup>29</sup>*dass* is strongly favored with root and embedded exclamatives.

- i. Root and embedded exclamatives (Penner and Bader (ms., Ch.4(13e)))  
 wo dass dää scho überau isch gsi!  
 where that he already everywhere is been

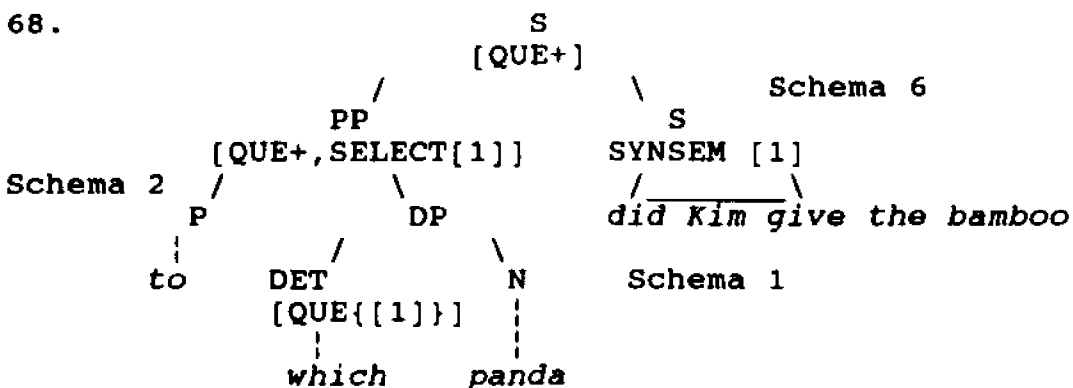
and multi-word fillers which are [LEX-]; I show below that this feature is not needed.)

66. Spanish (Torrego 1984 (2-3))
- a. *Qué querían esos dos?* [LEX+, ARG+, INV+]  
 what want those two  
 'What did those two want?'
- b. \**Qué esos dos querían?* [LEX+, ARG+, INV-]  
 what those two want
- c. *Con quién vendrá Juan hoy?* [PHR+, ARG+, INV+]  
 With who will-come John today  
 'With who will John come today?'
- d. \**Con quién Juan vendrá hoy?* [PHR+, ARG+, INV-]  
 With who John will-come today
67. (Torrego 1984(15), modified)
- a. *Cuándo Juan consiguió por fin abrir la puerta*  
 When Juan got finally to open the door  
 ayer?  
 yesterday [LEX+, ARG-, INV-]  
 'When did John finally get the door open  
 yesterday?'
- b. *Cuándo consiguió Juan por fin abrir la puerta*  
 When got Juan finally to open the door  
 ayer? [LEX+, ARG-, INV+]  
 yesterday  
 'When did John finally get the door open  
 yesterday?'
- c. *En que medida la constitución ha contribuido a eso?*  
 In what way the constitution has contributed to  
 that [PHR+, ARG-, INV-]  
 'In what way has the constitution contributed to  
 that?'
- d. *En que medida ha contribuido la constitución a eso?*  
 In what way has contributed the constitution to  
 that [PHR+, ARG-, INV+]  
 'In what way has the constitution contributed to  
 that?'

So fillers, whether lexical or phrasal, must also be distinguished as to whether they are arguments or adjuncts.<sup>30</sup> These examples show that phrasal fillers do indeed impose selectional requirements on the head daughter of Schema 6, just as lexical fillers do; since these requirements vary from language to language, they must be learned.

Moreover, since only the phrasal filler is in the same local tree as the head daughter of Schema 6, SELECT must be an appropriate feature for phrasal interrogatives. Note that *which* is not a daughter of the PP in (68).

68.



However, since every interrogative filler has a lexical descendent, the question is whether the phrasal category or the lexical category does the selection.

o If the SELECT value of phrasal fillers is stored in association with a phrasal node, syntactic learning is involved.

---

<sup>30</sup>I will not consider here how the information about the distinct selectional properties of arguments and adjuncts should be represented.

○ If the SELECT value of the phrasal filler is stored in the lexical entry of the interrogative daughter, lexical learning is involved.

Recall from Chs. 1 and 2 that I have adopted the hypothesis that all learning is lexical, which is consistent with the hypothesis that all language differences (except for parameter settings) are lexical. Since lexical learning is evidently required (regardless of the analysis of phrasal fillers), I adopt a conservative research strategy: I make the initial assumption that all learning is lexical, and I explore to what extent this permits the learning of apparently syntactic phenomena. I will conclude that syntactic learning is required only if linguistic evidence shows that lexical learning does not suffice. In Section 6.4.2 I show that SELECT values of interrogative phrasal fillers be stored lexically and thus acquired by lexical learning. In Section 6.4.3, I show that default unification allows the filler values to be stored compactly, and in Section 6.5.3, I briefly consider a potential linguistic challenge to the hypothesis that all SELECT values are stored lexically.

#### 6.4.2 ACQUIRING THE VALUE OF SELECT

The SELECT value of an interrogative word which is a (head or non-head) daughter of an interrogative phrase can be acquired in virtually the same way as the SELECT value of a lexical interrogative filler can, except that the learning



(70) eliminates the need for the independent SELECT Principle (32) by requiring that the value of SELECT on the non-head daughter of Schema 6 be structure-shared with the SYNSEM value of its head daughter. To ensure that the SELECT value of the entire filler phrase is structure-shared with the SELECT value of the interrogative daughter, I propose the Phrasal SELECT Principle.

71. Phrasal SELECT Principle  
 When a phrase and its daughter both bear the feature SELECT, the SELECT value of the phrase and its daughter must be structure-shared.

This recursive principle applies if the daughter is lexical, as in (72.a), or phrasal, as in (72.b).

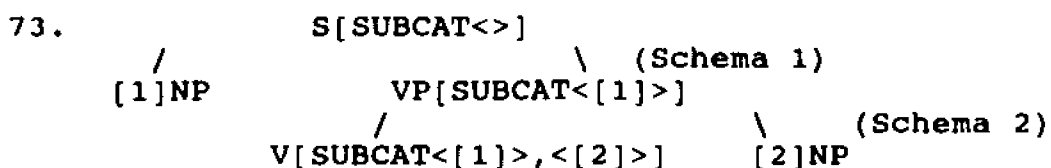
72. a. NP[Who] will Kim consult about the sick panda?  
 b. PP[Near the NP[feeding bowl PP[of which panda]]]  
 did Kim drop her hat?

Since the SELECT value of the filler phrase must be structure-shared with the lexical interrogative, whether or not it is the head-daughter, SELECT is not a head feature. Since SELECT is appropriate only for features with a marked value for QUE, it will not pass indefinitely up through the tree because the NONLOCAL Feature Principle, which constrains the distribution of QUE, indirectly constrains the distribution of SELECT.

With these modifications, the learning model works as it did previously. In learning, the SELECT value is structure-shared by the filler phrase and the lexical interrogative; the fully specified lexical node is transferred out of the syntax and into the lexicon for

storage in accordance with the Lexical Feature Principle (Ch.5(11)). This mechanism allows the selectional properties of phrasal fillers to be associated with individual lexical entries, whether or not the interrogative is the head daughter of the filler phrase.<sup>32</sup> In parsing and generation, the SELECT value of the interrogative word is transferred out of the lexicon to the lexical node in the syntactic tree, from where it percolates up to the correct phrasal level, carried along with the QUE feature.

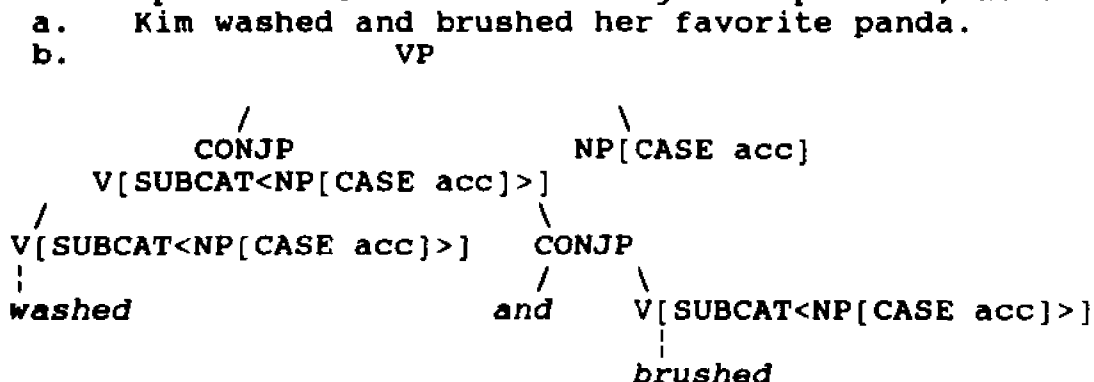
Note that the value of SUBCAT is also structure-shared between a VP, and the V in whose lexical entry it is stored.



Since the value of SUBCAT is the concatenation of the SYNSEM values of its complements (as in Ch.2(16)), the value of

---

<sup>32</sup>This percolation mechanism is similar to that which Gazdar (1981) showed to be required for SUBCAT in coordinate structures: the subcategorization requirements of a lexical conjunct must percolate up to the level of the conjoined phrases, as in (i.b).

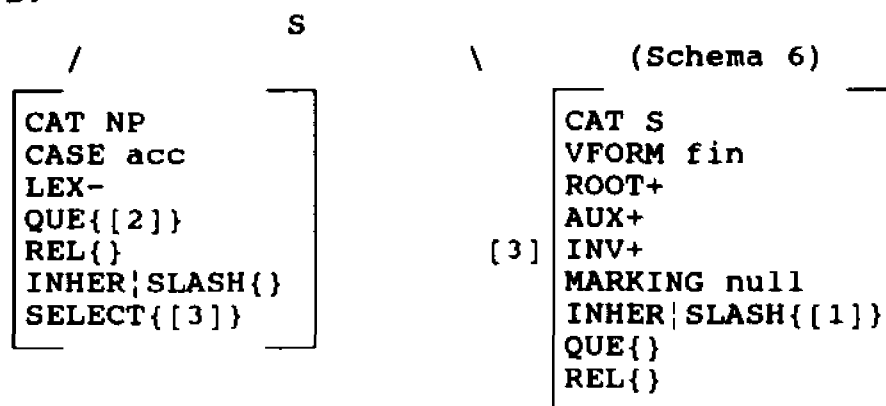


*Kim washed* could be intransitive, but unification with *brushed* forces it to be transitive in (i.a).

SUBCAT can be acquired in much the same way that the values of SELECT can.<sup>33</sup>

To see how the value of the lexical interrogative is learned when it is the daughter of a phrasal filler, consider (74).

74. a. Which panda will Kim see?  
b.



The NONLOCAL Feature Principle is satisfied in (74) because the QUE value of the NP is structure-shared with the interrogative determiner *which*; the Phrasal SELECT Principle is satisfied since the SELECT value of the NP is structure-shared with *which*.<sup>34</sup> The information in (75) is copied from the syntax into the lexicon, sans universal syntactic

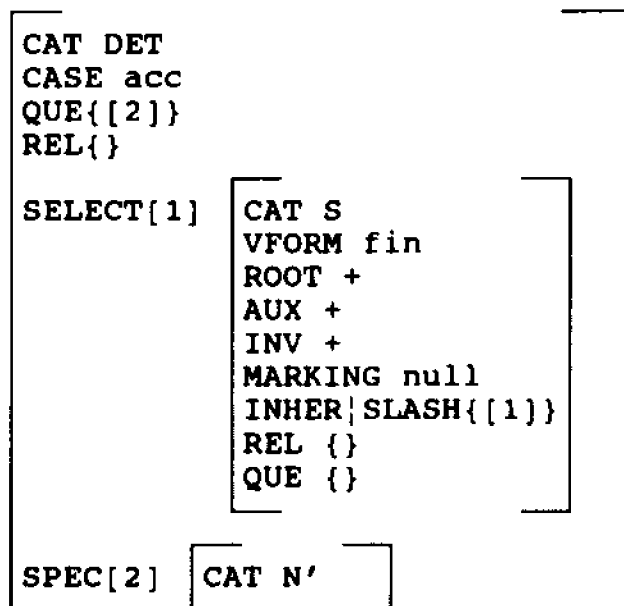
<sup>33</sup>I set aside the problem of the status of determiners, which has been an unstable aspect of phrase structure grammar. P&S-94 analyze determiners (including possessive 's), as head categories for which SPEC is appropriate. An alternative analysis is that determiners are non-heads; this raises semantic problems discussed in P&S-94:47ff.

<sup>34</sup>Note that the Phrasal SELECT Principle is a cost of the requirement that language-particular information is stored lexically. If syntactic learning were permitted, as in GPSG, a language-particular phrase structure rule could be acquired directly from (74.b). However, since the Phrasal SELECT Principle is innate, it does not need to be learned.

properties, in accordance with the Lexical Feature Principle (Ch.5(11)).

However, the [LEX-] of the phrasal filler is not appropriate for the interrogative determiner which is a lexical item; moreover, LEX is not a percolating feature. However, lexical SELECT values provide an alternative to Henry's claim that in Belfast Irish multi-word interrogatives occur with *that*, but single word ones do not; instead, the distinction should be attributed to the fact that determiners and nouns with the same phonological representation and otherwise identical features (at least for those features under consideration here) may have different selectional properties. (75) thus represents the lexical entry for the determiner *which* (without the feature LEX).

75.



When *which* is a determiner, it will necessarily combine with an N', thus forming a multi-word NP. (Recall from Chapter 3

that determiners are also associated with the feature SPEC, by which they specify SYNSEM values of the N' with which they may combine. I include the SPEC value in (75) for the sake of illustration, but I then set it aside for the remainder of this discussion.) However, when *which* is a noun, it will not combine with any adjectives or determiners, and so will necessarily be a single word.

#### 6.4.3 COMPACT STORAGE OF PHRASAL SELECT VALUES

For the lexical entry to be stored compactly in the lexicon, without predictable information, default values must be factored out. In order to give a concrete example of how this works, I now consider default values for the filler daughter (excluding the SELECT value). As above, I suggest default values on relatively little evidence in order to make a concrete proposal. The context-free defaults previously proposed for the features appropriate for the filler are in (76). (SELECT does not have a default; its value is universally determined by Schema 6 and by the Phrasal SELECT Principle.)

76. cf. (35.a)

|      |     |
|------|-----|
| CAT  | S   |
| CASE | acc |
| QUE  | { } |
| REL  | { } |

I now consider whether these context-free defaults are appropriate, or whether a context-sensitive default is required for any of these features.

With regard to CAT, it seems that fillers may be NPs in all or most languages with fillers, so I conjecture that the default for interrogative fillers is N. I therefore posit the context-sensitive default in (77), where QUE{[1]} represents a singleton value of QUE.<sup>35</sup>

77. [QUE{[1]},CAT N]

(77) says that an interrogative lexical item (i.e. one with the feature [QUE{[1]}) is by default a noun. (77) requires a modest extension of the restriction proposed in Chapter 5, that the context for all context-sensitive defaults is a lexical category. I suggested that lexical category features were safe contexts because they can never be factored out of a lexical entry. In this model, Schema 6 forces one of the NONLOCAL features to have a non-empty value, but no default specifies that one type of filler is more typical than the other.<sup>36</sup> Therefore the singleton value of QUE, while not a lexical category, is similarly safe because it is a marked feature which cannot be factored out by default unification; as a consequence, the value of

---

<sup>35</sup>In a system with syntactic defaults, Schema 6 might include the specification that all its non-head daughters are by default [CAT NP]; this is a case where the requirement that all defaults are lexical entails a loss of generalization.

<sup>36</sup>If syntactic defaults were permissible in this model, one might conjecture that there are 'schema-particular' defaults. In this case, the filler daughter of Schema 6 might specify the singleton value for QUE (or, alternatively, REL or a topicalized constituent) as a default value; the singleton value of SLASH would be specified as a default value on the head daughter .

QUE will always be specified in the lexical entry, so that the correct default value can be determined.

Note that the singleton value for QUE imposes a default value on the same node that it occurs on, rather than on its sister node.

78.

|                                     |
|-------------------------------------|
| CAT N<br>QUE{[1]}<br>SELECT [CAT S] |
|-------------------------------------|

(78) specifies that an interrogative will be an N by default, but that its sister node, the clause with the gap, will be an S by default. So two different defaults for CAT apply in (78): the context-sensitive value is instantiated on the filler node itself, because of (77), and the context-free default is instantiated on its sister. For CASE, I know of no reason to reconsider the default [CASE acc] proposed in Chapter 4. For the NONLOCAL features, I suggested above that the empty set is the context-free default; here too, this seems correct.

In summary, the context-sensitive defaults applicable to interrogative fillers (except for the values of SELECT) can be default unified with the context-free defaults, as in (79).

79. Default values for filler daughter that is [QUE {[1]}] (excluding values of SELECT).

| a.<br>CONTEXT-SENSITIVE<br>DEFAULTS | b.<br>CONTEXT-FREE<br>DEFAULTS                                                                                                                                                                      | c.<br>APPLICABLE<br>DEFAULTS |          |        |        |                                                                                                                                                                                           |       |          |        |        |
|-------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|----------|--------|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|----------|--------|--------|
| [CAT N]                             | $\% \sqcup$ <table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td>CAT S</td></tr> <tr><td>CASE acc</td></tr> <tr><td>QUE {}</td></tr> <tr><td>REL {}</td></tr> </table> | CAT S                        | CASE acc | QUE {} | REL {} | = <table border="1" style="display: inline-table; vertical-align: middle;"> <tr><td>CAT N</td></tr> <tr><td>CASE acc</td></tr> <tr><td>QUE {}</td></tr> <tr><td>REL {}</td></tr> </table> | CAT N | CASE acc | QUE {} | REL {} |
| CAT S                               |                                                                                                                                                                                                     |                              |          |        |        |                                                                                                                                                                                           |       |          |        |        |
| CASE acc                            |                                                                                                                                                                                                     |                              |          |        |        |                                                                                                                                                                                           |       |          |        |        |
| QUE {}                              |                                                                                                                                                                                                     |                              |          |        |        |                                                                                                                                                                                           |       |          |        |        |
| REL {}                              |                                                                                                                                                                                                     |                              |          |        |        |                                                                                                                                                                                           |       |          |        |        |
| CAT N                               |                                                                                                                                                                                                     |                              |          |        |        |                                                                                                                                                                                           |       |          |        |        |
| CASE acc                            |                                                                                                                                                                                                     |                              |          |        |        |                                                                                                                                                                                           |       |          |        |        |
| QUE {}                              |                                                                                                                                                                                                     |                              |          |        |        |                                                                                                                                                                                           |       |          |        |        |
| REL {}                              |                                                                                                                                                                                                     |                              |          |        |        |                                                                                                                                                                                           |       |          |        |        |

Recall that Schema 6 specifies that its non-head daughter must bear the feature SELECT whose value is structure-shared with the head daughter of Schema 6. The defaults proposed in (47.c) for the SELECT values of interrogative fillers are repeated here as (80).

80. DEFAULTS FOR SELECT VALUE OF INTERROGATIVE FILLERS  
(=(47.c))

|                   |
|-------------------|
| CAT S             |
| VFORM fin         |
| ROOT -            |
| INV -             |
| AUX -             |
| MARKING null      |
| INHER SLASH {[1]} |
| QUE {}            |
| REL {}            |

The applicable defaults for interrogative fillers are as in (81), the result of unification of (79.c) with (80).

## 81. Defaults for interrogative fillers:

|                                                                                                         |           |                                                                                                                                                                                                                                                                                                   |                                                                                                         |
|---------------------------------------------------------------------------------------------------------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| CAT N<br>CASE acc<br>QUE {}<br>REL {}                                                                   | SELECT[1] | <table border="1"> <tr> <td style="padding: 5px;">           CAT S<br/>           ROOT +<br/>           MARKING null<br/>           VFORM fin<br/>           AUX -<br/>           INV -<br/>           INHER SLASH {{1}}<br/>           QUE {}<br/>           REL {}         </td> </tr> </table> | CAT S<br>ROOT +<br>MARKING null<br>VFORM fin<br>AUX -<br>INV -<br>INHER SLASH {{1}}<br>QUE {}<br>REL {} |
| CAT S<br>ROOT +<br>MARKING null<br>VFORM fin<br>AUX -<br>INV -<br>INHER SLASH {{1}}<br>QUE {}<br>REL {} |           |                                                                                                                                                                                                                                                                                                   |                                                                                                         |

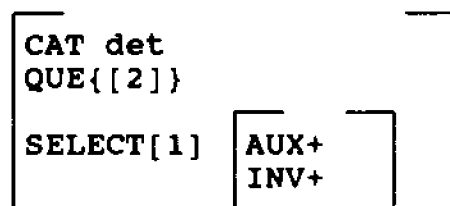
The lexical node that is passed into the lexicon from the fully specified lexical node is (82).

82. FULLY SPECIFIC LEXICAL ENTRY FOR *which* (= (75))

|                                                                                                        |           |                                                                                                                                                                                                                                                                                                  |                                                                                                        |
|--------------------------------------------------------------------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE acc<br>QUE{{2}}<br>REL{}                                                               | SELECT[1] | <table border="1"> <tr> <td style="padding: 5px;">           CAT S<br/>           VFORM fin<br/>           ROOT +<br/>           AUX +<br/>           INV +<br/>           MARKING null<br/>           INHER SLASH{{1}}<br/>           QUE {}<br/>           REL {}         </td> </tr> </table> | CAT S<br>VFORM fin<br>ROOT +<br>AUX +<br>INV +<br>MARKING null<br>INHER SLASH{{1}}<br>QUE {}<br>REL {} |
| CAT S<br>VFORM fin<br>ROOT +<br>AUX +<br>INV +<br>MARKING null<br>INHER SLASH{{1}}<br>QUE {}<br>REL {} |           |                                                                                                                                                                                                                                                                                                  |                                                                                                        |

When default unification is applied to (81) and (82) to solve for the acquired generalization, what the learner must store is (83).

83. (cf.49.b)  
Stored information about *which*[acc].



(83) is a compact lexical entry from which all features which are predictable because they are supplied by UG have been factored out. Since all interrogative phrases have a lexical interrogative descendant, this example demonstrates that the selectional properties of phrasal fillers can be stored lexically.

#### 6.4.4 PHRASAL SELECT: LEXICAL OR SYNTACTIC

In the previous section, I showed that interrogative and relative phrasal fillers must be associated with SELECT values, and that these SELECT values can be stored lexically; on this analysis, phrasal SELECT values can be acquired by lexical learning. In this section, I consider a linguistic question: is the likely ultimately to support the hypothesis that phrasal selection is exclusively lexical? Unfortunately it is not clear that all storage of SELECT values is exclusively lexical; if it is not, the hypothesis that all grammatical learning (except parameter settings) is strictly lexical is too strong.

One English example, noted by Fodor (p.c.) does suggest that phrasal selection might be lexical.

84. a. *What* socks Leslie was wearing!  
b. \**Which* socks Leslie was wearing!

However, this selection might be semantic. Moreover, in general, the kind of selectional distinctions that would be expected if selection were lexical do not appear to occur in English; these would include distinctions like the following non-facts in (85), pointed out by Fodor (p.c.).

85. NON-FACTS ABOUT ENGLISH
- a. SELECT [VFORM fin]
    - i. *Whom*
    - ii. *To whom*
    - iii. *Of whom*
    - iv. *With whom*
  - b. SELECT [VFORM inf]
    - i. *which boy*
    - ii. *to which boy*
    - iii. *of which boy*
    - iv. *with which boy*
  - c. SELECT [VFORM base]
    - i. *what boy*
    - ii. *to what boy*
    - iii. *of what boy*
    - iv. *with what boy*

However, this evidence is based on the non-appearance of a phenomenon, and although suggestive, it is not conclusive. Further research is required to determine whether in some language, the particular lexical item determines the type of *SELECT* value that is selected. If such a distinction among lexical items can be found in even one language, then lexical storage of *SELECT* values associated with phrasal fillers is linguistically justified.

However, even if such evidence of lexical selection could be found, it is also not clear that lexical selection is sufficient. The fact that Spanish distinguishes between adjuncts and argument fillers, as shown in (66) and (67)

above, but is (apparently) blind to the particular interrogative words, suggests that at least some phrasal learning is involved. The evidence is not definitive, and this is a point that requires further investigation. However, since I have adopted the conservative research strategy that all language-particular information is stored lexically unless irrefutable evidence to the contrary is discovered, I continue to assume that the SELECT values of all interrogative filler phrases are indeed stored lexically.

#### 6.5 GENERALIZATION WITH POINTERS

Learners must acquire an additional lexical entry for each use of *which* differs from (82), and hence (83), in any respect. As noted above, this problem is distinct from that of reducing individual lexical entries by omission of predictable features. The problem here is how the learner can collapse the similar lexical entries that it will acquire, without having to store them all separately. One possibility is conservative disjunction, although, as noted in Chapter 5, it is not an entirely attractive option. In this section, I suggest, very briefly, that a pointer system provides a way to model this process that is more psychologically plausible than disjunction.

In addition to (74), a learner of English will encounter other local trees with interrogative non-head daughters and finite heads. I list in (86.b)-(93.b) the fully speci-

fic value of the lexical interrogative node. (I use judgements from my dialect, but note certain dialectical variations.)

86. a. I wonder [which panda [Kim will feed]].

b.

|                                                                                                        |           |                                                                                                                                                                                                                                                                                                  |                                                                                                        |
|--------------------------------------------------------------------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE acc<br>QUE{[2]}<br>REL{}                                                               | SELECT[1] | <table border="1"> <tr> <td style="padding: 5px;">           CAT S<br/>           VFORM fin<br/>           ROOT -<br/>           AUX +<br/>           INV -<br/>           MARKING null<br/>           INHER SLASH{[1]}<br/>           QUE {}<br/>           REL {}         </td> </tr> </table> | CAT S<br>VFORM fin<br>ROOT -<br>AUX +<br>INV -<br>MARKING null<br>INHER SLASH{[1]}<br>QUE {}<br>REL {} |
| CAT S<br>VFORM fin<br>ROOT -<br>AUX +<br>INV -<br>MARKING null<br>INHER SLASH{[1]}<br>QUE {}<br>REL {} |           |                                                                                                                                                                                                                                                                                                  |                                                                                                        |

87. a. I wonder [which panda [Kim fed]].

b.

|                                                                                                        |           |                                                                                                                                                                                                                                                                                                  |                                                                                                        |
|--------------------------------------------------------------------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE acc<br>QUE{[2]}<br>REL{}                                                               | SELECT[1] | <table border="1"> <tr> <td style="padding: 5px;">           CAT S<br/>           VFORM fin<br/>           ROOT -<br/>           AUX -<br/>           INV -<br/>           MARKING null<br/>           INHER SLASH{[1]}<br/>           QUE {}<br/>           REL {}         </td> </tr> </table> | CAT S<br>VFORM fin<br>ROOT -<br>AUX -<br>INV -<br>MARKING null<br>INHER SLASH{[1]}<br>QUE {}<br>REL {} |
| CAT S<br>VFORM fin<br>ROOT -<br>AUX -<br>INV -<br>MARKING null<br>INHER SLASH{[1]}<br>QUE {}<br>REL {} |           |                                                                                                                                                                                                                                                                                                  |                                                                                                        |

88. a. [Which panda [will Kim give the food to]]?  
 b.

|                                          |           |                                                                                                        |
|------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE obl<br>QUE{{2}}<br>REL{} | SELECT[1] | CAT S<br>VFORM fin<br>ROOT +<br>AUX +<br>INV +<br>MARKING null<br>INHER SLASH{{1}}<br>QUE {}<br>REL {} |
|------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------|

89. a. I wonder [which panda [Kim will give the food to]].  
 b.

|                                          |           |                                                                                                        |
|------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE obl<br>QUE{{2}}<br>REL{} | SELECT[1] | CAT S<br>VFORM fin<br>ROOT -<br>AUX +<br>INV -<br>MARKING null<br>INHER SLASH{{1}}<br>QUE {}<br>REL {} |
|------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------|

90. a. I wonder [which panda [Kim gives the food to]].

b.

|                                          |           |                                                                                                        |
|------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE obl<br>QUE{[2]}<br>REL{} | SELECT[1] | CAT S<br>VFORM fin<br>ROOT -<br>AUX -<br>INV -<br>MARKING null<br>INHER SLASH{[1]}<br>QUE {}<br>REL {} |
|------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------|

91. a. To which panda will Kim give that food?

b.

|                                          |           |                                                                                                        |
|------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE obl<br>QUE{[2]}<br>REL{} | SELECT[1] | CAT S<br>VFORM fin<br>ROOT +<br>AUX +<br>INV +<br>MARKING null<br>INHER SLASH{[1]}<br>QUE {}<br>REL {} |
|------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------|

92. a. I wonder to which panda Kim will give that food.  
Note: ungrammatical in some dialects of English

b.

|                                                           |                                                                                                        |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE obl<br>QUE{[2]}<br>REL{}<br><br>SELECT{1} | CAT S<br>VFORM fin<br>ROOT -<br>AUX +<br>INV -<br>MARKING null<br>INHER SLASH{[1]}<br>QUE {}<br>REL {} |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------------|

93. a. I wonder to which panda Kim gave that food.  
Note: ungrammatical in some dialects of English

b.

|                                                           |                                                                                                        |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE obl<br>QUE{[2]}<br>REL{}<br><br>SELECT{1} | CAT S<br>VFORM fin<br>ROOT -<br>AUX -<br>INV -<br>MARKING null<br>INHER SLASH{[1]}<br>QUE {}<br>REL {} |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------------|

The learner must also acquire the selectional properties of fillers vis a vis infinitives in English. For example, learners will be exposed to sentences like (94).<sup>37</sup>

<sup>37</sup>A learner exposed to my dialect of English will not encounter (i)(=61.b).

i. \*Kim asked to which panda to give the bamboo \_\_\_\_.  
This is because oblique *which* does not select infinitive VPs.

94. a. Kim asked *which* panda to give the bamboo to.

b.

|                                          |           |                                                                                                         |
|------------------------------------------|-----------|---------------------------------------------------------------------------------------------------------|
| CAT DET<br>CASE obl<br>QUE{{2}}<br>REL{} | SELECT[1] | CAT VP<br>VFORM inf<br>ROOT -<br>AUX +<br>INV -<br>MARKING null<br>INHER SLASH{{1}}<br>QUE {}<br>REL {} |
|------------------------------------------|-----------|---------------------------------------------------------------------------------------------------------|

The number of lexical entries required for the relatively small grammar fragment under consideration here is an illustration of the variety of feature structures manifested in natural language. Non-empty values for REL potentially could double the number of lexical entries, as would each additional Boolean feature. Nevertheless, it is a fact that learners eventually acquire this large set of trees. Obviously it is not plausible to suppose that they record each one separately and in full.

Generalization by factoring out of default values provides considerable economy, as in (95.a)-(95.i).

95. a. (cf.74)  
       [CAT DET, QUE{{2}}, SELECT[AUX+,INV+]]
- b. (cf.86)  
       [CAT DET, QUE{{2}}, SELECT[ROOT-,AUX+]]
- c. (cf.87)  
       [CAT DET, QUE{{2}}, SELECT[ROOT-]]
- d. (cf.88)  
       [CAT DET, QUE{{2}}, CASE obl, SELECT[AUX+,INV+]]

- e. (cf.89)  
[CAT DET, QUE{{2}}, CASE obl, SELECT[ROOT-,AUX+]]
- f. (cf.90)  
[CAT DET, QUE{{2}}, CASE obl, SELECT [ROOT-]]
- g. (cf.91)  
[CAT DET, QUE {{2}}], SELECT [AUX+,INV+]]
- h. (cf.92)  
[CAT DET, QUE {{2}},CAT PP, SELECT [ROOT-,AUX+]]
- i. (cf.93)  
[CAT DET, QUE {{2}}, CAT PP, SELECT[ROOT -]]
- j. (cf.94)  
[CAT DET, QUE{{2}},CASE obl] SELECT [CAT VP,VFORM  
inf,ROOT-,AUX+]]

However, there is still considerable redundancy in (95.a)-(95.j), just as there was in (54.b)-(57.b),. According to the principle of conservative disjunction proposed earlier, these lexical entries can be combined by conservative disjunction only when two feature structures differ by exactly one feature value. This solves the problem technically, but produces an extremely unwieldy structure that is too complex to represent here and doesn't meet any reasonable criteria for simplicity of representation.<sup>38</sup> So although it works technically, conservative disjunction that generalizes over multiple lexical entries is not a satisfactory solution.

However, pointers offer another, more psychologically plausible solution. Note that in (54.b)-(58.c), for lexical fillers, and (95.a)-(95.j), for phrasal fillers, there are

---

<sup>38</sup>Berwick 1985:233 also notes that processing disjunction is very time consuming.

only four SELECT values, which I represent arbitrarily as **S1-S4**.

|                             |               |           |                                      |
|-----------------------------|---------------|-----------|--------------------------------------|
| 96. <sup>39</sup> <b>S1</b> | <b>S2</b>     | <b>S3</b> | <b>S4</b>                            |
| INV+<br>AUX+                | ROOT-<br>AUX+ | [ROOT-]   | CAT VP<br>VFORM inf<br>ROOT-<br>AUX+ |

I discussed in Chapter 2 the fact that syntactic pointers point to token-identical values. Suppose that **S1-S4** represent four SELECT value types, each of which is stored in the lexicon exactly once. Each **S**-value type is pointed to by as many SELECT as it is selected by. For example, (95.a), (95.d), and (95.g) will each point to the same SELECT value.

97. a. [CAT DET, QUE{[2]}, SELECT [**S1**]]  
 b. [CAT DET, QUE{[2], CASE obl, SELECT [**S1**]]  
 c. [CAT DET, QUE {[2]}], SELECT [**S1**]]

Pointers thus preclude the necessity of storing multiple copies of type-identical SELECT values, or of combining these SELECT values by conservative disjunction.

Note that in English none of the head daughters selects all of the default values in (80), as can be seen from the ungrammaticality of the sentences in (59.a, 98).

98. a. \*Who Kim sees?  
 b. \*Which panda Kim feeds?

---

<sup>39</sup>Note that [AUX+] is responsible for a lot of the redundancy in **S1-S4**. If AUX could be replaced by a different feature (or features), some of this redundancy might be eliminated.

How then does the learner distinguish between a language which licenses a head with only default values (i.e. where a feature structure with all default values is grammatical) and a language like English? Initially, the learner assumes that SELECT points to the default values in (80) that the filler imposes on its sister, which I will label **S\_def**, and repeat here as (99).

99. **S\_def**=(47.c)  
 DEFAULTS FOR SELECT VALUE OF INTERROGATIVE FILLERS

|             |       |
|-------------|-------|
| CAT         | S     |
| VFORM       | fin   |
| ROOT        | -     |
| INV         | -     |
| AUX         | -     |
| MARKING     | null  |
| INHER SLASH | {[1]} |
| QUE         | {}    |
| REL         | {}    |

Since (99) consists only of default values, it can be economically stored as the maximally empty list, as in (100).

100. **S\_def**  
 []

In all lexical entries for interrogative fillers, SELECT value initially points to **S\_def** ([]). However, since **S\_def** does not license a grammatical sentence in English, the learner must replace **S\_def** with some S value that is grammatical. Additional pointer values may be added as the learner encounters positive evidence that they are needed.

An advantage of the pointer system is that it makes psychological sense: once a feature description has been acquired and stored, adding another pointer to the same type

is not very costly. (Fodor (1992a) made a similar point in a slightly different technical framework.) So it is not surprising that a single SELECT value is pointed to by many fillers. But, conservative learning is needed to explain why the simpler pattern, in which all feature descriptions select all other feature descriptions, does not characterize natural language.

In summary, pointers, which enable the learner to store each value of SELECT only once, produce considerable economy of storage. The pointer system also eliminates the need for disjunction in storing local tree values, with its awkwardly complex representation.<sup>40</sup> Since disjunction requires active comparison of the feature structures to be collapsed, I conclude that pointer representation is preferable to disjunction. However, pointers (or at least the type of pointers used in (97)) do not completely eliminate the need to store duplicate information. It may be that it is simply a fact about natural language that learning involves acquiring this type of untidy generalization. However, it is undesirable to fall back on this solution until other possibilities are exhausted. One possible linguistic solution is that the syntactic theory might be able to discover different features which somehow fit the facts

---

<sup>40</sup>Further research is needed to see how much disjunction a pointer system could eliminate. Without implementation of a reasonably sized fragment of a lexicon, it is difficult to anticipate what the trade-offs will be.

better, and therefore allow more compact generalization; unfortunately it is not immediately clear how this might be done. A possible computational solution would be the discovery of a way to represent these generalizations with the right cost profiles, so that common subparts of lexical entries could be stored economically; such a solution would make collapsing of lexical entries effortless, but would not entail over-generalization that cannot be repaired without negative evidence.

## 6.6 CONCLUSION

In this chapter I have shown that the selectional properties of interrogative fillers, are learnable in exactly the way that selectional properties of complementizers are, with the additional complication that fillers can be phrasal. Therefore, the language-particular further specifications of Schema 6 associated with interrogatives can be acquired by the same process by which SPEC values were acquired and stored. The primary difference is that in phrasal learning, the SELECT value on the filler daughter of Schema 6 must be structure-shared with the lexical interrogative. The fully specified lexical node is transferred out of the syntax into the lexicon where default values are factored out off by default unification so that only marked values needed to be stored. Thus correct item-particular generalizations are stored as lexical entries that are compact, linguistically sound, and

consistent with the principles of markedness theory which are central to the mechanism of acquiring generalizations proposed here. These lexical entries, which are language-particular further specifications of Schema 6, are acquired conservatively, and automatically, from only positive evidence.

The modifications that I have proposed to P&S-94 in this chapter include the following:

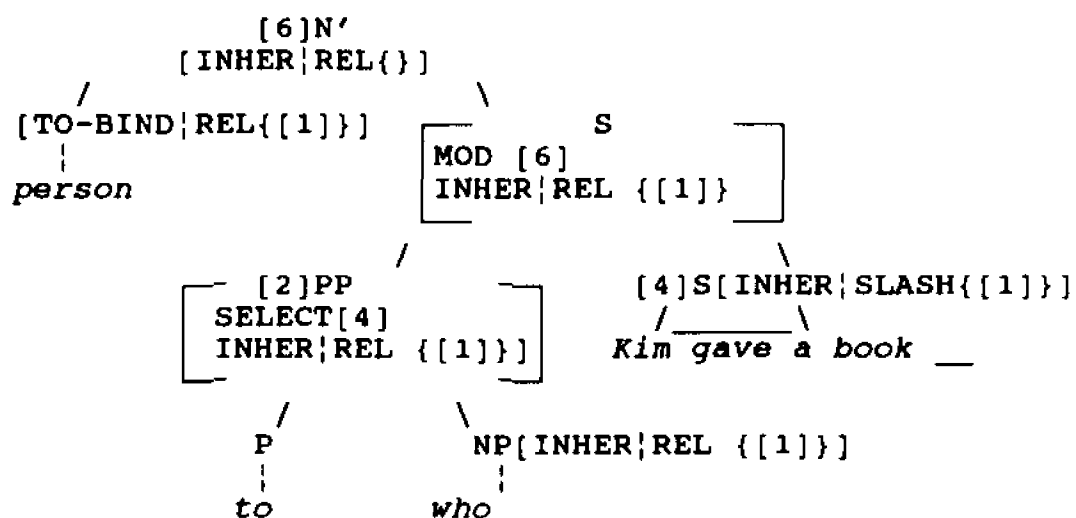
- Universal version of Schema 6. (See (11) and (70).)
- Interrogative fillers have a feature SELECT, formally similar to SPEC, which is appropriate for complementizers.
- SELECT can percolate, like SUBCAT, between a phrasal node and an appropriate lexical daughter.
- The SELECT value of an interrogative phrasal filler is stored in the lexical entry of the interrogative word which is a descendant of the filler phrase.
- Context-free defaults eliminate the need for the Clausal REL Constraint and Singleton REL Constraint, which, without negative evidence, are unlearnable.
- Context-sensitive defaults appropriate for interrogative fillers are these:
  - [QUE{[1]}, SELECT [INHER|SLASH {[1]}]] (= (46))
  - [QUE{[1]}, CAT N] (= (77))
- Pointers allow compact storage of SELECT values in the lexicon, eliminating the need to collapse them by conservative disjunction.

In Chapter 7, I will consider some further extensions to this learning model. I will show that the MOD values of lexical adjuncts can be acquired in the same way that SPEC and SELECT can. However, I will also show that the fact that headless adverbial adjuncts and topicalized constituents, which are not associated with a lexical item, have selectional properties poses a challenge to the hypothesis that all learning is lexical, and I will explore the types of solutions that are available.



complement for which the head is subcategorized, as with raising verbs. P&S-94 suggest a few solutions to this problem, all of which involve revision of the principle which constrains the semantic role that the first member of a SUBCAT list must make. However, if relative pronouns are like interrogatives in that they have a feature SELECT by which they specify the SYNSEM value of the head, the value of SUBCAT is simply irrelevant.

103.



The PP *to who* specifies properties of the head daughter of Schema 6; SELECT, rather than SUBCAT, is the appropriate feature selectional feature, because relative pronouns do not select semantic arguments. Semantically, the adjunct relative phrase combines with the head in accord with the Semantics Principle (discussed in Chapter 2). The Phrasal SELECT value ensures that the SELECT value of the filler will be structure-shared with the SELECT value of the relative pronoun. Relative phrases are thus like interrogative phrases in terms of their selectional properties,

and the need for a null relativizer with a SUBCAT value is eliminated; requirements about the semantic contribution made by members of a SUBCAT list to the semantics of the entire phrase are irrelevant. The SELECT value of the PP is stored in the lexical entry of the relative pronoun *who*; its SELECT value can be acquired in exactly the same way that the SELECT value of interrogative fillers can.

## CHAPTER 7 LEXICAL LEARNING AND LEXICAL DEFAULTS

In Chs. 5 and 6, I showed that the adoption of innate defaults consistent with the criteria proposed in Chapter 3 leads to a model of learning in which learners acquire compact language-particular generalizations from positive evidence without over-generalization. In Chapter 5, I showed that the phrasal properties of complements can be acquired by lexical learning because they are mediated by a selection feature in the lexical entry of the complementizers. In Chapter 6 I showed that co-occurrence relations between interrogative fillers, both lexical and phrasal, and their sister clauses, also can be acquired by lexical learning. Whether this is the correct treatment of phrasal fillers is something that future linguistic research must decide. In this chapter, I first consider co-occurrence constraints between adjuncts, which P&S-94 treat as non-head daughters of Schema 5, and their heads. Adjuncts have a feature MOD, by which they specify properties of the head that they modify. At issue here is whether the value of MOD is stored in the lexicon in the same way that the value of SPEC and SELECT are, so that the language-particular further specification of Schema 5 can be acquired by lexical learning.

I then consider two phenomena which pose a challenge to the claim that all relevant language-specific constraints are lexical: headless adjuncts, which are non-head daughters

of Schema 5: and topicalized constituents, which are non-head daughters of Schema 6. Headless adjuncts and topicalized constituents are phrases that are sensitive to properties of the head daughter of their respective schemas, even though they have no lexical daughter with which the selectional feature can be associated. In Section 7.2, I consider what the implications for the learning model proposed here will be if it turns out that phrases do have to have selectional features not inheritable from a lexical item. I conclude that if it should turn out that syntactic learning of sideways selectional properties is required, it is likely to be on a small scale relative to the amount of lexical learning required. However, the consequences for the formal theory and for the learning theory would not be small. Unless syntactic defaults could be re-introduced into the linguistic theory, such facts would be unlearnable from positive data.

In Section 7.3, I review the selecting properties of heads, complements, and non-head non-complements including fillers, markers, and adjuncts, and I show that all categories except complements come 'equipped' with a selectional feature that allows horizontal selectional properties to be stored in the lexicon. With a few modifications, the theory of P&S-94 therefore provides an interesting and consistent model of how syntactic relations between sister constituents are mediated; a unified account

of how the values of the selectional features are acquired is therefore possible.

It must be emphasized that the success of this learning method for acquiring compact and syntactically perspicacious generalizations crucially depends on the existence of universal defaults. At the end of this chapter, I review the implications, for HPSG, of the claim that defaults are important linguistically, and for acquisition.

### 7.1 THE SELECTIONAL PROPERTIES OF ADJUNCTS

As noted in Chapter 2, P&S-94 treat adjuncts such as adjectives and adverbs as non-heads that select properties of their head via the feature MOD(IFIED) in a configuration licensed by Schema 5.

1. Schema 5: Adjunct-Head Schema (P&S-94(Ch.1(60)))  
A phrase with DTRS value of sort *head-adjunct-structure* (*head-adj-struc*) such that the MOD value of the adjunct daughter is token-identical to the SYNSEM value of the head-daughter.

An adjective modifies an N'; an adverb may modify another adverb, an adjective, or a verbal category.<sup>1</sup> Adjuncts also select semantic properties of the heads that they modify. So the lexical entry for an adjective like *red* includes the specification that it modifies an N', as well as semantic selectional facts, as discussed in Chapter 2.

---

<sup>1</sup>Whether adjuncts select other syntactic properties of the heads that they modify requires further investigation.

2. (cf. Ch.2(33))



MOD is 'built in' to Schema 5, so that it is an obligatory feature for every adjunct daughter. There is no other schema for introducing adjuncts. Since the value of MOD is structure-shared with the SYNSEM value of the head daughter in Schema 5, the value of MOD associated with a lexical head can be transferred to the lexicon and stored compactly there by the same learning process that I have shown to be effective for acquiring the value of SELECT. The challenge for this model, as always, is selectional relationships between phrases and phrases, where there is no lexical head to serve as the focus for the facts that the grammar has to record and the learner has to learn. In this section, I explore some facts concerning phrasal adjuncts, to determine how well existing mechanisms work.

As an example of a headed adjunct phrase, P&S-94 give (3).

3. P&S-94, Ch.1(63)

With Kim gone, the project fell apart.

I call *with*, the head of the adjunct phrase, an adjunct preposition. *With* may also select a small clause whose predicate is a PP.

4. With his hat on his head, he left the room.

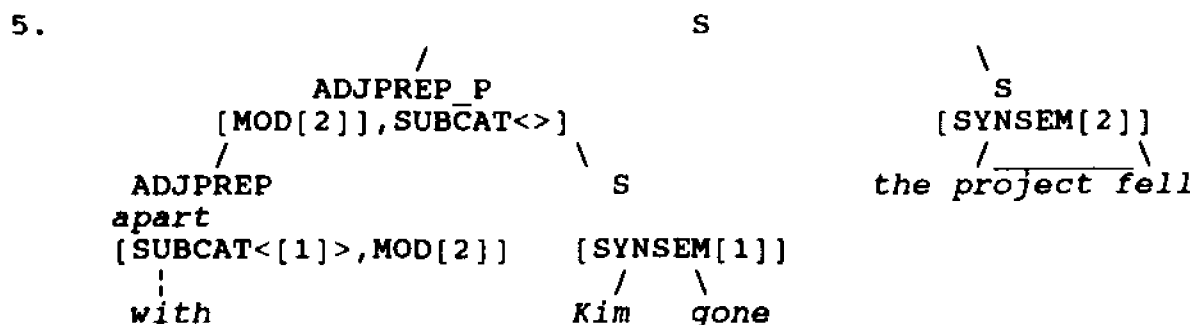
P&S-94 suggest that in sentences such as (3), the adjunct phrase is associated with two selectional features.

Roughly, the idea is to treat *with* as the head of the adjunct, letting it select the small clause *Kim gone* via the SUBCAT feature and the finite clause by the MOD feature.

P&S-94:56

This analysis treats both the selection relation between *with* and its sister within the phrasal adverbial, and the selection relation between the entire phrasal adverbial and the head that it modifies, as lexical.

On P&S-94's analysis, (3) has a structure like (5).



The phrase containing a MOD head, when fully saturated, combines with a head that satisfies the requirements in MOD. So in (5), the adjunct preposition *with* is subcategorized for a small clause, via SUBCAT, and the adjunct phrase with *Kim gone* selects its head, the main clause, via MOD. The value of MOD is structure-shared by *with* and the entire adjunct phrase, just as the value of SELECT in interrogative constructions is structure-shared by the lexical interrogative and the entire interrogative phrase. The value of MOD thus passes through trees in the same way that the values of SUBCAT and SELECT do, as discussed in Chapter 6.

In order to outlaw sentences like (6), P&S-94 conjecture that the requirement that the SUBCAT value of the adjunct phrase be saturated should be added to Schema 5, as shown in (5).<sup>2</sup>

6. \*With, the project fell apart.

On this analysis, *with* is not licensed without some sort of complement, such as a small clause, as in (3), or an NP, as in (7).

7. With good fortune, the project succeeded.

Since P&S-94 treat the selectional facts as lexical, the learning can be lexical, just as it was for the constructions in the previous chapter.

At least two other kinds of adjunct phrases lend themselves to the sort of lexical storage of horizontal select values that follows from P&S-94's analysis of the adjunct preposition *with*. Subordinating adverbs such as *although* and *after* resemble the adjunct preposition *with* in that they select gerunds (*although because* does not).

8. After/Although/\*Because leaving, Kim spoke to the veterinarian.

---

<sup>2</sup>If *with* is the head of the phrase, it is surprising that *with* is sometimes optional, but the small clause itself is not.  
 i. (With) the battle won, the army moved on.  
 P&S-94 choose this analysis, which pushes the theory in the direction of phonologically null elements, for semantic reasons. An alternative analysis, consistent with the approach of P&S-87, is that adverbials like *with* and *although* are markers and therefore non-head constituents that specify properties of the head of Schema 4 by SPEC, rather than heads of Schema 2 that specify properties of a complement by SUBCAT.

Subordinating adverbs are also like *with in* that they may not be intransitive.

9. \*After/\*Although/\*Because, the project fell apart.

But unlike *with*, subordinating adverbials such as *after*, *although*, and *because* select finite clauses, as in (10).

Thus they differ in their SUBCAT values.

10. After/although/because/\*with Kim left, the project fell apart.

Subordinating adverbials also differ from *with in* that they do not select small clauses.

11. a. \*After/\*Because/\*Although Kim gone, the project fell apart.

b. \*After/\*Because/\*Although his hat on his head, he left the room.

The phrases with which subordinating adverbials combine do select special types of clauses at least in some cases, as shown in (12).

12. a. After/Before Leslie left, Kim fed the pandas.

b. \*After/\*Before Leslie left, Kim was feeding the pandas.

c. \*After/\*Before Leslie left, Kim will feed the pandas.

It is not clear whether this selection is syntactic or semantic, but I follow P&S-94 in assuming that the value of MOD is structure-shared with the SYNSEM value of the head that it modifies.<sup>3</sup> Therefore the selectional relations between words like *after* and their sister clauses, and the selectional relations between the adjunct phrases thus

---

<sup>3</sup>An alternative might be that the value of MOD is structure-shared with the CONTENT and/or CONTEXT values of the head that it modifies.

created and the matrix clauses that they modify, can both be stored lexically, via SUBCAT and MOD respectively. The learner will acquire values for them as long as SUBCAT and MOD are universally specified as appropriate for adjunct prepositions and subordinating adverbials (by the UFAP). Then the values of SUBCAT and MOD can be acquired in much the same way that the value of SELECT can for fillers, given appropriate defaults, which I assume are as above. So the examples of adjuncts headed by adjunct prepositions, and of subordinating adverbs show that in many cases, the selectional properties of adjunct phrases can be acquired by lexical learning, without modification of the syntactic theory of P&S-94.

## 7.2 PHRASES WITH NO OVERT LEXICAL SELECTOR

I turn now to some examples where the selecting phrase is not associated with an overt lexical item in whose lexical entry the selectional properties of the phrase can be stored.

### 7.2.1 PARTICIPIAL ADJUNCTS

P&S-94:57 observe that "with-less" predicate adjuncts occur with and without subjects.

13. (=P&S-94,1(65))

- a. His hands trembling violently, Sandy loomed in the doorway.
- b. Trembling violently, Sandy loomed in the doorway.

The sentences in (14) are also grammatical, at least for some speakers of English.

14. a. The battle lost/ended, Sandy started to cry.

- b. The battle having been lost/ended, Sandy started to cry.
- c. Defeated by the enemy, Sandy started to cry.

Since (13.a) and (14.a) are both grammatical with *with*, they might be argued to have a non-overt *with*. But this argument is less plausible for (13.b), (14.b) and (14.c), as shown in (15.b), (16.b) and (16.c).

- 15. a. With his hands trembling violently, Sandy loomed in the doorway.
- b. \*With trembling violently, Sandy loomed in the doorway.
- 16. a. With the battle lost/ended, Sandy started to cry.
- b. ?With the battle having been lost/ended, Sandy started to cry.
- c. \*With defeated by the enemy, Sandy cried.

In the absence of a lexical item with which the MOD feature can be associated in (13) and (14), a possible analysis is that MOD is universally appropriate for participial morphology such as *-ing* and *-ed*, (perhaps because participles are verbal adjectives) and that the value of MOD is stored in the lexical entries for these suffixes.<sup>4</sup> If it were

---

<sup>4</sup>The ablative absolute of classical Latin is a similar phenomenon. Ablative absolutes are adjunct phrases that obligatorily consist of two or more words that are in the ablative case and that agree in number and gender (typically, they consist of a perfect passive participle and an NP subject of the participle).

- i. Caesar, *dē Bellō Gallicō* 1,6,4 (cited in Hale and Buck 1966 par.421)
- Omnibus                      rēbus                      comparātis
- all.ABL.PL.NEUT thing.ABL.PL.NEUTER    prepare.PPP.ABL.PL.NEUT
- diem                              dīcunt
- day.ACC.MASC.SING appoint.3P.PL.PRES
- 'Everything being ready, they set a day.'
- (Lit. 'Everything having been prepared, they set a day.')

Latin adverbial adjuncts such as the ones which may be translated as *with*, *after*, or *since* are never used with ablative

empirically correct that the only phrases that are sensitive to values of their sister contained some (any) morpheme to which the selection might be ascribed, this would be a very plausible alternative. However, since other kinds of phrases appear to be sensitive to the values of their sister nodes, as will now be discussed, a more general solution is evidently needed. Thus other analyses of participial phrases also need to be considered.

### 7.2.2 TOPICALIZED PHRASES

Topicalized constituents, which are non-head daughters of Schema 6, differ from interrogative and relative fillers, and from markers, in that they contain no special word in whose lexical entry selectional properties could be stored; they are unlike participials in that they are characterized by no morphological form with which their selectional properties could be associated in the lexicon. Topicalized phrases are thus selectors that do not have an overt descendant in whose lexical entry the SELECT value can be

---

#### absolutes.

The Ablative Absolute is (like the English Nominative Absolute, as in "this having been done") strictly a neutral construction, telling nothing about the real relation between the facts stated in it and the facts stated in the rest of the sentence. In English, however, we must ordinarily translate so as to show these relations. (*italics in original*)

(Hale and Buck 1903, par.421)

The selectional properties of Latin ablative absolutes might also be stored in the lexical entry of the morphological suffix of the perfect passive participle.

stored.<sup>5</sup> Under these circumstances, it might be expected that topicalized constituents would not select, but that is not the case. In English, topicalized constituents select non-inverted clauses only.

17. a. That book I will give to John.  
b. \*That book will I give to John.

In Spanish, by contrast, inversion is obligatory with topicalized constituents, as in (18).<sup>6</sup>

18. Torrego 1984 (24) (Cf. Ch.6(66 and 67))  
a. Un viaje a las Canarias hizo Antonio este verano.  
A trip to the Canaries made Antonio last summer  
b. \*Un viaje a las Canarias Antonio hizo este verano.  
A trip to the Canaries Antonio made last summer  
'A trip to the Canary Islands Anthony made last summer.'

The fact that the selectional properties of topicalized constituents may vary from language to language proves that grammars must include language-particular specification

---

<sup>5</sup>In principle, it is possible to preserve the hypothesis that selectional properties are stored lexically if any lexical item that happens to occur in a topicalized phrase is allowed to store whatever selectional properties are relevant. However, this analysis misses the linguistic generalization that topicalized phrases select a type of sister node that most phrases with *wh*-descendants do not. Moreover, particular nouns do not appear to have different selectional properties, e.g. an NP head by the noun *person* does not select a different type of clause than an NP headed by noun *mechanism*. So although storing topicalized SELECT values is technically possible, it is not linguistically satisfactory.

<sup>6</sup>In Spanish all verbs can be inverted, not just auxiliaries. Torrego (1984 fn.19) says that "frontings" such as (18) are not necessarily the Spanish equivalent of topicalization, but instead might be consider "*wh*-focus" constructions. Nevertheless, these constructions raise the same sorts of issues as topicalization does, because language-particular information about the selectional properties of *wh*-focus constructions must be stored in the grammar of Spanish and therefore must be learnable.

about the kinds of clauses which may undergo topicalization, and about which properties of their sisters the topics specify.

However, it is not clear from these examples whether selectional properties of topicalized clauses also vary within languages, depending on some feature of the filler phrase. If they do not, the selectional properties of topicalized phrases might be shown to follow from some other property of the language; a single parameter might be sufficient. However, if the selectional requirements of topicalized phrases (or other such phrases) do vary within a language, learners must be able to acquire the selectional properties of phrases. Note that it would take only one example of differing selection by topicalized phrases to show that these selectional properties must be registered somewhere in the grammar, and acquirable by learners. If the limited distribution of topicalized phrases in embedded clauses is indeed evidence that they are generally ungrammatical in English, as suggested by Koster (see Chapter 4), then topicalization in English would constitute proof that the selectional properties of topicalized phrases vary within languages, i.e. they select [ROOT+, INV-] but not [ROOT-, INV-]. But even if topicalization is indeed grammatical in embedded clauses, as is sometimes claimed, at least some topicalized fillers that they are not grammatical with infinitival clauses or VPs.

19. a. Who VP[to talk to \_\_\_] is obvious.  
 b. \*Kim VP[to talk to \_\_\_] is essential.  
 c. \*A solution [to demand \_\_\_] is essential.<sup>7</sup>

So topicalized phrases select S[VFORM fin] but not VP[VFORM inf]. It appears, then, that there is irreducible phrase-phrase selection in natural language. This is, of course, a linguistic issue rather than a learnability one *per se*, and more work is obviously needed to establish exactly what features this selection is sensitive to, and whether there could be other ways entirely of describing it, using existing formal mechanisms of HPSG, or extending them in some fashion.

Without attempting to decide the issue here, I simply note that some suitable descriptive mechanism will be needed if any genuine examples of phrasal selection (without a lexical constituent as a plausible source of the selection) occur in any language. Three types of solutions are in principle available: i) selectional properties of topicalized phrases are stored in the syntax, as language-particular syntactic rules; ii) selectional properties of topicalized phrases are stored in the lexicon in association with 'silent words', i.e., phonologically empty lexical items; or, iii) selectional properties of topicalized

---

<sup>7</sup>It is not clear the filler in (i) is a topicalized phrase or a relative phrase.

i. The person [to talk to \_\_\_] is obvious.  
 If it is a topicalized phrase, the case for phrasal selection by topicalized fillers is strengthened.

phrases are stored in an 'extended lexicon', in association with certain marked features. I discuss these in turn.

i) **SYNTACTIC SOLUTION:** Grammars could be permitted to contain language-particular schemas which must unify with the six universal schemas, in order to "further-specify" them. These language-particular schemas are syntactic, in that (like the universal schemas) they establish feature values on nodes but do not mention any individual lexical items.

On this analysis, at least some language-specific facts are irreducibly syntactic. So some learning is syntactic, and without negative evidence, this is possible only if UG provides suitable defaults, i.e. syntactic defaults. It would be urgent then to discover whether all syntactic defaults suffer from the transderivational problems that plagued GPSG's syntactic defaults. It might be possible to analyze even syntactic defaults as local in that they are confined strictly to local trees, and therefore not subject to recursion; only strict (i.e. falsifiable values, as defined in Chapter 4), may be passed out of the local tree by structure-sharing. However, this would call for many other adjustments, which I am unable to explore here. Without some such reconciliation, the syntactic solution to phrasal selection is not compatible with HPSG, or with the learning model proposed in this dissertation.

ii) **SILENT WORD SOLUTION:** It is imaginable that phrases with selectional properties that do not seem to have a lexical descendant for which MOD or SELECT is appropriate, do nevertheless have a descendant which is a phonologically empty word which has a genuine lexical entry. For example, topicalized constituents might be argued to have as a daughter a focus or topic word that has no phonological content. This approach is in keeping with the general theoretical assumptions of P&S-94, who assume the existence of phonologically empty words that have lexical entries, including trace (P&S-94 Ch.4(15)) and a phonetically null complementizer that specifies infinitival VPs (P&S-94:126-127). It is also consistent with the hypothesis investigated in this dissertation, that all properties that differ from language to language are stored in the lexicon. Therefore, the 'silent word' solution is consistent with the lexical learning model proposed in this dissertation, and with the hypothesis that all defaults are strictly lexical. How learners acquire facts about phonologically null elements is an important question that needs to be investigated.

iii) **EXTENDED LEXICON SOLUTION:** A third possibility is that not all selection facts that are stored in the lexicon are strictly lexical in the sense that they are associated with a word or morphological item. On this analysis, a (presumably) restricted amount of information is stored in

the lexicon in association with certain marked features. For example, the marked value of a feature like FOCUS might be stored in the lexicon in association with a SELECT value. With this solution, what I have been calling the lexicon would actually be an 'extended lexicon' which would be a repository not only for strictly word- or morpheme-specific information, but also for a restricted kind of information about language-particular feature co-occurrence.<sup>8</sup> This analysis involves a blurring of the strict distinction between lexical information and syntactic information, in that information which is arguably syntactic would be stored in the extended lexicon. Learning would therefore not be strictly lexical; however, defaults would be restricted to the extended lexicon, and therefore would be strictly local, as in the learning model presented in this dissertation.

I conclude that it remains to be determined whether all selectional restrictions are lexical, in either a strict or extended sense. Even if it proves impossible to maintain the strong lexical hypothesis under study in this dissertation because of its computational advantages, it seems that there may be some modest extension of the strong hypothesis which would make place in the grammar for linguistic facts that differ across languages, in such a way

---

<sup>8</sup>GPSG had FCRs (Feature Co-occurrence Restrictions) to encode such information. Fodor (1992) showed that language-specific FCRs are not learnable. HPSG replaced FCRs with typed feature structures, as discussed in Ch.2.

that the defaults essential for acquiring these facts can be included in UG without making the grammar computationally intractable.

### 7.3 TOWARD A THEORY OF SELECTING CATEGORIES

According to the analysis I have suggested, there are four selecting features, SUBCAT, SPEC, and MOD, as in P&S-94, and SELECT. In this section, I compare the linguistic similarities and differences between the selecting features, as a step toward the development of a unified theory of selecting features, and I consider whether it is really the selectional features that differ, or merely that they are associated with different categories.

As a result of the modification of the theory of P&S-94 by the addition of the feature SELECT, which is appropriate for filler daughters of Schema 6, each schema is now associated with a feature that stores selectional properties. Configurations of heads and complements are licensed by Schemas 1, 2 and 3, and all language-particular further specifications of these schemas are stored lexically in SUBCAT. Allowable configurations of selectors and heads are licensed by Schemas 4, 5 and 6 and further specifications of these schemas are stored in SPEC, SELECT and MOD. Markers, adjuncts and fillers thus resemble heads in that they may specify syntactic and semantic properties of their sister node. Moreover, the value of each selecting feature is structure-shared with the SYNSEM value of a sister node.

I have shown above that these selectional properties are sufficiently similar that their properties can be acquired by a uniform learning process. I therefore consider the differences between these features with a view toward determining whether the selecting features can be combined into one. I conclude that since the selecting features differ in the mechanism by which the value associated with a phrasal node is structure-shared with a lexical daughter, they cannot be identified at this time.

Many of the differences between SUBCAT, SPEC, MOD, and SELECT follow from properties of the signs with which they are associated, and therefore do not preclude identification of the selecting features.

○ Heads, for which SUBCAT is appropriate, are the syntactic and semantic 'bosses' of the phrases that they are part of. Heads may license multiple complements, while the other selecting categories may not.

○ Markers, for which SPEC is appropriate, are non-head daughters of Schema 4.<sup>9</sup> Markers are lexical, in that they never project phrases, and functional, in that their *primary*

---

<sup>9</sup>I set aside the the status of the SPEC value associated with determiners. P&S-94 analyze determiners (including possessive 's), as head categories for which SPEC is appropriate. An alternative analysis is that determiners are non-heads; this raises semantic problems discussed in P&S-94:47ff. P&S-94 also note (Ch.1, fn.63) that there is a question as to which ID schema licenses signs like *Kim's*: either Schema 1 must be modified because 's is not phrasal, or another schema is required.

contribution is to signal the nature of the relationship between two clauses.<sup>10</sup>

○ Adjuncts, for which MOD is appropriate, are non-head daughters of Schema 5. Adjuncts may be lexical or phrasal, and their function is semantic modification of the head.

○ Fillers, for which SELECT is appropriate, are non-head daughters of Schema 6. Fillers may be lexical or phrasal, and they obligatorily select sisters with gaps, whose presence is imposed by Schema 6.

---

<sup>10</sup>As noted in Ch.3, fn.9, I assume that some markers do make at least a logical contribution to the clauses that they are part of.

The distribution of these properties is summarized in (20).

| 20.                         | Schema | Functional cat | Projects phrase | Licenses multiple complements |
|-----------------------------|--------|----------------|-----------------|-------------------------------|
| HEADS--SUBCAT               | 1,2,3  | No             | Yes             | Yes                           |
| MARKERS--SPEC <sup>11</sup> |        |                |                 |                               |
| Complementizer              | 4      | Yes            | No              | No                            |
| <i>whether, if</i>          | 4      | Yes            | No              | No                            |
| ADJUNCTS--MOD               |        |                |                 |                               |
| Relative phrase             | 5      | No             | NA              | No                            |
| Advbial adjunct phrase      | 5      | No             | NA              | No                            |
| FILLERS--SELECT             |        |                |                 |                               |
| Interrogative               | 6      | No             | If head         | No                            |
| Relative                    | 6      | ?              | If head         | No                            |
| Topicalized phr.            | 6      | No             | NA              | No                            |

This chart suggests that the different properties of the selecting features follow from the classes with which they are associated and therefore do not preclude identification of the selecting features. In addition, the selector categories differ in ways that are not correlated with the particular selecting features with which they are associated.

○ Embedded clauses selected by markers and by interrogative fillers may be subcategorized for by a higher

---

<sup>11</sup>P&S treat lexical adverbials such as *with* as heads which select complements via SUBCAT. But if they are analyzed as markers, as suggested in fn.2, they could be added to this section of the chart.

|                           |   |   |    |    |
|---------------------------|---|---|----|----|
| Lexical adverbial adjunct | 4 | ? | No | No |
|---------------------------|---|---|----|----|

lexical item.<sup>12</sup> An embedded finite clause such as *Kim fed the pandas* must thus be selected both by a higher verb and by the marker, as in (i.a) and (20).

21. I think that Kim fed the pandas.

On the other hand, neither adjuncts (including relative clauses) nor a clause with a fronted topicalized phrase are subcategorized for by a higher verb. The fact that interrogatives, for which SELECT is appropriate, contrast with relatives, for which SELECT is also appropriate, shows that the issue of whether a selecting category must be subcategorized for is not correlated with the selecting feature. [MARKING complementizer], [QUE{X}], *whether*, and *if* are special in that they introduce arguments, which are subcategorized for by a higher predicate.

○ Selecting categories make different semantic contributions to the clause that they select. P&S-94:45 take the position that a marker is "'functional' or 'grammatical'", as opposed to substantive, in the sense that its semantic content is purely logical in nature (perhaps even vacuous).<sup>12</sup> On the other hand, adjuncts modify the content of the head

---

<sup>12</sup>P&S-94 speculate as to whether *and* and *if* are markers or adjuncts, but do not take a position. I assume that *whether* and *if* are a kind of marker in sentences like those (i), since they select properties of their head node and select clauses which are subcategorized for by a lexical head.

i. a. I wonder *whether/if* Kim fed the pandas yet.

b. \*I wonder *that* Kim fed the pandas yet.

The use of *if* in (i.a) is distinguished from *if* in (ii), where *if* is a subordinating adverbial, like *because*.

ii. *If/because* Kim fed the pandas, it must be Tuesday.

by providing additional information that must be combined with the semantics of the head. For example, the adjective *red* imposes the semantic restriction [RELN *red*] on the head noun, as discussed in Chapter 2.

But the semantics of the combination of selecting categories and their heads is more complicated than the Semantics Principle suggests. Complementizers could be argued to be semantically empty, but *whether* and *if*, which I have suggested are markers, do contribute semantic information, as noted in Ch.3, fn.8. Adjunct prepositions and subordinating adverbs also contribute information about the relationship between the adjunct and the main clause.<sup>13</sup> Interrogative and relative fillers can contribute information about case, number, and gender, and human/non-human distinctions, but otherwise the LOCAL value of the filler is determined by the subcategorizing verb of the lower clause. P&S-94 propose the Semantics Principle for combining semantic information contributed by heads and complements, and heads and adjuncts (Ch.2(34)). I assume that this principle must be supplemented to handle the subtly different semantic contributions of the various selecting categories.

I conclude that neither the differences between selecting features summarized in (20) nor the differences

---

<sup>13</sup>This relationship could be temporal/aspectual (*before*, *during*), causal (*because*, *since*), involve presupposition (*whether*, *if*), or of some other sort.

with regard to subcategorization and semantics preclude identification of the selecting features. However, there is one important reason for not identifying the features: different principles constrain the way in which the features are passed through trees. Although SUBCAT is associated with heads, it is not a head feature; instead, the Subcategorization Principle ensures that each value of SUBCAT is realized in each sentence structure exactly once. SELECT is not a head feature; instead, the Phrasal SELECT Principle ensures that the value of a phrasal filler is structure-shared with the appropriate lexical daughter, whether it is a head or non-head. However, MOD and SPEC<sup>14</sup> are head features, according to the analysis of P&S-94, so the HFP ensures that the requirement that the value of these features be structure-shared by the mother node and its head daughter.<sup>15</sup> Moreover, complementizers, fillers, and ad-

---

<sup>14</sup>The reason that SPEC is a head feature has nothing to do with markers, which are non-head daughters of Schema 4. SPEC is a head feature because it is also appropriate for determiners, as noted in Ch.3. P&S-94(51-54) treat possessive 's as a determiner which combines with an NP sister in order to form a determiner phrase. Since they also analyze a word such as *Kim's* as a determiner which combines with an N' sister, it is crucially important that the SPEC value of 's be carried up to *Kim's* by the HFP. However, any analysis in which the value of SPEC is structure-shared between a determiner phrase or noun phrase and the determiner itself allows the value of SPEC associated with a determiner to be acquired by lexical learning.

<sup>15</sup>The status of determinersw and adjuncts has been an unstable aspect of phrase structure grammar. If the approach to SPEC and MOD were re-analyzed, the possibility that at least SELECT, SPEC and MOD could be identified would deserve reconsideration.

junctions do not normally co-occur in the same clause. And when they do, there is a typical hierarchy (for example, fillers select complementizers but complementizers do not select fillers). Therefore, I adopt the conservative position that the selecting features are different sorts of selecting features that are alike in that they specify SYNSEM values of their sister node. This position is consistent with the theory of P&S-94, which I have modified only by adding the feature SELECT.

#### 7.4 RECONSIDERING DEFAULTS

I suggested in Chapter 1 that this dissertation could be viewed as a formal experiment designed to investigate Fodor's hypothesis that markedness relations that account for cross-language variation and that capture the asymmetrical feature associations of natural language also explain how learners can acquire correct generalizations from positive evidence in accordance with the Subset Principle. The experiment has been successful in that I have shown that Fodor's general approach to solving learnability problems is viable within HPSG, and that her model of syntactic learning can be adapted to lexical learning. Both the syntactic modifications to P&S-94 that I have proposed, and the learning model that I have developed are consistent with P&S-94's theoretical approach, with the one exception of my claim that defaults are needed to account for successful acquisition of linguistically perspi-

cacious generalizations. I hope that my research has also helped to show that the HPSG framework provides a useful environment for explicitly modelling the process of grammatical learning.

Still, the learning model that I have proposed, while suggestive, accounts for only a small part of the learning required to acquire a grammar successfully. Even on the assumption that all universal linguistic properties are innate, much more language-particular learning remains to be accounted for. Among the 'syntactic' learning problems that can be fruitfully studied within the approach taken in this dissertation are the following:

- Which other apparently syntactic phenomena can plausibly be stored lexically? For example, P&S-94's proposal that markers and adjuncts specify properties of their sister nodes via the lexical features SPEC and MOD made it possible to show that these properties can be represented lexically and acquired by lexical learning. Perhaps other such mechanisms can be discovered.
- Is syntactic learning required, even if only for certain linguistic phenomena? If so, for what kinds of phenomena is it required? If syntactic learning is indeed required, can the model of lexical learning proposed here shed light on ways of constraining syntactic learning so that it can be successful?

- How can learners acquire generalizations about relations between elements that are not in the same local tree, as in binding?
- How can learners acquire selectional properties of fillers in languages that license multiple interrogatives, such as Bulgarian and Polish?
- Can some of the unwieldiness of linguistic generalizations be eliminated by modification of the features used to represent linguistic phenomena? In particular, the syntactic problems that I have considered suggest that the features AUX and INV should be reconsidered.

The learning model which I have developed depends crucially upon defaults, which are controversial in HPSG. I have argued, following Fodor, that defaults are needed in linguistic theory to account for linguistic tendencies that are not universal, and in learnability theory to account for how learners can acquire correct generalizations that neither miss significant generalizations nor license ungrammatical constructions. I have shown the following:

- A learning process in which grammatical generalizations are derived from the fully specific feature structure of input sentences would result in massive redundancy without defaults,
- Defaults can eliminate the need to store in the grammar more than once linguistic information that is universal and therefore predictable.

- Lexical defaults are free from the computational problems that make syntactic defaults intractable in GPSG.
- Lexical defaults can represent information that appeared to require syntactic defaults.

The conclusion that defaults are essential for learning may be unwelcome to HPSG, where defaults have been regarded as linguistically unnecessary and computationally problematic. My hope is that the issues that I have raised in this dissertation will lead to a re-opening of a discussion about defaults and the information that they represent. Linguistic research is needed to discover formal criteria for recognizing default values, to identify defaults that are cross-linguistically valid, to investigate other grammatical phenomena which might require a more powerful mechanism than linguistic defaults, and to try to understand the principles underlying defaults. In addition, there is an urgent need for computational research to determine which kinds of defaults are 'safe' and which are not, whether there is a way of implementing syntactic defaults that is not transderivational or otherwise computationally intractable, and whether there can be discovered another computational mechanism that could do the work of defaults without adding computational complexity.

Since the study of learnability is as yet in its infancy. the appropriate approach is to study a facet of the problem in the hope, as suggested by Pinker (1979), that

all of the pieces ultimately will fit together. It is my hope that the formal experiment that I have conducted in this dissertation sheds light on an important facet of the learnability problem.

## REFERENCES

- Akmajian, Adrian, Susan M. Steele, and Thomas Wasow (1979) "The category AUX in Universal Grammar", *Linguistic Inquiry* 10(1):1-64.
- Andersson, Anders-Börge and Östen Dahl (1974) "Against the Penthouse Principle", *Linguistic Inquiry* 5(3):451-453.
- Aoun, Joseph and Dominique Sportiche (1983) On the formal theory of government, *Linguistic Review* 2:211-236.
- Baker, C.L. (1979) "Syntactic theory and the Projection Problem", *Linguistic Inquiry* 10(4):533-581.
- Berwick, Robert C. (1985) *The Acquisition of Syntactic Knowledge*. MIT Press, Cambridge, MA.
- Borer, Hagit and Kenneth Wexler (1987) "The Maturation of Syntax". In Roeper and Williams, eds.
- Bouma, Gosse (1992) "Feature structures and nonmonotonicity", *Computational Linguistics*, Special Issue on Inheritance I, 18(2):183-203.
- Bresnan, Joan (1970) "On Complementizers: Toward a Syntactic Theory of Complement Types", *Foundations of Language*, 6:297-321.
- Calder, Jo (1994) "Feature-value logics: some limits on the role of defaults". In *Constraints, Language and Computation*, edited by C.J. Rupp, M.A. Rosner, and R. Johnson. London, Academic Press.
- Carlson, Lauri and Krister Linden (1987) "Unification as a grammatical tool", *Nordic Journal of Linguistics* 10:111-136.
- Carpenter, Bob (1993) "Skeptical and credulous unification with applications to templates and inheritance". In *Inheritance, Defaults, and the Lexicon*, edited by Ted Briscoe, Ann Copestake, and Valeria de Paiva, Cambridge University Press, Cambridge, England.
- Carpenter, Bob (1992) *The Logic of Typed Feature Structures*. Cambridge Tracts in Theoretical Computer Science No. 32. Cambridge University Press, Cambridge, England.
- Chomsky, Carol (1969) *Acquisition of Syntax in Children from 5 to 10*. Research Monograph 57, MIT Press, Cambridge, Mass.

- Chomsky, Noam (1993) "A minimalist program for linguistic theory". In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, edited by Kenneth Hale and Samuel J. Keyser, MIT Press, Cambridge, MA.
- Chomsky, Noam (1986) *Barriers*. MIT Press, Cambridge, MA.
- Chomsky, Noam (1985) *Knowledge of Language*. Praeger, N.Y.
- Chomsky, Noam (1981) "Markedness and core grammar". In *Theory of Markedness in Grammar*, edited by Adriana Belletti, L. Brandi, and Luigi Rizzi. Proceedings of the 1979 Glow Conference, Scuola Normale Superiore de Pisa, Pisa, 1981.
- Chomsky, Noam (1965) *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.
- Chomsky, Noam and Morris Halle (1968) *The Sound Patterns of English*, Harper and Row, N.Y.
- Clark, Robin (1990) *Papers on Learnability and Natural Selection*, Technical Reports in Formal and Computational Linguistics No. 1, Département de Linguistique, Université de Genève.
- Comorovski, Ileana (1971) "Multiple *Wh* movement in Romanian", *Linguistic Inquiry* 17(1):171-176.
- Copestake, Ann (1993) "Defaults in lexical representation". In *Default Inheritance in Unification Based Approaches to the Lexicon*, edited by Ted Briscoe, Ann Copestake, and Valeria de Paiva, Cambridge University Press, Cambridge, England.
- Culicover, Peter W. (1976) *Syntax*, Academic Press, N.Y.
- Daelemans, Walter, Koenraad De Smedt, and Gerald Gazdar (1992) "Inheritance in natural language processing", *Computational Linguistics* 18(2):205-218.
- Emonds, Joseph (1976) *A Transformational Approach to English Syntax: Root, Structure Preserving, and Local Transformations*. Academic Press, N.Y.
- Farkas, Donka and Draga Zec (1995) "Agreement and pronominal reference". In *Advances in Roumanian Linguistics*, edited by G. Cinque and G. Giusti, J. Benjamins, Philadelphia, PA.

- Flickinger, Daniel (1987) *Lexical Rules in the Hierarchical Lexicon*. Doctoral dissertation, Stanford University.
- Flickinger, Daniel, Carl Pollard, and Thomas Wasow (1985) "Structure-sharing in lexical representation" in *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*. Morristown, N.J., Association for Computational Linguistics.
- Fodor, Janet D. (1995) "Fewer but better triggers", *CUNYFORUM* 19. Also (in preparation) "Unambiguous Triggers", *Linguistic Inquiry*.
- Fodor, Janet D. (1994a) "How to obey the Subset Principle: binding and locality". In *Binding, Dependencies and Learnability*, edited by Barbara Lust, Gabriella Hermon, and Jaklin Kornfilt.
- Fodor, Janet D. (1994b) "Language acquisition: paradoxes of parameter setting". Talk at University of Delaware.
- Fodor, Janet D. (1992a) "Islands, learnability and the lexicon". In *Island Constraints: Theory, Acquisition and Processing*, edited by Helen Goodluck and Michael Rochemont, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Fodor, Janet D. (1992b) "Learnability of phrase structure grammar". In *Formal Linguistics: Theory and Implementation*, edited by Robert Levine. Oxford University Press, N.Y.
- Fodor, Janet D. (1990) "Parameters and parameter-setting in a phrase structure grammar". In Frazier and de Villiers, eds.
- Fodor, Janet D. (1989a) Principle-based learning. *CUNYForum* 14: 59-67.
- Fodor, Janet D. (1989b) "Learning the periphery". In *Learnability and Linguistic Theory*, edited by R.J. Matthews & W. Demopoulos. Kluwer Academic, Dordrecht.
- Fodor, Janet D. (1985) "The procedural solution to the Projection Problem". Talk delivered at the BU Conference on Language Development.
- Fodor, Janet D. and Crain, Stephen (1990) "Phrase structure parameters", *Linguistics and Philosophy* 13:619-659.
- Fodor, Janet D. and Stephen C. Crain (1987) "Simplicity and generality of rules in language acquisition". In

*Mechanisms of Language Acquisition*, edited by Brian MacWhinney. Lawrence Erlbaum Associates, Hillsdale, N.J.

- Fraser, Norman M. and Richard A. Hudson (1992) "Inheritance in Word Grammar", *Computational Linguistics*, Special issue on inheritance: I, 18(2):133-158.
- Frazier, Lyn and Jill de Villiers (editors) (1990) *Language Processing and Language Acquisition*. Kluwer Academic Publishers, Dordrecht, the Netherlands.
- Freidin, Robert and A. Carlos Quicoli (1989) "Zero-stimulation for parameter setting", *Behavioral and Brain Sciences* 12:338-339.
- Gazdar, Gerald (1981) "On syntactic categories", *Philological Transactions of the Royal Society of London*, B 295:267-283.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag (1985) *Generalized Phrase Structure Grammar*, Harvard University Press, Cambridge, MA.
- Gibson, Edward and Kenneth Wexler (1994) "Triggers", *Linguistic Inquiry* 25(3):407-454.
- Gold, E. Mark (1967) "Language identification in the limit", *Information and Control* 10:447-474.
- Green, Georgia M. (1994) "Modelling grammar growth: universal grammar without innate principles or parameters". Talk presented at HPSG Conference, Copenhagen.
- Green, Georgia M. (1976) "Main clause phenomena in subordinate clauses", *Language* 52(2):382-397.
- Green, Georgia M. and Jerry L. Morgan (to appear) "Auxiliary inversions and the notion 'default specification'", *Journal of Linguistics*, March 1966.
- Greenberg, Joseph H. (1966a) "Some universals of language with particular reference to the order of meaningful elements". In *Universals of Language*, edited by Joseph H. Greenberg, MIT Press, Cambridge, MA.
- Greenberg, Joseph H. (1966b) *Language Universals with Special Reference to Feature Hierarchies*, Mouton, The Hague.

- Grimshaw, Jane (1993) "Minimal projection, heads, and optionality", unpublished ms.
- Grimshaw, Jane and Steven Pinker (1989) "Positive and negative evidence in language acquisition", *Behavioral and Brain Sciences*, 12:341-342.
- Hale, William G. and Carl D. Buck (1903) *A Latin Grammar*. Alabama Linguistic and Philological Series, reprinted 1966. University of Alabama Press, University, Alabama.
- Halle, Morris (1962) "Phonology in generative grammar", *Word* 18:54-72.
- Hawkins, John A. (1983) *Word Order Universals*. Academic Press, N.Y.
- Henry, Alison (1992) "Infinitives in a for-to dialect", *Natural Language & Linguistic Theory* 10(2):279-301.
- Hermon, Gabriella (1992) "Binding Theory and Parameter Setting", *Linguistic Review* 9(2):145-181.
- Hornstein, Norbert and David Lightfoot (1980) Introduction. In *Explanation in Linguistics: The Logical Problem of Language Acquisition*, edited by Norbert Hornstein and David Lightfoot, 9-31. Longman, London.
- Hukari, Thomas E. and Robert D. Levine (1986) "Generalized Phrase Structure Grammar: A review article", *Linguistic Analysis* 16(Nos.3-4).
- Jacobson, Pauline (1987) "Evidence for gaps". In *The Nature of Syntactic Representation*, edited by Pauline Jacobson and Geoffrey K. Pullum, Reidel, Dordrecht, Holland.
- Jakobson, Roman (1966) "Implications of language universals for linguistics", in *Universals of Language*, edited by J.H. Greenberg. MIT Press, Cambridge, MA.
- Kasper, Robert T. (1993) "Typed feature constraint systems: structures and descriptions". In *Feature Formalisms and Linguistic Ambiguity*, edited by Harald Trost, Ellis Horwood, Chichester, UK.
- Kasper, Robert T. and William Rounds (1990) "The logic of unification in grammar", *Linguistics and Philosophy* 13:35-58.
- Kayne, Richard S. (1991) "Romance clitics, verb movement, and PRO", *Linguistic Inquiry* 22(4):647-686.

- Kean, M-L. (1981) "On a theory of markedness: Some general considerations and a case in point", In *Theory of Markedness in Grammar*, edited by Adriana Belletti, L. Brandi, and Luigi Rizzi. Proceedings of the 1979 Glow Conference, Scuola Normale Superiore de Pisa, 559-604.
- Keenan, Edward L. and Bernard Comrie (1977) "Noun phrase accessibility and Universal Grammar", *Linguistic Inquiry* 8(1):63-99.
- Kiparsky, Paul (1973) "'Elsewhere' in phonology". In *A Festschrift for Morris Halle*, edited by Stephen R. Anderson and Paul Kiparsky, 93-106. Holt, Rinehart and Winston, N.Y.
- Klein, Elaine (1990) *The Null-Prep Phenomenon in Second Language Acquisition*, doctoral dissertation, CUNY.
- Koster, Jan (1978) "Conditions, empty nodes, and markedness", *Linguistic Inquiry* 9(4):551-593.
- Kroch, Anthony (1994) "Morphosyntactic Variation". In *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*, edited by K. Beals.
- Lane, George M. (1903). *A Latin Grammar for Schools and Colleges*. Revised edition. Reprinted by AMS Press, N.Y., 1970.
- Lasnik, Howard and Mamoru Saito (1984) "On the nature of proper government", *Linguistic Inquiry* 15(2):235-289.
- Lebeaux, David (1990) "The grammatical nature of the acquisition sequence". In Frazier and de Villiers, eds.
- Lightfoot, David (1989) "The child's trigger experience: Degree-0 learnability", *Behavioral and Brain Sciences* 12:321-332.
- Lust, Barbara, Gabriella Hermon, and Jaklin Kornfilt (1994) *Binding, Dependencies and Learning*, Lawrence Erlbaum Associates, Hillsdale, N.J.
- Marcus, Gary M. (1993) "Negative evidence in language acquisition", *Cognition* 46:53-85.
- Maxfield, Thomas L. (1990) *The Learnability of a Version of Phrase Structure Grammar*, doctoral dissertation, CUNY.

- McCloskey, James (1991) "Clause structure, ellipsis and proper government in Irish", *Lingua* 85:259-302.
- McCloskey, James (1979) *Transformational Syntax and Model Theoretic Semantics*. Reidel Publishing Co., Dordrecht, Holland.
- McDaniel, Dana (1989) "Partial and multiple wh-movement", *Natural Language & Linguistic Theory* 7:565-704.
- Miller, Philip H. and Ivan A. Sag (1993) "French clitic movement without clitics or movement". Talk presented at LSA.
- Moravcsik, Edith and Jessica Wirth (1986) "Markedness: An overview". In *Markedness*, edited by Fred Eckman, Edith Moravcsik, and Jessica Wirth. Plenum Press, N.Y.
- Penner, Zvi and Thomas Bader (ms.) *Issues in the Syntax of Subordination*, unpublished manuscript, University of Berne, Switzerland.
- Pinker, Steven (1989) "Markedness and language development". In *Learnability and Linguistic Theory*, edited by Robert J. Matthews and William Demopoulos, 107-127. Kluwer Academic Publishers, Dordrecht, the Netherlands.
- Pinker, Steven (1984) *Language Learnability and Language Development*, Harvard University Press, Cambridge, MA.
- Pinker, Steven (1979) "Formal models of language learning", *Cognition* 7:217-283.
- Pollard, Carl and Ivan A. Sag (1994) *Head-Driven Phrase Structure Grammar*. CSLI, Stanford, CA.
- Pollard, Carl and Ivan A. Sag (1987) *Information-based Syntax and Semantics*, Volume 1, Fundamentals, CSLI Lecture Notes Number 13. CSLI, Stanford, CA.
- Pullum, Geoffrey K. (1984) "How complex could an agreement system be?", *Proceedings of the First Eastern States Conference on Linguistics*, 78-103, Columbus, Ohio.
- Pullum, Geoffrey K. (1982) "Syncategorematicity and English infinitival to", *Glossa* 16(2):181-215.
- Rivero, Maria Luisa (1994) "Clause Structure and V-movement in the languages of the Balkans", *Natural Language & Linguistic Theory* 12(1):63-120.

- Rizzi, Luigi (1995) "Early null subjects and root null subjects". In Lust, Hermon, and Kornfilt, eds.
- Rizzi, Luigi (1990) *Relativized Minimality*. MIT Press, Cambridge, MA.
- Roeper, Thomas and Jorgen Weissenborn (1990) "How to make parameters work: Comments on Valian". In Frazier and de Villiers, eds.
- Roeper, Thomas and Edwin Williams (1987) Introduction. In *Parameter Setting*, edited by Thomas Roeper and Edwin Williams, Reidel Pub. Co., Dordrecht, The Netherlands.
- Ross, John R. (1973) "The Penthouse Principle and the order of constituents". In *You Take the High Node and I'll Take the Low Node, Papers from the Comparative Syntax Festival: The Differences between Main and Subordinate Clauses*, edited by Claudia W. Corum, T. Cedric Smith-Stark, and Ann Weiser. Chicago Linguistic Society, Chicago, IL.
- Rothstein, Susan D. (1995) "Pleonastics and the interpretation of pronouns", *Linguistic Inquiry* 26(3):499-529.
- Rudin, Catherine (1988) "On multiple questions and multiple WH fronting", *Natural Language & Linguistic Theory* 6:445-501.
- Sag, Ivan A. and Janet D. Fodor (1994) "Extraction without traces". Talk at West Coast Conference on Formal Linguistics.
- Sag, Ivan A., Lauri Karttunen, and Jeffrey Goldberg (1992) "A Lexical Analysis of Icelandic Case". In *Lexical Matters*, edited by Ivan A. Sag and Anna Szabolcsi. CSLI Lecture Notes No. 24, CSLI, Stanford University.
- Sells, Peter (1985) *Lectures on Contemporary Syntactic Theories*, CSLI, Stanford, CA.
- Shieber, Stuart (1986a) *An Introduction to Unification-based Approaches to Grammar*, CSLI, Menlo Park, CA.
- Shieber, Stuart (1986b) "A simple reconstruction of GPSG", Technical Note 384, SRI, Menlo Park, CA. (Also published in *Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, W. Germany, 1986.)

- Shieber, Stuart, Lauri Karttunen, and Fernando C.N. Pereira (1984) *Notes from the Unification Underground: A Compilation of Papers on Unification-based Grammar Formalisms*. Technical Note 327, SRI International, Menlo Park, CA.
- Tomić, Olga M. (1989) *Introduction, Markedness in Synchrony and Diachrony*. Mouton de Gruyter, Berlin.
- Torrego, Esther. (1984) "On inversion in Spanish and some of its effects", *Linguistic Inquiry* 15(1):103-129.
- Travis, Lisa (1989) "Parameters of phrase structure". In *Alternative Conceptions of Phrase Structure Grammar*, edited by Mark Baltin and Anthony Kroch, University of Chicago Press, Chicago.
- Trubetzkoy, N.S. (1939) *Grundzüge der Phonologie*. 1958 edition. Göttingen: Vandenhoeck and Ruprecht. Translated as *Principles of Phonology* by A.M. Baltaxe. University of California Press, Berkeley, 1966.
- Valian, Virginia V. (1990) "Logical and psychological constraints on the acquisition of syntax". In Frazier and de Villiers, eds.
- Warner, Anthony R. (1992) "English auxiliaries: A lexical account in HPSG". Research Paper YLLS/RP 1992-1, University of York, York, England.
- Weinberg, Amy (1990a) "Child grammars -- radically different or more of the same? Comments on de Villiers, Roeper and Vainikka". In Frazier and de Villiers, eds.
- Weinberg, Amy (1990b) "Markedness versus maturation: the case of subject-auxiliary inversion", *Language Acquisition* 1(2):165-194.
- Weissenborn, Jorgen, Helen Goodluck, and Thomas Roeper (1992) "Introduction: Old and new problems in the study of language acquisition". In *Theoretical Issues in Language Acquisition*, edited by Jorgen Weissenborn, Helen Goodluck, and Thomas Roeper, Lawrence Erlbaum Associates, Hillsdale, N.J.
- Wexler, Kenneth and Peter W. Culicover (1980) *Formal Principles of Language Acquisition*. Kluwer, Dordrecht, Holland.
- Wexler, Kenneth and M. Rita Manzini (1987) "Parameters and learnability in binding theory". In Roeper and Williams, eds.

Williams, Edwin S. (1981) "Language acquisition, markedness, and phrase structure". In *Language Acquisition and Linguistic Theory*, edited by S. Tavakolian, MIT Press, Cambridge, MA.

Zaenen, Annie (1983) "On syntactic binding", *Linguistic Inquiry* 14(3):469-504.