Government and Binding
Unification Grammar

Adam Meyers

June 6, 1998
Contents

Preface vii
1 Introduction 1
2 Generalized Feature Structures 19
3 Constituent Relations and A Theory of Phrase Structure 45
4 Predicate Licensing 73
5 Agreement Licensing Relations 117
6 Scope Licensing Relations 151
7 Coindexing, Empty Categories and Structure Sharing 169
8 Command Relations and Binding Theories 203
9 Believe-type Verbs 219
10 Concluding Remarks 229
Bibliography 231
Preface

Government and Binding Unification Grammar (GBUG) is a Feature Structure based model for Principles and Parameters (P & P) theories (Minimlist Theory, Government and Binding Theory, Extended Standard Theory, etc.) based on previous work in other models of grammar using edge-labeled graphs as data structures. These frameworks include versions of: Relational Grammar, Arc Pair Grammar, Stratified Feature Grammar, Head-Driven Phrase Structure Grammar, Generalized Phrase Structure Grammar, Lexical Functional Grammar, Categorial Grammar and Tag Adjoining Grammar. This book outlines the GBUG model and proposes a GBUG-based version of P & P theory. In addition to its theoretical value, this book is intended as an introduction to unification-based formalisms for linguists better acquainted with P & P approaches, as well as an introduction to P & P for linguists better acquainted with edge-labeled-graph-based theories. I propose it be used as a textbook for graduate and advanced undergraduate students.

Compared with other P & P approaches, it is easy to compare GBUG-based P & P analyses with non-P & P analyses, or to cast aspects of non-P & P analyses in P & P terms (and vice versa). I develop parts of linguistic theory which were previously better developed in non-P & P linguistic frameworks. It turns out that these portions of linguistic theory are well-suited to unification-based models. My approach complements recent trends in formalization of linguistic theory and the renewed interest in analyses which cross-over from one linguistic framework to another. I believe that such an approach to P & P has been needed for some time.

My focus is different from previous work formalizing P & P approaches. For example, Stabler 1993 aims to provide a fairly complete formalization of the Barriers (Chomsky 1986a) version of P & P. Stabler’s approach and my own share the common goal of formalizing Prin-
ciples and Parameters theories. However, complete formalization of a
tory (to the extent possible) is central to his work and integration of
H & P with other frameworks is central to my work. Stabler’s first order
logic is more powerful than FS logics, including GBUG. As a result, there
are many GB constraints that Stabler can represent, but which I cannot
as of yet. When GBUG is augmented to include more powerful opera-
tions (perhaps in future work), I would like to represent higher order
constraints on paths of arcs that would cover some of the same ground
as Stabler. But first, I would like to explore the limits of GBUG without
complicating the logic.

This book includes the salient points of my 1994 dissertation A
Unification-based Approach to Government and Binding Theory (Meyers 1994),
as well as subsequent papers on related topics, including Meyers 1995a,
1995b. Some analyses are revised or augmented to account for linguistic
research which I read too late to incorporate into the original work.

I was working on my dissertation during my employment at the IBM
Thomas J. Watson Research Center. David Johnson and I were develop-
ing a parser for Stratified Feature Grammar (SFG), a unification-based
framework for Relational Grammar (RG) type analyses (Johnson et al. 1993).
This research influenced my dissertation immensely. As it turns out, the
goal of representing RG-like multistratal analyses in one Feature Struc-
ture is similar to the goal of representing multi-level (derivational) Prin-
ciples and Parameters (P & P) analyses in one Feature Structure. Some
ideas I borrowed outright from SFG, e.g., defining subsumption and unifi-
cation on arc labels. Other ideas are based extensively on joint work
with David and I try to be explicit about this.

Many people gave me helpful criticism and advise on various drafts
of my dissertation and this book. My thesis advisor Bay Dougherty and
David Johnson stand out as my two biggest influences. Paul Postal,
Ralph Grishman, Antonio Moreno-Sandoval, Dikran Karagueuzian of
CSLI and an anonymous CSLI reviewer also gave me significant com-
ments and criticisms on various drafts. I would also like to thank the fol-
lowing people for helpful comments: Mark Baltin, John Singler, Cather-
ine Mackeod, Bob Fiengo, Janet Dean Fodor, Carmen Piccallo, Leslie
Barrett, Beth Craig, Ruth Reeves, Scott Browne,Josef Fioretta, Roman
Yangarber, Jussi Karlgren and Zujezdana Vizić.

This book or my thesis would not have been possible without the
financial and intellectual support of New York University (my fellow-
ship with Mark Baltin and my current position with Ralph Grishman’s
Proteus Project), IBM’s T. J. Watson Research Center (under the di-
rection of Ezra Black (1989-1990) and David Johnson (1991-1992)) and
Teacher’s College (Lois Bloom’s Child Language Project).
I would also like to thank my close friends and family for their emotional support: Mom, Dad, Beth, Rosie, Happy, Nana, Tom, Mara, Marvin, Will, Muriel, Greg, Aviva, Casey, Omar, Molly, Neal, David, Fernando, Judy, Francesca, Joel, Jed, Eric, Ed and Fang.

My son Joshua (born May, 1996) was also very helpful during the later stages of writing my book. He rarely drudged on my keyboard and rarely added any of his own keystrokes to this publication (less than five typos resulted from his efforts.)

Most of all, I thank Jenny, the love of my life, who pampered me through the process of getting my degree and preparing this manuscript for publication. Without Jenny, this book would not have been possible.
1

Introduction

1.1 Problem Statement and Approach
This book formulates Government and Binding Unification Grammar (GBUG), a Feature Structure (FS) model for Principles and Parameters (P & P) theories including the Extended Standard Theory, Government and Binding Theory and the Minimalist Program. I show that GBUG's version of the FS is superior to the phrase structure tree as a model for representing P & P analyses. GBUG precisely represents fundamental P & P relations including constituent relations (adjunct, specifier, complement), predicate relations (theta, modification, etc.), agreement (abstract case, etc.) and scope relations. In contrast, the tree model assumed in previous P & P approaches can only represent constituent relations. These trees have unlabeled branches and, contrary to many P & P assertions, unlabeled trees are inadequate for representing constituent relations in P & P theories. This fact is not immediately obvious to many P & P linguists because these linguists do not base their claims on a fully specified version of P & P's theory of constituent structure ($X$ theory). In fact, fully specified versions of P & P's $X$ theory are hard to find, as noted in Pulley 1985 and elsewhere.\footnote{Jackendoff 1977 develops a detailed version of $X$ theory which is not widely adopted. The version of $X$ theory described here more closely fits what most P & P linguists seem to assume.}

In GBUG, fundamental P & P relations are modeled as LICENSING RELATIONS.\footnote{My usage of the term licensing should not be confused with the usage of Abney 1987. Abney posits a licensing relation between a head and its adjuncts, similar to the theta relation. Licensing relations in this book are used in a broader sense. Additionally, unlike Abney 1987, I assume that modifiers predicate of their heads, e.g. red is the predicate in the phrase red books.}
Each licensing relation $X(A, B)$ is modeled in an edge-labeled single-rooted directed acyclic graph as a pair of same source arcs, as shown in Figure 1.

\begin{figure}[h]
\centering
\begin{tikzpicture}
\end{tikzpicture}
\caption{Example of a licensing relation $X(A, B)$}
\end{figure}
A, the value of the X-licensor arc represents the licensor of relation X and B, the value of the X-licensee arc represents the licensee of the relation X. For example, if X is a theta relation, then A is the theta assigner and B is the theta recipient. If X is an adjunct relation then A is the head (adjunct licensor) and B is the adjunct (adjunct licensee). Similarly, Figure 1 is a schema for representing the case relation, the complement relation, the specifier relation, the modification relation and other licensing relations. This schema clearly identifies case assigners/assignees, heads and complements/specifiers, modifiers and modificees, etc.

Figure 2 provides a simplified GBUG representation of *The hedgehogs ate potatoes*. GBUG allows a single feature to instantiate more than one licensor feature and/or more than one licensee feature. For example, the feature *Head*\(^3\) represents that its value is the head of the phrase, which is the licensor of all constituent relations, as well as the licensee of the quantification relation, the modification relation, and others. The feature *QUANTIFIER* represents that its value is the licensor of the quantification relation, and the licensee of the specifier relation. Thus, in Figure 2, the relations Quantification(he, hedgehogs) and Specifier(hedgehogs, the) are represented by the paths *SPEC-CASE Head* and *SPEC-CASE QUANTIFIER*. The *INT-THETA* feature represents that its value is the licensee of one internal theta relation and one complement relation; the *EXT-THETA* feature represents that its value is the licensee of one external theta relation and the licensee of one specifier relation; and the *Minimal A - Chain (MAC)* feature repre-

---

\(^3\)The arc labels used in this book are shorter than the ones used in my thesis. The labels *HEAD-PROJ, EXTERNAL-THETA, INTERNAL-THETA, CATEGORY, PHONOLOGY, SPECIFIER-CASE* and *COMPLEMENT-CASE* have been abbreviated to *HEAD, EXT-THETA, INT-THETA, CAT, PHON, SPEC-CASE* and *COMP-CASE.*
sents that its value is simultaneously the licensees of an internal theta relation, a complement relation and a complement-case relation. By defining features as representing multiple licensor/licensee features, linguistic generalizations are incorporated into the grammar and lexicon. For example, internal theta roles are always assigned to complement positions because the internal theta relation is defined as a type of complement relation. COMP-CASE and MAC relations are also defined as types of complement relations.

The licensing relation approach explicitly represents all licensing relations in a FS model. This contrasts with previous phrase structure tree based P & P approaches, which only attempt to represent constituent relations, discussing other licensing relations informally in the text. Nearly all P & P approaches assume that unlabeled phrase structure trees are sufficient for representing constituent structures, and that the identification of constituent licensors (heads) and licensees (complements, specifiers, and adjuncts) follow from the configurations of nodes in a tree. Kayne 1994 makes a serious attempt to back up these claims. However, Kayne’s definitions are not based entirely on phrase structure tree configurations, but also rely on informal relations which are not part of the phrase structure tree. In Chapter 3, I show that a labeled phrase structure tree would be required to adequately represent Kayne’s analyses.

Predicate licensing relations are defined to hold between a predicate licensor \( A \) and a predicate licensee \( B \) if it can be shown that \( A \) anchors or justifies the constituent immediately dominating \( A \) and \( B \), and \( A \) imposes selection and word order restrictions on \( B \). Following Categorial Grammar (CG), Stratified Feature Grammar (SFG) and Keenan 1974, we assume that the predicate licensor (or functor) may be distinct from the head of the phrase. Chapter 4 presents an analysis, based on Meyers 1995a, in which determiners are predicate licensors which anchor noun phrases. This contrasts with the common assume-

\(^4\)In this book, accusative case, a morphological case is defined so that it is distinct from complement-case, a type of abstract case. An NP is a complement-case licensee if it satisfies some set of constraints which may include: satisfaction of an adjacency constraint, having accusative case morphology, and/or other criteria. (See Chapter 5.)

\(^5\)Some authors, such as Higginbotham 1985, connect nodes in their phrase structure trees with additional arcs representing relations other than constituent relations. The resulting data structures are no longer trees, but rather labeled directed acyclic graphs much like the FSs assumed here.

\(^6\)A constituent \( A \) anchors a constituent \( B \), if the lexical entry of \( A \) justifies the existence of \( B \). As discussed further in Chapter 4, phrases may also anchor other phrases.
FIGURE 2  The hedgehogs ate potatoes
tion in P & P approaches in which phrase structure is always projected from the head (See for example Abney 1987.)

This book distinguishes structure sharing analyses in which one phrase is the value of two constituent licensor/licensee features from empty category analyses in which one constituent binds an empty category. A structure sharing analyses, represents that a single constituent plays two independent roles in an analysis. Structure sharing is typically determined by lexical items. An empty category analysis represents that a constituent is coreferential with a phonologically null phrase. Binding of empty categories is constrained similarly to the binding of pronouns or anaphors. Structure sharing analyses are most common in Relational Grammar, Arc-Pair Grammar, Head-Driven Phrase Structure Grammar and most other edge-labeled-graph-based theories, whereas empty category analyses are more common in P & P theories.

In Chapter 7, I argue for structure sharing analyses of certain constructions and empty category analyses of others. I show that structure sharing analyses are preferred for raising, passive and equi structures. In each of these constructions, the lexical entry for a passive, raising or equi verb uniquely identifies one constituent as a part of two distinct phrases. These contrast with cases of anaphor coreference where the anaphor is not uniquely bound to an antecedent, but rather constrained by binding theory. For example, binding theory permits himself to be coreferential with either Bill or Tom in example 1. Similar ambiguities do not arise with NP-traces or obligatorily controlled PRO. In contrast, non-obligatorily controlled PRO cannot be characterized by structure sharing since it can take split antecedents in example 2. I therefore adopt an empty category analysis for non-obligatorily controlled PRO.

(1) Tom offered Bill a picture of himself

(2) Mary asked Betty if it would be appropriate PRO to buy each other dinner

This book makes an effort to identify what is represented by various theoretical devices including licensing relations, empty categories, structure sharing analyses, and coreference relations. These devices are given formal FS-based definitions, so that the theory as a whole can be fully analyzed (compared with other theories, implemented computationally, etc.). It seems to me that this sort of approach to P & P theory has been needed for a long time.

Some current P & P approaches fail to indicate what is represented by each part of their analysis. Chomsky's Minimalist approach (Chomsky 1992) assumes levels of PF for representing the phonetic form of a sentence and LF for representing its logical form. However, very little indication is pro-
vided regarding what is represented by individual branches and nodes in these levels. For example, it is clear that some "constituent positions" in LF are used to represent scope of quantification and that others represent surface word order or "spell-out" positions. However, some theorists ask us to simply accept that many (final) constituent positions are correct based on "general principles". The correctness of such syntactic positions and the principles they exemplify are unfalsifiable. See Chapters 5 and 6 further discussion. GBUG is a model for those P & P theories in which all "syntactic positions" can (should) be theoretically defined.

In Summary, GBUG includes the licensing principle schema, as well as other FS-based mechanisms for representing coreference, word order and linguistic constraints. Some of the more interesting results include: definitions of X categories in terms of constraints on FSs; arguments for eliminating many empty categories in favor of structure sharing analyses; and the elimination of redundancies between types of licensing relations, e.g., defining the Internal-Theta relation as a type of Complement relation. The account of word order and the internal structure of NP stand out as descriptive and explanatory analyses which are compatible with a variety of non-P & P theoretical frameworks.7

The framework and theory presented in this book is intended as a starting point for a version of P & P theory that is sufficiently detailed to be compared with non-P & P theories and to be implemented on a computer or used in other applied areas. I take the position that in the future, applied areas of linguistics will prove to be good testing grounds for theoretical work. Linguistic theories that cannot be modeled explicitly cannot be tested this way.

1.2 A GBUG-based Model of Grammar

This section provides a brief introduction to the model of grammar for those readers who are already familiar with FS formalisms. Other readers may want to skip this section and return to it after Chapter 2, where definitions of GBUG’s FS terminology are provided.

The GBUG model differs from previous P & P models in that many P & P principles thought to be clause-level or phrase-level principles are modeled as parts of lexical entries or constraints on the construction of lexical entries. For example, the assignment of case and theta roles, as well as specification of head, complement, adjunct and specifier positions are all determined by lexical entries. When a phrase unifies with the value of some set of licensor/licensee features, all the associated li-

---

7These accounts borrow many features of previous work in (SFG) with David Johnson.
lensing relations are instantiated immediately. Movement does not occur and there are no distinct levels of representation. This is typical of unification-based and other graph-based linguistic frameworks.

In GBUG, lexical entries, phrases, classes of lexical entries, classes of phrases, some linguistic constraints and other theoretical objects are modeled as FSs or as statements in FS logic.

Lexical items, classes of lexical items, and classes of phrases are organized in a (default) inheritance hierarchy in the grammar. I will refer to this hierarchy as the lexicon, even though it includes representations of both words and phrases. Every item \( X \) in the lexicon can be “looked up” to get all FSs anchored by \( X \) along with the associated word order constraints for each FS. The word order constraints for a FS \( F \) are modeled as a partial order on the surface arcs in \( F \). Thus the lexical entry for the transitive finite form of \( \text{eat} \) includes both the FS in Figure 3 and the word order (or linear precedence) restriction: \( \text{Head} \prec \text{MAC} \). This indicates that the verb \( \text{eat} \) must precede its surface complement. Similarly, the entry for the class of modifier PPs includes both Figure 4 and the word order constraint \( \text{Head} \prec \text{Modifier} \), indicating that modifiers such as \( \text{in England} \) must follow the heads they modify.

The relation between a class and a subclass in the lexicon can usually be modeled as subsumption. For example, Figure 5 subsumes Figure 3 (above).

Representations of phrases can be derived by combining representations of their constituents by UNIFICATION. For example, Figure 8 is derived by unifying Figure 6 with the value of the \( \text{MAC} \) arc in Figure 7. The FS in Figure 8 anchors an instance of Figure 4. In other words, Figure 4 is found by “looking up” Figure 8 in the dictionary (Figure 8 is a locative PP, a type of PP which anchors Figure 4 FSs.). After dictionary lookup, Figure 8 unifies with the value of the \( \text{MODIFIER} \) arc in Figure 4, yielding Figure 9. In the figures below, the arcs which have been unified are represented by dashed lines.

A constraint \( C \) is satisfied by a FS \( F \) if \( C \) describes \( F \). In simple cases, \( C \) describes \( F \) if a FS-based representation of \( C \) SUBSUMES \( F \). For example, Figure 10 represents a constraint which is instantiated as a FS. This constraint is used to define the set of complete noun phrases \( (N) \). The grammar asserts that the node label \( \text{INDEF} \) subsumes the node label \( \text{DEF} \). This FS subsumes all definite and indefinite noun constituents (the good book, a good book, good books, John), but not

---

8 A surface arc is an arc labeled with a licensor or licensee feature which is stipulated to be a surface relation in the grammar.

9 In cases of default inheritance, a special version of subsumption is needed. See Young and Rounds 1996 for a good approach.
Figure 3: Finite Lexical Entry for *eat*
Figure 4 Entry for PP Modifier
Figure 5: Finite Transitive Verb
Figure 8  *in england*
Figure 9: FS anchored by *in England*
noun constituents like *good book* which lack a value for *QUANT-VAL*. Some constraints require FS logic mechanisms which describe FSSs, but for which a FS representation is not appropriate. For these constraints, subsumption is not the appropriate relation. For example, Figure 11 reads, if there is a *THETA* feature whose value is of category *NOM* (nominal), the target of the *THETA* feature must be the beginning of an inverse path consisting of a *CASE* arc, where inverse paths start with the *TARGET* of a “normal” path and end with its source. The “inverse path” is marked with a $^{-1}$ superscript and means “go up one arc”. The notion “inverse path” originates with (Johnson and Moss 1993 and 1994). Defining constraint satisfaction is not quite as simple in cases involving negation, disjunction and quantification. Further details of constraint satisfaction will be provided throughout the book.

1.3 Overview

I define the FS and GBUG-specific concepts and then use these concepts to outline a GBUG-based version of P & P. One consequence of this approach is that Chapters 2 and 3 are filled with definitions. Readers that find these early chapters to be dense should keep in mind that the later chapters are easier reading.

Chapter 2 provides basic definitions of concepts of unification-based formalisms from the literature and shows how GBUG differs. The terms *Feature Structure*, *Subsumption*, and *Unification* are defined in terms of a model in which FSSs are described as directed acyclic graphs. It is explained that GBUG FSSs are actually Generalized (typed) Feature Structures (Johnson and Meyers 1996) because subsumption and unification are defined on non-identical node and arc labels. Examples of various applications of unification, subsumption and constraint satisfac-
tion are provided to illustrate the roles of these relations in parsing, the
structure of the lexicon and other areas.

Chapter 3 defines the role of constituent licensing relations in P & P
theories. In a FS representing a well-formed phrase, each licensor and li-
censee feature is assumed to be a subclass of some constituent licensor or
licensee feature. In feature structure representations of constraints, this
restriction need not hold. For example, a constraint on theta recipients
may constrain the occurrence of the feature THETA, the superclass of
features such as INT-THETA and EXT-THETA licensee features. Al-
though THETA is not a type of constituent licensor or licensee feature,
INT-THETA and EXT-THETA are types of COMPLEMENT and
SPECIFIER features respectively. In defining constituent relations, a
new definition of X theory is devised and the assumption that constituent
relations can be defined without arc (or branch) labels is shown to be
incorrect.

In Chapter 4, predicate licensing relations, including (internal and ex-
ternal) THETA, MODIFICATION and others, are defined in terms of
selection restrictions and word order restrictions imposed by the li-
censor on the licensee, as well as by the identity of the ANCHOR of the
phrase.

Chapter 5 defines agreement licensing relations, including abstract
CASE and N-AGREEMENT (adjective noun agreement or determiner
noun agreement) represent sets of conditions which must be satisfied for
predicate relations to be well-formed. These conditions can include mor-
phological agreement, adjacency constraints, among others. A surface
interpretation of agreement relations based on work in Government and
Binding (GB) Theory is shown to be superior to a logical form inter-
pretation along the lines of the Minimalist Program (MP). The GB approach
is shown to be conceptually more transparent and computationally more
efficient.

Chapter 6 introduces scope licensing relations, the relation between
any element that has a scope (quantifier, negative, etc.) and the phrase
representing the domain of the scope. The relations are crucial to GBUG
representations of logical form. It is shown that GBUG scope relations are
necessarily distinct from GBUG agreement relations. In a GBUG version
of the MP, these two types of relations would be merged. However, such
an approach is not supported by linguistic evidence.

Chapter 7 discusses the differences between structure sharing and
empty category approaches and introduces a mechanism for representing
coindexing, both for pronouns and empty categories. Coindexing in
GBUG is similar to coindexing in Pollard and Sag 1994, except that the
former can represent more than one kind of coindexing including split
antecedence (examples 2 and 3) and the antecedent/gap relation in VP-deletion sentences (e.g., in example 4, there are two separate instances of eating, each of which can have a different coreferent for his).

(3) John told Mary that they would leave together

(4) Tom ate his apple and so did Sam ______

Chapter 8 provides definitions of command and government relations. In contrast with previous P & P approaches (and similar to Arc-Pair Grammar), command relations are defined between arcs rather than between nodes. Command relations like c-command, defined on constituent licensor/licensee arcs, are compared to command relations defined on other sets of licensor/licensee arcs. Predicate command, a relation defined on all predicate licensor/licensee arcs, is shown to have interesting similarities to LFG’s f-command, HPSG’s o-command, and others. It is shown how these similarities can be exploited to compare LFG, HPSG and P & P analyses of binding phenomena.

Chapter 9 discusses believe-type verbs in light of the following GBUG assumptions: (1) all licensing relations are types of constituent relations; and (2) all syntactic positions are defined in terms of observable linguistic phenomena. I review data such as (5) which support the raising to object analysis and argue against exceptional case marking and raising to AGR-O analyses. Following Postal and Pullum 1988, sentences like (5a) show that complements need not be theta marked, since expletives can occur in complement position contrary to assumptions of Government and Binding Theory. The claim that all complements must be theta marked is behind previous objections to the raising to object. In order for the complements of the coordinated verbs in (5b) from Dougherty and Leacock 1993 to be assigned a single constituent structure, a raising to object analysis must be assumed. Under the exceptional case marking and raising to AGR-O analyses, the phrase the doctor in (5b) would have to be the surface (spell-out) object of persuade, but could not be the surface object of expect. Also under these analyses, the doctor would be the surface subject of the complement of expect, but not the surface subject of the complement of persuade. A raising to object analysis allows a single constituent structure to be assigned. The doctor is a surface object of the conjoined verbs and the non-surface subject of the shared complement. These facts and other evidence lead me to adopt a raising to object analysis rather than changing the fundamental GBUG assumption that all licensing relations are types of constituent relations.

(5) a. I can hardly believe it that you agree
    b. John both persuaded and expected the doctor to visit Bill
In Chapter 10, I conclude that the phrase structure tree model is inadequate for P & P frameworks and I advocate GBUG as a more viable model. The first nine chapters of the book support this position.
2

Generalized Feature Structures

2.1 Introduction
This chapter provides a brief overview of unification-based formalization, and reformulates a few key concepts for GBUG. The FS assumed in GBUG is actually an instance of a Generalized Feature Structure (GFS), as defined in Johnson and Meyers 1996, with origins in SFG (Johnson and Moss 1993, 1994 and Johnson et al. 1993). GFS models differ from “ordinary” unification-based models of grammar in that subsumption and unification relations are defined on nonidentical arc and node labels. GFS models resemble “typed” FS models (cf. Carpenter 1992), which allow nonidentical node labels (but not nonidentical arc labels) to enter subsumption and unification relations. Generalized Feature Structure Logic (GFSL) is introduced as a formal means for representing P & P linguistic constraints.

2.2 Feature Structures
Definition 1 A FEATURE STRUCTURE is either: (a) an atomic unit, which cannot be subdivided (ATOMIC FEATURE STRUCTURE); (b) a set of features and their values, each value a FS; or (c) A variable over FSs, notated NIL;

The Feature Structure (FS), defined in Definition 1, is the main data structure assumed in unification-based models of grammar.¹ FSs are viewed as partial descriptions of the linguistic objects they represent.

¹FSs, and variations on FSs, are known by a variety of names. The term feature structure comes from work on PATR-II (Shieber 1986). However, other terms used include functional description (Kay 1979, 1985) functional structure or f-structure in LFG (Bresnan 1982b, 1982a), category in GPSG (Gazdar and Pullum 1982 and Gazdar et al. 1985), term in Definite Clause Grammar (Pereira and Warren 1980), attribute value structure (Johnson 1988), and S-graph in SFG (Johnson and Moss 1993, 1994 and Johnson et al. 1993).
FIGURE 12 Finite Intransitive Lexical Entry for *cat*
Figure 13: Finite Intransitive Lexical Entry for *eat*
FIGURE 14 Finite Intransitive Lexical Entry for cat
Figures 12 and 13\(^2\) both partially describe the lexical entry for the finite intransitive sense of *cat*\(^3\). A FS can also partially describe or SUBSUME another FS. For example, Figure 12 subsumes Figure 13.

Although FSs are often diagrammed as attribute value matrices like Figure 14 (completely equivalent to Figure 13), a graph-based model is usually assumed and graph theory vocabulary is usually employed. As Sheiber notes:

Feature structures can be viewed as rooted, directed, acyclic graph structures (from which the term “dag” is derived as an acronym) whose arcs are labeled with feature names. Each arc points to another such dag or an atomic symbol. Underlying the graph-theoretic view is a twofold rationale. First, graph theory provides a simple and mathematically well-defined vocabulary with which to model the various feature systems of potential linguistic theories. Second, it leads to a coherent framework for investigating potential structure-combining operations.

Such operations on graph structures abound. Notions of unification, generalization, disjunction, negation, overwriting, and other more idiosyncratic operations can all be formally defined. (Sheiber 1986 (pp. 20–21))

For expository reasons, I will draw the graphs directly, rather than using the matrix notation popular in the FS literature.\(^4\) I see two main advantages of the graph notation. First of all, the phrase structure tree assumed by many P & P linguists is a type of DAG. Thus GBUG based P & P analyses drawn as DAGs should be easily recognized as more precise versions of their tree-based counterparts. Secondly, the phenomenon known as structure sharing is most clearly represented graphically. In a STRUCTURE SHARING configuration, multiple features share the same value or multiple arcs share the same target. Structure sharing is also known as MULTI-ATTACHMENT, NODE SHARING or GRAPH REENTRY. DAGs clearly represent that an item that is structure shared occurs only once in a FS but is the value of more than

---

\(^2\)The label \textit{THIRD-SING} should be viewed as an atomic label, not the negation of the label \textit{THIRD-SING}. As discussed below, the fact that only a finite number of features can be the value of \textit{MORPH-AGR} can be used to circumvent certain complexities of negation.

\(^3\)In most P & P frameworks, it is assumed that verbal inflection heads an Inflection (INFL) phrase and VP is the complement of the head INFL. These entries represent the word formed by combining the inflection with the uninflected or base form.

\(^4\)I will follow the following notational conventions: (1) arc labels will be in italics; (2) node labels will be in bold; and (3) the phonological representations of actual words will be surrounded by a box.
one feature. The numbered indices used to represent structure sharing in the matrix notation are deceptively similar to empty category indices used in P & P analyses.

The representation of linguistic structure with DAGs and other types of edge-labeled graphs originates with work in RG and APG, especially Johnson 1974. Many of the more recent frameworks (e.g., LFG and HPSG) have adopted RG-based and APG-based analyses, including their own versions of the APG (cf. Johnson and Postal 1980) analyses of raising and control phenomena. These accounts use FSs in place of RG and APG graphs. Further, the RG term multi-attachment is equivalent to structure sharing in unification-based accounts.

2.3 Subsumption

Definition 2 A FS A SUBSUMES a FS B (A ⊆ B) if the set of all objects described by B form a subset of the set of all items described by A.

Definition 3 If A and B are labeled nodes (e.g., atomic FSs), then A subsumes B if A and B have the same label.

A DAG A SUBSUMES a DAG B (A ⊆ B) iff:

a. For each arc A_n with source r_A, the root of A, there is a corresponding arc B_n with source r_B, the root of B.

b. The labels of A_n and B_n are the same.

c. The subgraph rooted at the target of A_n subsumes the subgraph rooted at the target of B_n.

Subsumption, defined above, is a relation between two elements X and Y, such that if X SUBSUMES Y (notated X ⊆ Y), X is more general than Y. Subsumption refines the notion 'Y is a subtype of X' as X ⊆ Y. For example, consider a lexicon organized into various classes related by subsumption. Given FSs: FS_v describing any member of the set of lexical entries of type verb FS_i describing any member of the set of lexical entries of type intransitive verb and FS_cat representing the lexical entry for nonfinite intransitive instances of the verb eat, FS_v ⊆ FS_i and FS_i ⊆ FS_cat. Also, FS_v ⊆ FS_cat, since subsumption is a transitive relation.

Figures 15, 16 and 17 represent FS_v, FS_i and FS_cat respectively, where FS_v ⊆ FS_i ⊆ FS_cat.\(^5\) Trivially, all structure in FS_v is included

\(^5\)These figures represent uninflected verbs, in contrast with the entries for finite verbal entries given above. The lexical entry for a finite verb, e.g., Figure 13, can be derived by unifying Figure 17 with the value of the INT-THETA feature in a lexical entry for an INFL.
FIGURE 15  A Verb Entry

FIGURE 16  An Intransitive Verb Entry
FIGURE 17 An Intransitive Lexical Entry for *cat*
in both $FS_i$ and $FS_{ext}$. For every arc $A$ and every node $n$ in $FS_i$, there is an identical arc $A'$ and node $n'$ in $FS_{ext}$. All structure sharing in $FS_i$ is found in $FS_{ext}$ as well. For each graph $G'$ (e.g., $FS_i$), subsumed by a graph $G$ (e.g., $FS_{ext}$), the additional arcs in $G'$ represent additional feature value pairs, narrowing the set of linguistic objects represented by $G'$ over $G$. Shieber 1986 (pp. 14–17) defines subsumption for FSs as follows:

There is a natural lattice structure for feature structures that is based on subsumption—an ordering on feature structures that roughly corresponds to the compatibility and relative specificity of information contained in them... a feature structure $D$ subsumes a feature structure $D'$ (notated $D \sqsubseteq D'$) if $D$ contains a subset of the information in $D'$. More precisely, a complex feature structure $D$ subsumes a complex feature structure $D'$ if and only if $D(l) \sqsubseteq D'(l)$ for all $l \in \text{Dom}(D)$ and $D'(p) = D'(q)$ for all paths $p$ and $q$ such that $D(p) = D(q)$. An atomic feature structure neither subsumes nor is subsumed by a different atomic feature structure. Variables subsume all other feature structures, atomic or complex, because, as the trivial case, they contain no information at all...

Subsumption is only a partial order—that is, not every two feature structures are in a subsumption relation with each other. This can come about because the feature structures have differing but compatible information [He gives examples of two FSs with different values for the feature Agreement—one singular NP and one 3rd person NP. They are not incompatible, but neither subsumes the other]... or because they have conflicting information [e.g., a 3rd person NP and a 1st person NP]...

A path is a sequence of one or more features (or arcs). The value of the path of features (or arcs) is the value of the final feature in the path. In Shieber’s notation, $D$ is a function from paths to their values, so that $D(X)$ is the value of a path $X$. $\text{Dom}(X)$ the set of features in a FS $X$, the domain of $X$. The equal sign is used to represent that its arguments (two features or paths) have the same value (structure sharing). The variable subsumes all other FSs. Following common usage, I represent the variable as $NIL$. Notationally, if a feature labeling an arc $A$ has a value of $NIL$, either the word $NIL$ or nothing at all appears at the target

---

$^6$$NIL$ presumably gets its name because it is a FS containing no information. Atomic labels and feature value pairs constrain the set of objects a FS represents. A FS with no such information is maximally unconstrained, describing any object.
of A. In both cases it should be assumed that the target of A bears the node label NIL.\(^7\)

### 2.4 GFS Subsumption

**Definition 4** Given \( FS_A \) and \( FS_B \), two generalized feature structures such that \( r_A \) is the root node of \( FS_A \) and \( r_B \) is the root node of \( FS_B \), \( FS_A \subseteq FS_B \) iff: \( r_A \subseteq r_B \).

Given two nodes \( n \) and \( n' \), \( n \subseteq n' \), iff: (1) \( \text{label}(n) \subseteq \text{label}(n') \), where \( \text{label} \) is a function from arcs or nodes to their labels; and (2) each arc \( A \) with source \( n \) subsumes an arc \( A' \) with source \( n' \); and (3) For each pair of paths \( P \) and \( Q \), such that \( P = Q \) and \( n \) is the target of both \( P \) and \( Q \), there exist paths \( P' \) and \( Q' \) sharing target \( n' \) such that: \( P' = Q' \) where \( P \subseteq P' \) and \( Q \subseteq Q' \).

Given two arcs \( A \) and \( A' \), \( A \subseteq A' \), iff: (1) \( \text{label}(A) \subseteq \text{label}(A') \); and (2) \( \text{Target}(A) \subseteq \text{Target}(A') \), where \( \text{Target} \) is a function from an arc to its target node.

Given path \( P \), consisting of arcs \( A_1, \ldots, A_n \) and \( P' \), consisting of arcs \( A'_1, \ldots, A'_n \), we say that \( P \subseteq P' \) if \( A_i \subseteq A'_i \) for all values of \( i \), \( 0 \leq i \leq n \).

Subsumption relations are defined between GFS arc labels and between GFS node labels based on partial orders on these labels. The ordinary FS is a special case of a GFS for which label subsumption is determined solely by the identity relation. In this book, partial orders on nonidentical labels are sets of statements of the form \( \text{Label}X \subseteq \text{Label}Y \).

---

\(^7\)Kay 1979 represents the variable as ANY.

\(^8\)In this book, only leaf nodes will be labeled. This suggests that the above clauses should be of the form: (1) or (2) and (3)). The definition as stated, however, can easily be applied to typed feature structures, assuming that types are modeled as node labels.
For example, consider a grammar, **Grammar 1**, which includes the following statements:

**Grammar 1**

\[
\begin{align*}
NOM & \subseteq NOUN \\
NOM & \subseteq GERUND \\
THETA & \subseteq INT-THETA \\
THETA & \subseteq EXT-THETA
\end{align*}
\]

where NOM, NOUN, and GERUND are node labels, and THETA, INT-THETA and EXT-THETA are arc labels. The following relations hold between the FSs in Figures 18 and 19: \( FS_{NOM} \subseteq FS_{NP} \), \( FS_{NOM} \subseteq FS_{GP} \), \( FS_{\theta} \subseteq FS_{INT-\theta} \) and \( FS_{\theta} \subseteq FS_{EXT-\theta} \).

### 2.5 Unification

**Definition 5** \( C \) is the **unification** of two items \( A \) and \( B \) (notated \( A \cup B \)) iff: (1) \( A \subseteq C \); (2) \( B \subseteq C \); and (3) there is no item \( D \) subsumed by \( A \) and \( B \), but not subsumed by \( C \). If no item can be subsumed by both \( A \) and \( B \), then unification fails.

**Grammar 2**

\[
\begin{align*}
\text{COMPLEMENT} & \subseteq \text{INT-THETA} \subseteq \text{MAC} \\
\text{COMPLEMENT} & \subseteq \text{COMP-CASE} \subseteq \text{MAC} \\
\text{SPECIFIER} & \subseteq \text{EXT-THETA}
\end{align*}
\]

---

\(^9\) Alternatively, there can be an internal syntax of arc and/or node labels, which defines a partial order. Meyers 1994 discussed the subsumption of disjunctive labels. For example, \( \text{INT-THETA} \cup \text{EXT-THETA} \subseteq \text{EXT-THETA} \) where the label \( \text{INT-THETA} \cup \text{EXT-THETA} \) is equivalent to the label \( \text{THETA} \) discussed in the text. Johnson and Moss 1993, 1994 and Johnson et al. 1993 define a syntax for “stratified labels” in SFG, such that the the label \([1] \) subsumes the label \([1,0] \).

\(^{10}\) Unification is a partial function from two elements \( A \) and \( B \) to a third element \( C \). The function is partial because unification is undefined (or fails) for pairs of elements that are in incompatible.
Unification of GFSs differs from the unification of "ordinary" FSs because nonidentical pairs of arc labels or node labels can unify. LABEL UNIFICATION (including arc label unification and node label unification) requires a set of statements in the grammar clearly defining which labels subsume which. Given Grammar 1 and Grammar 2, INT-THETA \cup MAC = MAC because MAC (Minimal A-Chain) is the least specific feature which is subsumed by both itself and INT-THETA. Similarly, SPECIFIER \cup EXT-THETA = EXT-THETA. INT-THETA \cup COMP-CASE = MAC given that MAC is the only feature subsumed by both INT-THETA and COMP-CASE. Labels can also unify under identity as in "normal" FS unification. If X = Y then X \subseteq Y and Y \subseteq X. Thus X \cup Y = X = Y. See the previously cited literature on SFG, GBUG and typed feature structure formalisms for more details about label unification.

2.5.1 Examples of GFS Unification

Figure 20 is a type hierarchy for a partial set of English verb types. A FS representing any item \( L \) in Figure 20 is the unification of FSs representing all the superclasses of \( L \). For example, Figure 24 represents transitive verbs as having three independent superclasses: verbs assigning one case (Figure 21), verbs assigning one internal theta role (Figure 22) and verbs assigning one external theta role (Figure 23). The set of transitive verbs is the intersection set of these classes. The FSs representing the superclasses unify together to yield Figure 24, i.e., Figure 21 \cup Figure 22 \cup Figure 23 = Figure 24.

Figure 25 represents the NP \textit{pickled tomatoes}, which unifies with the value of the MAC feature in Figure 26, yielding Figure 27. This may be stated as: Figure 25 \cup FS_{MAC} = FS_{MAC'}, where \( FS_{MAC} \) and \( FS_{MAC'} \) are the values of the feature MAC in Figures 26 and 25. This unification relation represents the instantiation of the licensing relations COMP-CASE, INT-THETA and the constituent relation COMPLEMENT, as all three of these licensee features are subsumed by MAC. Instantiation of these licensing relations are equivalent to case assignment, theta role assignment, and the combination of a head with its complement to form a larger constituent.

\footnote{Let's suppose that some other feature \( L \) distinct from MAC is also subsumed by both INT-THETA and COMP-CASE. If \( L \subseteq MAC \), then \( INT-THETA \cup COMP-CASE = MAC \). If \( MAC \subseteq L \), then \( INT-THETA \cup COMP-CASE = L \).}
\textbf{FIGURE 20} A Type Hierarchy for English Verbs
FIGURE 21  Verb that Assigns One Case
Figure 22 Verb that Assigns one Internal Theta Role
2.5.2 Unification Procedures

The unification of GFSs may be derived by slightly revised versions of previously defined procedures for unifying FSSs.\textsuperscript{12} Procedure 1 is one example of a GFS unification procedure. To derive Figure 24, a GFS unification procedure could first unify Figures 21 and 22, and then unify the result with Figure 23. In the first unification, the algorithm attempts to match the arcs and nodes in Figures 21, and 22. The elements which can be unified are unified (\textit{COMP-CASE} \textcup \textit{INT-THETA} = \textit{MAC}, \textit{VERB} \textcup \textit{VERB} = \textit{VERB}, etc). The remaining (unique) elements of Figures 21 and 22 (e.g., \textit{MORPH-CASE}, the morphological case feature and its value) are also included in the result.\textsuperscript{13} Then the same procedure applies again to incorporate Figure 23. Most arc labels unify under identity, except \textit{SPECIFIER} \textcup \textit{EXT-THETA} = \textit{EXT-THETA}.

\textbf{Procedure 1} \textit{FS}_C, the unification of \textit{FS}_A and \textit{FS}_B is derived as follows:

1. Let \( r_a \) be the root of \textit{FS}_A, \( r_b \) be the root of \textit{FS}_B and \( r_c \) be the root of \textit{FS}_C. \( \text{label}(r_a) \textcup \text{label}(r_b) = \text{label}(r_c) \).

\textsuperscript{12}See, for example, Karttunen 1986 for a simple unification algorithm. See, for example, Godden 1990 for an efficient unification algorithm.

\textsuperscript{13}This intermediary result looks exactly like Figure 24, except that it has a \textit{SPECIFIER} arc in place of the \textit{EXT-THETA} arc.
FIGURE 24  Transitive Verb
2. Let $A_d, \ldots, A_n$ be the set of all arcs $A$ with source $r_a$, $A'_1, \ldots, A'_m$ be the set of all arcs $A'$ with source $r_b$, and let $PAIRS$ be the list of all pairs of arcs $[A_d, A'_e]$ for which $\text{label}(A_d) \sqcup \text{label}(A'_e)$ is defined, where $i \leq d \leq n$ and $j \leq e \leq m$.

3. For each arc $A_d$ such that there is no pair $[A_d, A'_e]$ in $PAIRS$ for any value of $e$, create an arc with source $r_c$ with the same label as $A_d$ and same the value.

4. For each arc $A'_e$ such that there is no pair $[A_d, A'_e]$ in $PAIRS$ for any value of $d$, create an arc with source $r_c$ with the same label as $A'_e$ and same the value.

5. For each pair $[A_d, A'_e]$ in $PAIRS$, create an arc $A''$ with source $r_c$, with label equal to $\text{label}(A_d) \sqcup \text{label}(A'_e)$. The value of $A''$ is equal to the unification of the values of $A_d$ and $A'_e$. If the unification of the values of $A_d$ and $A'_e$ fails, then the procedure fails. The procedure also fails if there are two arcs $A_{d1}$ and $A_{d2}$, both of which unify with $A'_e$. As discussed in Section 2.6.3, this last condition insures that GFS unification has a unique result.
FIGURE 26 Nonfinite Transitive Lexical Entry for *cat*
FIGURE 27  eat pickled tomatoes
2.6 Properties of Generalized Feature Structures

2.6.1 Well-typed Feature Structures

It is assumed that GBUG FSs are well-typed in the sense of Carpenter 1992. This means that each feature in GBUG has a well-defined set of possible values \( v \). \( v \) is either a member of finite set of possible atomic FSs or \( v \) is a complex FS consisting of feature value pairs such that each feature belongs to a finite set of possible features. An advantage of well-typed FSs is that some negations reduce to disjunctions. If the values of a feature are limited in this way, then the negation of any value is equivalent to the disjunction of the remaining possible values. For example, suppose the agreement feature \( PERSON \) can only take the values \( 1ST, 2ND \) or \( THIRD \). Then \( PERSON : \neg THIRD \) is equal to \( PERSON : 1ST|2ND \).\(^{14}\) As a result, satisfaction of negative constraints can be determined more simply. See Nakazawa 1991, p.46 for the formal details of reducing FS negation to disjunction.

2.6.2 Functionality

Well-formed FSs must be functional, following Kay 1979 (p.143), where a FS is FUNCTIONAL iff there is no pair of arcs \( A \) and \( A' \), such that \( A \) and \( A' \) have the same source and the same arc label. Two distinct labels \( L \) and \( L' \) in a GFS can describe the same feature if they are unifiable. Therefore I assume a stricter functionality requirement for GFSs. In a well-formed GFS, two arcs with the same source cannot have labels which can describe the same feature. For example, a GFS \( G \) is ill-formed if \( G \) contains one \( COMP\)-\( CASE \) arc and one \( INT\)-\( THETA \) arc with the same source. Given that \( INT\)-\( THETA \subseteq MAC \) and \( COMP\)-\( CASE \subseteq MAC \), each of these arcs partially describe a MAC arc, and their unification would result in a MAC arc. Without this functionality requirement, it would be possible after a few unifications for a GFS to have two identical labels and thus violate Kay's version of functionality.\(^{15}\)

2.6.3 Unique Unifier

**Condition 1** Given two arcs \( A \) and \( B \) with source \( r \), the root of FS \( F \), and arc \( AB' \) with source \( r' \), the root of FS \( F' \),

\[
\text{If } \text{label}(A) \cup \text{label}(AB') \text{ succeeds and } \text{label}(B) \cup \text{label}(AB') \text{ succeeds,}
\]

\(^{14}\)I handle a few cases like this by introducing "special" labels. For example, I assume the label \( \neg THIRD \), where \( \neg THIRD \subseteq 1ST \) and \( \neg THIRD \subseteq 2ND \).

\(^{15}\)It is assumed that heads universally take at most \( n \) complements. My research on constructing a Syntactic dictionary (COMLEX Syntax: cf. Macleod et al. 1996a) suggests that \( n \) is probably equal to three or four in English. Multiple theta roles (etc.) assigned by the same verb are numbered and different numbers are incompatible, e.g., \( MAC_1 \) will not unify with \( INT\)-\( THETA_2 \).
Then \( F \sqcup F' \) fails.

**Grammar 3**

\[\text{SPECIFIER} \sqsubseteq \text{SPEC-CASE}\]
\[\text{COMPLEMENT} \sqsubseteq \text{COMP-CASE}\]
\[\text{CASE} \sqsubseteq \text{COMP-CASE}\]
\[\text{CASE} \sqsubseteq \text{SPEC-CASE}\]

Figure 28 fails to unify with Figure 29 given Condition 1 and **Grammar 3**. Condition 1 insures that the unification of any two GFSs is unique. Either the **SPECIFIER** arc or the **COMPLEMENT** arc in Figure 28 can unify with the **Case** arc in Figure 29 yielding either a **SPEC-CASE** or a **COMP-CASE** arc. Therefore Condition 1 is violated.\(^{16}\)

\(^{16}\)The unification of two well formed GFSs is guaranteed to be a well-formed GFS. Arcs which violate Condition 1 will not arise as the result of unification. Consider the following proof.

**Given:**

1. arc labels \( a, a', a'', b, b', b'', c, c'' \)
2. \( a \sqcup a' = a'' \)
3. \( b \sqcup b' = b'' \)
4. \( a \sqcup b \) is undefined
5. \( a' \sqcup b' \) is undefined
6. \( a'' \sqcup b'' = c'' \) (is defined)

**Then:**

7. \( a'' \sqsubseteq c'' \)
8. \( b'' \sqsubseteq c'' \)
9. \( a \sqsubseteq a'' \sqsubseteq c'' \) (given 2)
10. \( a' \sqsubseteq a'' \sqsubseteq c'' \) (given 2)
2.7 Logical Constraints

A Feature Structure Logic (FSL) is a logical language used to describe FSs. This section introduces Generalized Feature Structure Logic (GFSL), the FSL used for GBUG. GFSL is formulated in Johnson and Meyers 1996 based on previous work in SFG.\textsuperscript{17} Other previous work on FSL includes: Kasper and Rounds 1986, 1990, Moshier and Rounds 1987, Nakazawa 1991, Carpenter 1992, Moss 1992, Young and Rounds 1993 and Keller 1993.

A GFSL statement $S$ defines the set of FSs described by $S$. If $S$ represents a linguistic constraint, then only members of the set defined by $S$ satisfy that constraint. Conjunctions of path value pairs, such as Figure 30, are simple statements of GFSL which can themselves be modeled as FSs like Figure 31. Figure 30 describes all FSs subsumed by Figure 31. In Chapter 3, GFSL statements define the $X$ theory terminals ($X^0$s), complete phrases ($X$s), and other classes of nonterminals ($\overline{X}$s). $N$s in English are as defined the set of constituents which satisfy Figure 30, where the grammar stipulates that $INDEFINITE \subseteq DEFINITE$.\textsuperscript{18}

This book will use the following GFSL logical devices to formulate linguistic constraints:

The "special atoms" $TRUE$ and $FALSE$, symbols which can be values of paths. $NIL \subseteq TRUE$ and $TRUE$ subsumes all FSs except

\begin{enumerate}
\item $b \subseteq b'' \subseteq c''$ (given 3)
\item $b' \subseteq b'' \subseteq c''$ (given 3)
\item $\alpha \cup b = d''$ (\alpha and $b$ both subsume $d''$, given 9-12. Therefore, they can unify with each other.)
\item $\alpha' \cup b' = d''$ (\alpha' and $b'$ both subsume $d''$, given 9-12. Therefore, they can unify with each other.)
\end{enumerate}

\textbf{A Contradiction:} Lines 13 and 14 contradict lines 4 and 5.

\textbf{Therefore:} Hypothesis (6) ($a'' \cup b'' = c''$) must be false.

\textsuperscript{17}See Johnson and Moss 1993, 1994 and Johnson et al. 1993.

\textsuperscript{18}This account of $N$ is based on unpublished work by David Johnson and myself.
NIL. NIL ⊆ FALSE and FALSE is incompatible with (cannot satisfied by) any FS except for itself and NIL.

The negative operator ¬, the disjunctive operator |, the conjunctive operator ∧, and the implicative operator ⊃. These logical operators may have path value pairs or FSs as arguments.¹⁹

A superscript −1 representing that its argument is an INVERSE PATH. An inverse path P⁻¹ is a (not necessarily unique) path from the target of P to the source of P.

These logical operators provide a precise language for representing most of the linguistic constraints discussed in this book. For example, the statement SPECIFIER : TRUE → HEAD : TRUE is a feature cooccurrence condition (as in GPSG) satisfied by all FSs F with root r such that either r is not the source of any SPECIFIER arc or r is the source of one SPECIFIER arc and one HEAD arc. Figure 32, has the interpretation that the feature THETA can only have a nominal value if there is some CASE feature with the same value as THETA. This version of the case filter states that if you have a THETA arc with a nominal value, then you must be able to “travel up” one CASE arc from the value of the THETA arc. The inverse unary path Case⁻¹ describes a path from the target to the source of an arc labeled CASE. This path begins at the target of the THETA arc. Therefore the nominal is the value of one CASE and one THETA arc. For example, a MAC arc with a nominal value satisfies the case filter because MAC subsumes one type of CASE (COMP-CASE) and one type of THETA (INT-THETA). The path THETA [CASE]⁻¹ describes a path down MAC and then back up MAC.²⁰

¹⁹ PATH : FALSE is equivalent to ¬PATH : TRUE.
²⁰This version of the case filter is based on a different analysis from most current P & P accounts. A-chains are sets of one or more arcs rather than sets of nodes. CASE and THETA arcs share NP values—there are no NP-traces. I adopt essentially Arc-Pair Grammar’s analyses of constructions like passive, raising to object (or exceptional case marking), etc. See Chapter 7 for details.
In the future, GFSL formulas may be expanded to include quantifiers. However, in this book, P & P constraints requiring quantification will be stated in plain English.

2.8 Further Remarks

A language is provided above for modeling P & P concepts and analyses, as well as P & P versions of analyses from other theories. GBUG-based P & P analyses are easy to compare and combine with analyses in non-P & P unification-based frameworks due to the similarity between the GBUG model and the models used in other frameworks, especially SFG and HPSG. There are many P & P principles which would require a richer logic than described here, in particular principles involving quantification. I believe that many of these principles could be added in the future by adding elements of Stabler 1993 first order logic representations to GBUG. For the present, these principles appear in the text (as appropriate) in ordinary English.

The more exotic mechanisms used above derive either from previous work in SFG or are discussed in Johnson and Meyers 1996. FSs, operations on FSs and FSlSs are discussed in a variety of sources. Shieber 1986 provides a good introduction to FSs and Keller 1993 provides a good overview of FSL.
3

Constituent Relations and A Theory of Phrase Structure

3.1 Introduction

This chapter presents a GBUG-based approach to constituent structure in which: (1) constituent relations are represented as pairs of same-source arcs, one constituent licensor arc (a HEAD arc) and one constituent licensee arc (a SPECIFIER, ADJUNCT or COMPLEMENT arc); and (2) levels of projections $X^0, \overline{X}$ and $\overline{\overline{X}}$ denote sets of linguistic entities that satisfy FS logic-based constraints.

GBUG constituents are defined relationally, as in Relational Grammar (RG), rather than structurally as in previous P & P approaches. I show that structural definitions, definitions based on tree configurations, are inadequate for unambiguously defining $\overline{X}$-theory terminology given the full range of P & P analyses and assuming that all analyses must meet $\overline{X}$ constraints. Analyses are excluded in which $\overline{X}$ theory generates some initial structure, which is then transformed into a structure that could not itself be generated by $\overline{X}$-theory. For example, under many P & P approaches, both VP adjuncts and small clause subjects (specifiers) may be the sister of one $\overline{X}$ and the daughter of another $\overline{X}$, the former sometimes being the result of a transformation and the latter being “base generated” by $\overline{X}$-theory based phrase structure rules. Therefore, a phrase occurs in this position cannot be unambiguously identified as an adjunct or a specifier based purely on the tree configuration. The relational approach is unproblematic in this respect.

This theory of phrase structure is more descriptively adequate than previous P & P approaches because, unlike previous approaches, all phrase structures assumed in P & P analyses are accounted for. This theory of phrase structure is more explanatory than previous P & P ap-
approaches due to generalizations between constituent relations and other licensing relations. Most constituent licensing relations are represented using arcs that also represent other relations, e.g., an Int − Theta arc represents that (the constituent represented by the subgraph rooted at) its target is both a complement and an internal theta recipient. Therefore distinct arcs for representing these constituent relations are not needed.

### 3.2 Constituent Licensor and Licensee features as Superclasses

A constituent relation is a licensing relation between the value of the feature HEAD and the value of any licensee feature from the set \{SPECIFIER, COMPLEMENT, ADJUNCT\}. Word order constraints are represented separately, as discussed in Chapter 4.5. For example, the licensing relation Complement(H, C) is represented graphically as a pair of same source arcs: one HEAD arc (the constituent licensor arc) with value H and one COMPLEMENT arc with value C, as shown in Figure 33. The same representation is used whether the head precedes or follows its complement.

Constituent licensor and licensee features sort other licensor and licensee features into four types: HEAD, COMPLEMENT, SPECIFIER and ADJUNCT. Figure 34 is a subsumption hierarchy of the licensor and licensee features that are defined in the course of this book.\(^1\) Each arc-

\(^1\)Figure 34 is a revised version of Figure 7.1 from Meyers (1994). Scope licensor/licensee features and N-Agr licensor/licensee features have been added. Additionally, I no longer list the supertypes THETA and CASE as constituent licensee features because these types of relations are conceptually different. However, given that all subtypes of case features, theta features (and others) are constituent licensee features, the consequences of the difference between the view presented here and the view presented in my thesis is small.
row connects a feature to one of the features it subsumes. For example, \(\text{COMPLEMENT} \subseteq \text{INT-THETA}\) and \(\text{THETA} \subseteq \text{INT-THETA}\). Alternatively, \(\text{INT-THETA}\) is a subtype of \(\text{COMPLEMENT}\) as well as a subtype of \(\text{THETA}\). Table 1 matches each terminal feature in Figure 34 with a list of licensor and licensee features it represents. It is assumed that each licensor and licensee feature in a structurally comparable GBUG representation of a word or phrase is subsumed by some constituent licensor or licensee features, where the term STRUCTURALLY COMPARABLE means “comparable with structural approaches”. At this stage of research, I feel that comparability with structural or phrase structure based approaches is extremely important. So most FS representations of words and phrases used in this book will be structurally comparable. Other supertypes of licensor and licensee features, e.g., \(\text{THETA}\), will be used mainly in linguistic constraints. Of course FSs can describe words or phrases in varying degrees of generality. A FS consisting of a single \(\text{THETA}\) arc and a same source head arc describes a phrase that is unspecified as to whether it is a \(\mathbf{X}\) or an \(\mathbf{X'}\), containing either a specifier or complement respectively. However, such a FS would be difficult to compare with structurally based P & P accounts.
This subsumption hierarchy captures linguistic generalizations and otherwise simplifies GBUG representations of linguistic objects. For example, stating that COMPLEMENT $\subseteq$ INT-THETA represents that internal theta roles must be assigned to complement positions, or alternatively that the INT-THETA feature is a type of COMPLEMENT feature. In contrast, Chomsky's (e.g., Chomsky 1981) projection principle claims not only that all internal theta roles are assigned to complement positions, but also the converse, that all complement positions are assigned theta roles. This latter position does not in fact hold, as shown in Postal and Pullum 1988 and discussed further in Lasnik and Saito 1991 and Chapter 9. Expletives are known not to be theta recipients, yet it is clearly a complement of believe in Example 1. Therefore expletive it is the value of another type of complement feature in sentences like Example 1.

(1) I can hardly believe it that you think so.

Following Johnson and Postal 1980, p. 41, if arc $A$ and arc $A'$ have the same source and target, we say either that $A$ and $A'$ are PARALLEL or that $A$ is PARALLEL to $A'$. If an arc labeled with a feature $F$ must always be parallel to an arc labeled with feature $F'$, $F'$ is redundant. The generalization may be characterized as $F' \subseteq F$. For example, COMPLEMENT $\subseteq$ INT-THETA. Whenever $F (\text{INT–THETA})$ appears in a FS, $F'$ (COMPLEMENT) is completely predictable. If the implication (and therefore the subsumption) goes in both directions, then the two features are equivalent, i.e., if $F \subseteq F'$ and $F' \subseteq F$, then $F = F'$.

Some features in complementary distribution may be instances of the same feature. For example, the grammar above assumes that the feature HEAD is equivalent to a large number of licensor and licensee features that are in complementary distribution. HEAD is the licensor feature of the specifier, complement and adjunct relations, as well as the licensee of several other relations. If many projections of phrase structure are assumed (especially binary branching), then arcs labeled with these features would never have the same source. These features are all subtypes of HEAD, each HEAD arc representing that the FS rooted at its source is a projection of the FS rooted at its target. Since these subtypes are in complementary distribution, there is no reason to distinguish them. Furthermore, if we assume flat analyses of phrase structure in which complements and specifiers are sisters (as Farmer 1984 proposes for languages like Japanese), it would be unprincipled to differentiate between

\footnote{The equal sign, as used here, represents an identity relation. The equal sign is used elsewhere in this book to represent path equivalence (structure sharing).}
a Specifier licensor and a Complement Licensor. In those languages arcs labeled with these features would always be parallel.

Given the subsumption hierarchy in Figure 34, many constituent licensor/licensee arcs may be eliminated from FS representations because they are completely predictable from the otherwise parallel predicate (e.g. theta) licensor/licensee, agreement (e.g. case) licensor/licensee arcs and scope licensor/licensee arcs.\(^3\)

For each licensing relation \(L\) and constituent licensing relation \(C\): (1) if \(C\)-licensor \(\subseteq\) \(L\)-licensor, then \(C\)-licensee \(\subseteq\) \(L\)-licensee (and vice versa); and (2) if \(C\)-licensor \(\not\subseteq\) \(L\)-licensor, then \(C\)-licensee \(\not\subseteq\) \(L\)-licensor (and vice versa).\(^4\) We will assume that there is a partial function \(\text{CONSTITUENT}\), which maps nonconstituent licensing relations to constituent licensing relations according to these specifications. The function \(\text{CONSTITUENT}\) for can be derived from the subsumption hierarchy in Figure 34. Given this function, constituent relations have the purpose of sorting other licensing relations into types. The complement relation looks a lot like the object relation in RG. The specifier and adjunct relations cover some of the same ground as subject and oblique relations in RG.

### 3.3 The Constituent Licensor or Head of a Phrase

Following Meyers 1995a, the constituent licensor can be identified by the following definition:

**Definition 6** The CONSTITUENT LICENSOR of a phrase \(XP\) is the constituent that is obligatory, barring ellipsis, and that bears the features \(F\) that are subject to the subcategorization and selection restrictions imposed on \(XP\). For example, for a phrase \(XP\) to be an internal theta recipient of the verb eat, the constituent licensor of \(XP\) must be a non-abstract noun like fish rather than an abstract noun like idea or a non-noun like if. (Meyers 1995a, Definition 1)

Thus the constituent licensor, or head of the phrase has the function of determining the external distribution of the phrase, as represented by the fact that the head determines the category of the phrase (for example).\(^5\) This definition of head is equivalent to the term base defined in Zwicky 1993. In later chapters, as we discuss the role of other

---

\(^3\)Since FSs can describe phrases in varying degrees of generality, constituent licensor and licensee features may appear in FSs that do not represent predicate, agreement or scope relations.

\(^4\)It is assumed that licensing relations are irreflexive and therefore that (1) and (2) cannot hold for the same instances of \(L\) and \(C\).

\(^5\)See Baltin 1989 for a previous account in which items selection for the heads of phrases. That paper presumably assumes a similar definition of head.
types of constituent licensors in P & P analyses, it will become clear that
Definition 6 may be controversial for some analyses. It will be argued
that one reason some linguists will hesitate to accept Definition 6 is that
they incorrectly assume that the head of the phrase plays the role of ev-
ery type of licensor discussed in this book. A similar point is made in
Zwicky 1985, 1993. It will be maintained that the only necessary func-
tion of the head is to determine the distribution of the nonterminal it
heads and Definition 6 seems to assure this function.

3.4 Constituents and “Bar Levels”
A terminal constituent or word is modeled as a GBUG FS that represents
no constituent licensing relations, e.g., Figure 36. A nonterminal consti-
tuent is modeled as a FS that includes representation of a constituent
relation, e.g., Figure 35 includes one HEAD arc (constituent licensor)
and one QUANTIFIER arc (a type of SPECIFIER licensee arc) repre-
senting that the values of these arcs form a constituent. Constituents
are distinguished by whether they are terminals (X^0) consisting of a
single word, whether they may be independent phrases (maximal projec-
FIGURE 36 Lexical Entry for She

<table>
<thead>
<tr>
<th>Category</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^1$</td>
<td>CAT: TRUE ∧ HEAD: FALSE</td>
</tr>
<tr>
<td>$\overline{X}$</td>
<td>CAT: TRUE ∧ SPECIFIER: FALSE</td>
</tr>
<tr>
<td>$\overline{X}$</td>
<td>CAT: TRUE</td>
</tr>
<tr>
<td>$\overline{V}$</td>
<td>CAT: VERB ∧ SPECIFIER: TRUE</td>
</tr>
<tr>
<td>$\overline{F}$</td>
<td>CAT: VERB ∧ SPECIFIER: TRUE</td>
</tr>
<tr>
<td>$\overline{N}$</td>
<td>CAT: NOM ∧ QUANT-VAL: INDEF</td>
</tr>
</tbody>
</table>

Table 2 $\overline{X}$ Constraints for English

...tions or $\overline{X}$s) and whether they lack specifiers (intermediate projections or $\overline{X}$). $\overline{X}$, $\overline{X}$ and $X^0$ are defined as sets of FSs satisfying linguistic constraints, some of which vary by category type and by language. I assume that $\overline{X}$ constraints are the only ones that vary. For English, I propose the $\overline{X}$ constraints listed in Table 3.4. Given a constituent of English $C$ represented by a FS $F$, $C$ belongs to every category in the left-hand column of Table 3.4 such that $F$ satisfies the corresponding constraint in the right-hand column. If more than one constraint applies, as for $\overline{N}$, $\overline{V}$ and $\overline{F}$, all constraints must be satisfied. These constraints assume that the following subsumption relations hold: NOM \:⊆\: NOUN; NOM \:⊆\: GERUND; and INDEF \:⊆\: DEF (indefinite and definite quantificational properties). Since bar levels are derived by constraints, bar levels are not represented explicitly in GBUG FSs.\(^6\)

\(^6\)The specifiers contained in $\overline{V}$ and $\overline{F}$ are not required to be surface constituents, as defined below.
Figure 36 represents the lexical entry of the pronoun *she*. The value of the CAT feature in Figure 36 is $NOUN$. Figure 36 satisfies the constraints for $X^0$, $\overline{N}$ and $\overline{N}$ in Table 3.4. Therefore *she* belongs to all three of these categories, as do all proper nouns, pronouns, generic nouns, plural nouns and mass nouns. In contrast, Figure 35 is an $\overline{N}$, but not an $\overline{N}$, nor an $N^0$. Figure 37 is an $\overline{N}$, but neither an $N^0$, nor an $\overline{N}$.

$X^0$s must obey the constraint $HEAD(FALSE)$, assuming that the notion $X^0$ is identical to the notion “word”. Since all constituent relations are represented as one $HEAD$ arc and one constituent licensee arc, $HEAD(FALSE)$ effectively rules out all nonterminals.\(^7\)

A universal category-neutral language-neutral GBUG definition of an $\overline{X}$ would only be possible if a single constraint could be found that defines all $\overline{X}$s. Unfortunately, $\overline{X}$s obey constraints that vary by language and by

---

\(^7\)This definition of $X^0$ constrains the possible treatments of verb particle constructions. Under the popular analysis in which the verb and its particle form one complex $V^0$ (the phrasal verb analysis), neither the verb, nor the particle could be the head without violating the $X^0$ constraint. A variant of the phrasal verb analysis is possible in which the verb and particle form a $\overline{V}$ of which the verb is the head. This $\overline{V}$ would take the set of complements for that phrasal verb. Similar considerations may arise with idioms.
part of speech category. The term \( \overline{N} \) therefore does not refer to a class of phrases with common internal structure, but rather a set of phrases which can function in the same way — An \( \overline{N} \) is a constituent that can function as a complete phrase (i.e., maximal projection).

Most \( \overline{N}s \) in English must obey the constraint \( QUANT-VAL(INDEF) \), where \( INDEF \subseteq DEF \). Thus indefinite nominals like *good books* and definite nominals like *the book or she are \( \overline{N}s \), but nominals not marked for definiteness at all like *good book* are not \( \overline{N}s \). This analysis originates in some unpublished work by David Johnson and myself.

The above constraint holds only for \( \overline{N} \) specifiers and complements. Other \( \overline{N}s \) may violate \( QUANT-VAL(INDEF) \), but must satisfy other constraints.\(^8\) \( \overline{N}s \) that modify nouns cannot be plural,\(^9\) cannot contain certain determiners\(^10\) and cannot be pronouns as in (2). (cf. Radford 1988, p. 207) Following Sager 1981 (p. 87) and some unpublished work by Catherine Macleod and me, it is assumed that adverbial \( \overline{N}s \) headed by common nouns must contain a member of a defined class of determiners, adjectives, and/or other modifiers to be complete phrases, as in (3) and (4).\(^11\) For temporal \( \overline{N}s \), the class of non-head items includes: the adjective *last*, the determiners *each, every* and *one*, certain \( \overline{P}s \), the adverb *ago* and relative clauses. For the locative \( \overline{N}s \), the class of non-head items includes: the adverb *apart*, locative \( \overline{P}s \) and the same determiners as with the temporal \( \overline{N}s \). I assume that the non-head items have some special (foot) feature that is shared by dominating projections (just like \( QUANT-VAL \)), e.g., TAG-FEATURE(TMP) for the adjective *last* that only completes temporal \( \overline{N}s \), TAG-FEATURE(LOC) for the adverb *apart* that only completes locative \( \overline{N}s \) and TAG-FEATURE(LOC-or-TMP) for the determiners that complete either \( LOC \subseteq LOC \text{ or } TEMP \) and \( TEMP \subseteq LOC \text{ or } OR \text{ or } TEMP \).\(^12\)

(2) a. The three dolphin show

\(^8\)See Ross 1995 for further categorization of noncanonical \( \overline{N}s \).
\(^9\)Unless the plural (e.g., *the Dolphins*) is taken to refer to a singular entity like a football team.
\(^10\)The allowed determiners include numbers, *all, half, no, each, any* and *every*.
\(^11\)Proper adverbial nouns do not require these non-head items to be complete, e.g., days of the week, place names, etc.
\(^12\)Jackendoff 1973 provides an alternative analysis of examples like (3c) and (4a,c) in which the prepositional phrases and adverbs are heads and the \( \overline{N} \) (our head) is a specifier. The account presented here describes the data below better than Jackendoff’s account.
b. The one dolphin show
c. The all dolphin show
d. The Cecil Albatross show
e. The Mary and Paul show
f. The dolphin show
g. *The the dolphin show
h. *The the dolphins show
i. *The the it show

(3)  a. Last night, I saw you
    b. One night, I saw you
    c. A night ago, I saw you
d. The night of the election, I saw you
e. The night that they were elected, I saw you
f. *A night, I saw you
g. *The night, I saw you

(4)  a. We placed the signs five miles apart
    b. We placed the signs every five miles
c. We placed the signs five miles into the ground
d. *We placed the signs five miles

For some part of speech categories, there is no specific $X$ constraint.\footnote{The minimally defined $X$ constraint in Table 3.4 insures that only FSs with some part of speech category can be $X$s and not some arbitrary FS e.g., the atomic FS INDEF in Figure 35.} Single words of these categories are complete phrases with or without complements, adjuncts or specifiers, e.g., an adjective by itself can be an $Adj$. The requirement that all of a word’s (obligatory) complements be filled is taken care of in the lexicon. The lexical entry for an item $L$ includes $F$, a FS representation of the nonterminal ($\overline{X}$ or $\overline{N}$) that $L$ anchors or justifies. $F$ includes schematic representations of all the complements of $L$. These schematic representations must unify with actual complements for $F$ to be “complete”. Details are given in Johnson et al. 1993 and Chapter 4.

Following current P & P usage, we define $\overline{X}$ as a constituent lacking a specifier. $\overline{X}$ is the constituent that includes the head, all its complements, and possibly one or more of its adjuncts. An $\overline{N}$ plus determiners and/or possessives can form an $\overline{N}$. A $\overline{V}$ plus its (VP-internal) subject forms a $\overline{V}'$. For both nouns and verb phrases, there are some specifiers that can only be the sisters to $\overline{X}$, and others that can be sisters of either $\overline{X}$ or $\overline{N}$. A verb can have only one ext-theta licensee (VP-internal subject). Some
determiners including a and numbers cannot be followed by other noun specifiers. Some intermediate projections contain specifiers, for example, the determiner three in [the [three puppies]] and determiners inside of N “small clauses”, e.g., a in John considered [Mary [a linguist]].\footnote{This informal bracketing is neutral between a raising to object and exceptional case marking account of John considered Mary a linguist. In both these analyses, Mary is an external theta recipient (a type of GBUG specifier) relative to the head a linguist.} This suggests that, as assumed in the Jackendoff 1977 version of X theory, there may be more than one class of intermediate projections relevant to linguistic theory. We propose that each class of intermediate projections can be defined in terms of a linguistic constraint. X is one such class. I propose that other intermediate classes, also represented as linguistic constraints, are idiosyncratically selected for by particular lexical items. For example, a nominal constituent following the, this or these must satisfy the constraint QUANT-VAL : ¬Def. Thus the one book is grammatical, but the this book is not.

Our definitions of $X^0$, $\overline{X}$ and $\overline{\overline{X}}$ differ from previous FS based definitions of these category types. GPSG (Gazdar et al. 1985, p. 22) assumes a FS-based version of X theory for which there is a BAR feature with possible values $\{0, 1, 2\}$. HPSG assumes a set of IMMEDIATE DOMINANCE (ID) schemata. These schemata are like the $\overline{X}$ rules usually assumed in P & P frameworks (see Section 3.7) in that they constrain the possible combinations of constituents in a general way. HPSG’s definitions of $X^0$ and $\overline{X}$ in Pollard and Sag 1994 (p.362) are similar to our own. Pollard and Sag define $\overline{X}$ as a FS whose SUBCAT list (list of subcategorized items) has been exhausted. They assume that nouns like book subcategorize for their determiner. Good book is not an $\overline{X}$ because the FS for good book selects a determiner phrase in its SUBCAT list. Unlike the approach presented here, the Pollard and Sag approach does not seem to account for why some $\overline{N}$s have more than one determiner.

Table 3.4 summarizes our initial defining constraints for $X^0$, $\overline{X}$ and $\overline{\overline{X}}$. It is assumed that bar levels are completely derived from constraints and need not be represented explicitly in GBUG FSs.

$\overline{X}$ has no universal definition. The notion $\overline{X}$ is used synonymously with “maximal projection” and is defined by its function, rather than its internal properties. The set of constituents that can serve as an $\overline{X}$ in a particular language for a particular syntactic category must be defined idiosyncratically. For example, Figure 3.4 lists some of these constraints in English. The lack of any constraints for a particular $\overline{X}$ suggests that...
there is no such constraint, and therefore any constituent of that category
should be viewed as an $\overline{X}$.

3.5 Properties of $\overline{X}$ Grammars

This section discusses GBUG versions of properties of $\overline{X}$ rules originating
is observed that coordinate, and perhaps sentential constituents, should
not be viewed as $\overline{X}$ theory constituents at all.

The LEXICALITY property requires that each nonterminal $N$ has a
unique daughter $H$ called the head, such that $N$ and $H$ have the same
category. Lexicality is encoded in GBUG as the constraint:

$$\text{HEAD} \rightarrow \text{CAT} = \text{HEAD} : \text{CAT}.$$  

According to WEAK SUCCESSION, for each nonterminal $X^n$ that is
the immediate projection of a constituent $X^m$, the bar level of $X^n$
cannot be lower than the bar level of $X^m$ ($X^n \Rightarrow X^m, n \geq m$). This condition
is satisfied by the proposed theory of constituent structure as follows:
(a) $X^0$, as discussed above, must obey the constraint $\text{HEAD} : \text{FALSE}$,
thus $X^0$ cannot be the projection of anything, let alone a higher bar
level projection; (b) $\overline{X}$, the highest bar level constituent, can be the
projection of a constituent of any bar level; (c) $\overline{X}$ constituents obeying
weak succession must obey the constraint:

$$\text{SPECIFIER} : \text{FALSE} \land \text{HEAD} : \text{TRUE} \rightarrow \text{HEAD} : \text{SPECIFIER} : \text{FALSE}$$

This constraint requires that if an $\overline{X}$ is a nonterminal, its head con-
stituent must also be an $\overline{X}$. All $X^0$ constituents are also $\overline{X}$ constituents
according to the definitions above. $\overline{X}$ constituents that lack specifiers
are also $\overline{X}$ and/or $X^0$ constituents by the above definitions.

The UNIFORMITY property requires that every maximum projection
is the same bar level. This requirement is obeyed trivially in GBUG
representations since $\overline{X}$ is the only potential maximal projection by def-
inition.

---

15GBUG's Head Feature Convention (Gazdar et al. 1985 pp. 94-99) requires a set
of constraints like this one for each HEAD feature. A more general way of
capturing the above constraint would be to say that the feature CAT is a member of
the set HEAD - FEATURE and that for each feature $F \in$ HEAD - FEATURE
the constraint holds:

$$\text{HEAD} \rightarrow F = \text{HEAD} : F$$

Other head features (in English) include morphological case and agreement features
(MORPH-CASE and MORPH-AGR),

16The SUCCESSION property assumed in the literature requires that the bar level of
$X^n$ is one greater than the bar level of $X^m$. This weak version of succession combined
with the lexicality property is equivalent to Radford's version of the endocentricity
constraint (Radford 1988, pp. 250-251).
The MAXIMALITY property requires that every non-head daughter of a constituent is a maximal projection. This constraint is assumed above.

Two other \( \lambda \) properties reported in the literature are CENTRALITY and OPTIONALITY, both of which we are rejecting for GBUG \( \lambda \) rules. CENTRALITY is the requirement that the start symbol(s) is a maximal projection. I assume that all maximal projections belong to the set of start symbols, in addition to the sentential node, which is usually viewed as the “start symbol” of a phrase structure grammar for natural languages. It is by no means clear across linguistic frameworks that the sentential node fits the \( \lambda \) schema at all.\(^{17}\) No evidence has ever been provided that sentences conform to \( \lambda \) theory. Therefore I hesitate to ascribe centrality to GBUG’s version of \( \lambda \) theory.

Above, it is assumed that “non-branching” nodes are superfluous.\(^{18}\) Thus optionality, which assumes that all non-head daughters are optional, is not a requirement of the version of phrase structure assumed here. When a non-head constituent does not appear, no new bar level is projected.

Following Pullum 1985 and various work in FS formalisms, I assume that there may be some constituents that cannot be described by \( \lambda \) grammars. For example, coordinate conjunctions raise a host of problems that (I believe) are most easily dealt with by assuming that \( \lambda \) constraints do not apply. Lexicality is violated because the category of a phrase consisting of a coordinate conjunction and a set of coordinate phrases is determined by the set of coordinate phrases. In simple cases, this set shares the same category. In more complex cases, the category of the coordinated phrase must be derived by generalizing over the set. For example, according to Gazdar et al. 1985 (p.175) postcopular phrases like those in (5) are of category [+PRD], a feature that is common to FSs describing all post-copular complements. There is not a single head constituent that determines this category as the lexicality condition would require.

(5)  a. The president was [a good woman, but very cruel]
    b. He seemed [moody and out of his mind]

\(^{17}\) P & P approaches like Stowell 1981 suggest that it is theoretically desirable to assume that sentences conforms to \( \lambda \) theory. Most other linguistic theories (RG, HPSG, etc.) assume that this is not a desirable move.

\(^{18}\) See for example, Baltin 1989 and Chomsky 1994.
3.6 Surface Constituent Relations

It is proposed that \textit{SURFACE} is a set of licensing relations (SURFACE RELATIONS), which in English, includes \textit{DEGREE}, \textit{QUANTIFIER}, \textit{MODIFIER}, \textit{COMP-CASE}, \textit{SPEC-CASE}, and possibly others. Licensor and licensee arcs representing surface relations are SURFACE arcs. Surface arcs are ordered by word order constraints, as discussed in Chapter 4.5. These constraints determine the word order of constituents consisting of surface licensors and licensees (SURFACE CONSTITUENTS). In contrast, word order is irrelevant to nonsurface constituents.

Surface constituents obey constituency tests reported in the linguistics literature, but constituents consisting of licensors and licensees of nonsurface licensing relations may not. (6) to (10) demonstrate the effects of various constituency tests. Adjacent items form a surface constituent if this set of adjacent items act like a unit with respect to: (1) participation in productive alternations like the active/passive alternation in (6); (2) the filling of gaps in various contexts, including VP deletion constructions like (7); (3) pronominal coreference as in (8); (4) ability to serve as answers to questions, e.g., (9); and/or (5) coordination with other constituents as in (10). Elements within a surface constituent $C$ may have nonsurface licensing relations with a constituent outside of $C$. For example in Figure 38, the surface subject (Spec-Case licensee) of \textit{can} is linked to the passive verb by an Int-Theta relation. These two elements do not form a surface constituent, as evidenced by the fact that there are no constituent test that show they form a unit.

\begin{itemize}
  \item[(6)]  \begin{itemize}
    \item a. Sam can eat eggs.
    \item b. Eggs can be eaten by Sam
  \end{itemize}

  \item[(7)]  \begin{itemize}
    \item a. Sam [ate eggs] and so did I  \\
    \item b. Sam [ate eggs] and I did too  \\
    \item c. [Eat eggs] is what Sam did
  \end{itemize}

  \item[(8)]  \begin{itemize}
    \item a. Sam [ate eggs], and I did it, too
    \item b. [The big green monster], ate itself
  \end{itemize}

  \item[(9)]  \begin{itemize}
    \item a. What did Sam do? [Eat eggs]
    \item b. What did Sam eat? [eggs]
  \end{itemize}

  \item[(10)] \begin{itemize}
    \item a. John [ate eggs] and [gave blood] on the same day
    \item b. Do you want to [sleep in a hotel] or [drive all night long]
    \item c. Underdog will either [swallow his secret energy pill] or [be eaten by monsters]
  \end{itemize}
\end{itemize}

In GBUG, there must be at least two distinct Int-Theta relations, such that both Int-Theta licensee features are subsumed by \textit{INT-THETA}. 
FIGURE 38 They can be seen
INT-THETA\(^N\) takes nominal constituent licensees and is a nonsurface relation. INT-THETA\(^A\) takes other licensees and is a surface relation. The INT-THETA\(^N\) licensee feature subsumes MAC (MINIMAL – \(A = \text{CHAIN}\)), a surface feature.\(^{19}\) For simplicity of exposition, I will use the term INT-THETA to represent either INT-THETA\(^N\) or INT-THETA\(^A\), but will make the distinction when it is relevant to the discussion.\(^{20}\)

Given the set of surface constituents, nonsurface constituents can be identified based on theoretical considerations. (6b) can be represented informally as (11), where \(t, t’\) and \(t”\) are informal place-holders for nonsurface positions occupied by the surface subject eggs.\(^{21}\) A GBUG analysis of (11) is given in Figure 38 and the following relations are represented: INT-THETA(V, N), Specifier(V, N) and SPEC-CASE(T, N). The Spec-Case licensor and licensee form a surface constituent, testable by the methods described above (extraction, coordination, etc.). The Int-theta licensor and licensee are assumed to form a constituent on the analogy of finite sentences like (6a). The Specifier licensors and licensees are assumed to form a constituent for a more complex reason. Following Pullum and Wilson 1977 and Stowell 1978, modals and auxiliaries (including passive be) are analyzed as subject to subject raising predicates, sharing their subjects with their complements.\(^{22}\)

When a \(\text{V}\) is attached as the complement of a \(\text{V}\) or \(\text{I}\),\(^{23}\) the specifier of the head is identified with the specifier of its complement. The intermediate specifier licensee arcs (or \(t’\) and \(t”\)) make it possible to identify the Internal-theta licensee of the passive verb eaten with the Spec-case licensee of can.

(11) \([\text{Eggs}], \text{can} t’ t’[\text{V} t’[\text{V} \text{eaten} t’] \text{by Sam}]\)

\(^{19}\)MAC is a surface feature because it is subsumed by the surface feature COMP-CASE (see Figure 34).

\(^{20}\)There are a number of complications that I am ignoring. It appears that predicate nominals are INT-THETA\(^A\) licensees because they do not need case. Predicate nominal complements of verbs like be, seem, consider, etc. cannot be passive subjects and need not be inflected for objective case when they are pronouns. Predicate nominals can also occur in other non-case marked positions. For example, predicate nominals can occur as secondary predicates in He left college [a genius] or prenominally in a biology student.

It is also possible that there is some case-like relation that applies to non-nominal constituents. This would further suggest that INT-THETA\(^A\) \(\subseteq\) MAC or that INT-THETA\(^A\) and MAC are equivalent.

\(^{21}\)t’, t’ are not necessarily intended as P & P traces.

\(^{22}\)Some modals may be subject control predicates as shown in Pullum and Wilson 1977.

\(^{23}\)For the purposes of discussion, we assume that modals belong to the category \(\text{I}\) and passive be belongs to the category \(\text{V}\).
To summarize, surface constituents are subject to constituency tests and word order restrictions, and hence easily identifiable. In contrast, nonsurface constituents are modeled as constituents in order to maintain parallels with surface constituent relations and to align raised (and controlled) subjects. These parallels enable nonsurface relations to be modeled in a constituent structure tree. These parallels also provide a model for sorting licensor and licensee features, as per Section 3.2.

3.7 Previous P & P Versions of X Theory

3.7.1 Introduction

Under the approach presented in this chapter \( X^0 \), \( X \) and \( \overline{X} \) are derived from FSL constraints and \( \textit{SPECIFIER}, \textit{COMPLEMENT} \) and \( \textit{ADJUNCT} \) sort other licensing relations into types. These definitions, and generalizations based on these definitions, rely crucially on a relational account of constituent structure, one in which the head is in a constituent licensing relation with each of its sisters. In contrast, most phrase structure based P & P approaches assume that the terms \( \textit{Head}, \textit{Specifier}, \textit{Complement}, \textit{Adjunct} \), \( X^0 \), \( X \) and \( \overline{X} \) are variables over constituents. In these approaches, each type of constituent is defined in terms of the configurations of phrase structure trees in which they can appear.

To my knowledge, configurational definitions proposed in the P & P literature do not cover all the linguistic analyses assumed by any P & P account. Small clauses, noun phrases with multiple determiners, and phrases with multiple complements are among the phrase types that obfuscate configurational definitions of \( \overline{X} \) terminology. For example, both small clause subjects (specifiers) and adverbial adjuncts of \( V \)'s seem to occupy the position: daughter of \( X \), sister of \( \overline{X} \). Therefore an element in this position cannot be unambiguously identified as an adjunct or specifier simply by looking at its position in a phrase structure tree.

A new type of structurally based definition is proposed in Kayne 1994. Kayne eliminates the specifier/adjunct distinction and takes great pains to enforce the structurally based definitions he proposes. I show that Kayne's adjunct/head distinction rests on information not found in the phrase structure tree and that Kayne's analysis of phrase structure would become well-defined if he admitted arc labels that mark the head of the phrase.

This section demonstrates that adequate configurational definitions have not been formulated within P & P theory. It is suggested that GBUG's relationally based approach to phrase structure is preferred to configurational P & P approaches.
CONSTITUENT RELATIONS AND A THEORY OF PHRASE STRUCTURE / 63

1. \( \overline{X} \Rightarrow X^0 \) COMPLEMENT
2. \( \overline{X} \Rightarrow \overline{X} \) SPECIFIER
3. \( X^n \Rightarrow X^n \) ADJUNCT

FIGURE 39 \( \overline{X} \) Rules from Radford 1988

3.7.2 Chomsky’s \( \overline{X} \) Theory

As noted by Pullum 1985 (among others), specific details are not provided for the versions of \( \overline{X} \) Theory that are most widely assumed in the P & P literature. Chomsky 1970 is usually cited as the source of P & P’s version of \( \overline{X} \) theory and partial lists of \( \overline{X} \) schemata based on Chomsky 1970 are found in various sources. The \( \overline{X} \) rules listed as Figure 39 are adapted from Radford 1988 (pp. 255, 277).24 Most sources only list rules 1 and 2 which seem to originate with Chomsky 1970. Few besides Radford provide \( \overline{X} \) rules for the full variety of adjuncts, although Radford’s characterization of adjuncts is representative of P & P analyses.25

In most P & Pacounts, these are unordered phrase structure rules, word order being derived from independent constraints. COMPLEMENT, SPECIFIER and ADJUNCT are variables over other \( \overline{X} \) level constituents. The optionality of the specifier allows for a non-branching \( \overline{X} \) node. The Kleene star next to the complement variable allows zero or more complements, so that \( \overline{X} \) can either be branching or nonbranching. The adjunct constituent is not optional to avoid vacuous recursion, as Radford notes. Instead the adjunct rule is optional.26 Examples of phrases generated from these rules are provided in (12). The \( \overline{V} \) in (12a) is derived from the rule \( \overline{V} \Rightarrow V^0 \) COMPLEMENT; the \( \overline{T} \) in (12b) is derived from \( \overline{T} \Rightarrow T \) SPECIFIER and the \( \overline{N} \) in (12c) is derived from \( \overline{N} \Rightarrow N \) ADJUNCT.

(12) a. John [\( \overline{V} \) saw [\( \overline{X} \) the man]]
   b. [\( \overline{T} \) [\( \overline{N} \) the man] [\( \overline{T} \) walked home]]
   c. the [\( \overline{N} \) [\( \overline{N} \) very friendly] [\( \overline{N} \) professor of linguistics]]

\(^{24}\)Radford proposes three separate rules for adjuncts, one for each bar level. I combined these three rules into one rule, where \( n \) is a variable over bar levels.

\(^{25}\)The first version of \( \overline{X} \) theory dates at least back to Harris 1951. The Jackendoff 1977 version of \( \overline{X} \) theory was popular in the Extended Standard Theory stage of P & P, but is not representative of current P & P analyses.

\(^{26}\)If the adjunct constituent were optional, then the rule \( \overline{N} \Rightarrow \overline{N} \) would be an instance of the adjunct rule. In a one word \( \overline{X} \) like John, the path from the \( \overline{X} \) node to the \( N^0 \) node could include any number of \( N \) nodes from one to an infinite number. It would be infinitely structurally ambiguous.
Table 3

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Sister-of</th>
<th>Daughter-of</th>
<th>Terminal/Nonterminal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^0$</td>
<td>$\overline{Y}$ or null</td>
<td>$X$</td>
<td>Terminal</td>
</tr>
<tr>
<td>$\overline{X}$</td>
<td>$\overline{Y}$ or null</td>
<td>$X$ or $\overline{X}$</td>
<td>Nonterminal</td>
</tr>
<tr>
<td>$\overline{X}$</td>
<td>$\overline{Y}$ or $\overline{Y}$</td>
<td>$\overline{X}$</td>
<td>Nonterminal</td>
</tr>
<tr>
<td>Complement</td>
<td>$X^0$</td>
<td>$\overline{X}$</td>
<td>Nonterminal</td>
</tr>
<tr>
<td>Specifier</td>
<td>$\overline{X}$</td>
<td>$\overline{X}$</td>
<td>Nonterminal</td>
</tr>
<tr>
<td>Adjunct</td>
<td>$X$</td>
<td>$X$</td>
<td>Nonterminal</td>
</tr>
</tbody>
</table>

Let's assume, counterfactually, that that Figure 39 lists the complete set of schemata for all phrase structure rules assumed in P & P theory. Furthermore, if we assume that $n = 1$ in Rule 3, then adjuncts must be sisters of $X$. Given these assumptions, Table 12 defines all $X$ symbols in Figure 39 configurationally in terms of the relation $SISTER-OF$, the relation $DAUGHTER-OF$, and their status as a terminal/nonterminal. The relation $DAUGHTER-OF(X, Y)$ holds if $Y$ is a constituent structure tree, $X$ is either a subtree of $Y$ or a leaf node of $Y$, and there is a one branch (arc) $A$ from the root of $Y$ to $X$ (if $X$ is a subtree, the target of $A$ is the root of $X$). The relation $SISTER-OF(X, Y)$ holds if $DAUGHTER-OF(X, Z)$ and $DAUGHTER-OF(Y, Z)$. According to the table, $X^0$ is a terminal constituent that must be the daughter of $\overline{X}$, and the sister of an optional non-head maximal projection (the complement); and complements are defined as daughters of $\overline{X}$ and sisters of $X^0$.

Note that adjuncts, complements and specifiers are types of $\overline{X}$s, since these definitions are compatible with the table's definition of $\overline{X}$.

Some linguists might object immediately to the above formulation of structural definitions. These definitions, unlike the original $X$ rules do not generate phrase structure, but constrain it like node admissibility conditions (cf. McCawley 1968a: McCawley cites R. Stanley as the originator of node admissibility conditions). Table 12 requires every part of every analysis to obey $\overline{X}$ theory. For nonderivative theories like the one presented here, this means that every constituent is defined by which constraints it meets, e.g., only those constituents meeting the above definition for specifier are specifiers. For derivational theories, this means that every tree in a derivation must satisfy $\overline{X}$ theory. For example,

---

27 A derivation may be viewed as a sequence of trees $T$ with a partial function $D$ that maps each tree $T^n$ in the sequence to the following tree $T^{n+1}$. $D$ must include a partial mapping from the set of nodes in $T^n$ to the set of nodes in $T^{n+1}$. One can view $T$ and $D$ as a single representation. Similarly, one might view a GBUG FS as a single representation or the unification of several different partial representations. It is conceivable that particular derivational and nonderivational approaches can be shown...
a specifier must meet the definition for specifier both before and after the application of each \( X \) rule, each pruning operation, each adjunction operation, each movement, etc.\(^{28}\)

If \( n \) can equal any bar level in Figure 39, then analyses such as (13) are possible where \( \text{even, green and up} \) occupy adjunct positions. The definition of adjunct becomes [sister-of \( X^n \), daughter-of \( X^n \)]. It would be difficult to account for these constructions in P & P without this modification. However, the structural distinction between \( X \)'s and \( \bar{X} \) is blurred, since both can be in the position [Daughter-of \( \bar{X} \), Sister-of \( \bar{X} \)], e.g., the friendly people in (13a) is an \( \bar{N} \) occupying this position and very friendly professor of linguistics in (12b) is an \( \bar{N} \) occupying this position.

(13) a. \( [\bar{N} \text{ the friendly people}] \) got angry
b. the \( [\bar{N} \text{ big green book about fish}] \)
c. Mary will [\( \bar{V} \) call \( \bar{N} \) John]

Under the most common P & P analyses of small clause constructions (see Stowell 1983) specifiers, like adjuncts, can occur in the position [sister-of \( \bar{X} \), daughter-of \( \bar{X} \)], as in (14). The \( \bar{N} \)'s in (15), which contain multiple determiners each have at least one determiner in a [sister-of \( X^n \), daughter-of \( X^n \)] position, where \( n \) could equal either 1 or 2 depending on one’s analysis. Assuming that \( n = 2 \), one determiner (a specifier) occurs in the structural position [sister-of \( \bar{X} \), daughter-of \( \bar{X} \)]. Assuming \( n = 1 \), at least one determiner occurs in the position [sister-of \( \bar{X} \), daughter-of \( \bar{X} \)]. In particular, note that the adjunct only can occur before, after or between any two determiners. Given these data, a structurally based distinction between adjuncts and specifiers is difficult to maintain.\(^{29}\)

(14) a. I saw [\( \bar{X}_\theta \) John [\( \bar{X}_\theta \) angry]]
b. I consider [\( \bar{N} \) John [\( \bar{N} \) a nice guy]]
c. He made [\( \bar{V} \) John [\( \bar{V} \) leave]]

to be equivalent. This may lead one to suppose that the derivational/nonderivational distinction is a superficial one.

\(^{28}\)A discussion with Bob Fiengo helped me clarify the relation between nonderivational and derivational approaches.

\(^{29}\)In Meyers 1994 I assumed that only is an unusual determiner. However, only (also just) appears to be a focus marker that can modify a wide range of constituent types, as in (i). I assume here that focus markers are types of adjuncts.

(i) a. I was only buying a turnip
b. Only five people
c. Only slightly bigger
d. Only like his mother
(15) a. \[\text{the} \ [N - \text{only} \ [N - \text{three} \ [N \text{- green books}]]] \]
    b. \[\text{only} \ [\text{the} \ [N \text{- three} \ [N \text{- green books}]]] \]
    c. \[\text{the} \ [N \text{- only} \ [N \text{- green book}]] \]

Assuming that all constructions are “binary branching” or that there are no non-branching nodes causes further difficulties for structurally based distinctions. Example (16a) assumes the Chomsky 1975 (p. 246) “binary branching” analysis of multiple complements.\(^\text{30}\) Under the binary branching analysis, complements can be in the position [sister-of \(X\), daughter-of \(X\)], just like adjuncts. If we further assume that there can be no non-branching nodes when complements and specifiers don’t occur (Baltin 1989 and Chomsky 1994), then complements, like specifiers, can occupy the position [sister-of \(X\), daughter-of \(X\)], as shown in (16b), in which a nonbranching \(N\) is pruned from a determiner-less \(N\). The no-non-branching-node assumption also allow adjuncts in the [sister-of \(X\), daughter-of \(X\)] position, as in (17). The syntactic position [sister-of \(X^0\), daughter-of \(X\)], which is only possible once non-branching nodes are eliminated, can be occupied by specifiers, complements or adjuncts, as in (18).

(16) a. \[\text{John} \ [\text{past} \ [\text{t} \ [\text{give} \ [\text{Mary}]] \ [\text{a book}]]] \]
    b. \[\text{[N of gifts [of books]] [from Mary]} \]

(17) \[\text{Complicated [N medical examinations]} \text{ upset most people.} \]

(18) a. \[\text{She completed the [N of examinations]} \]
    b. \[\text{She completed [N of examinations] [of her patients]} \]
    c. \[\text{She completed [N of complicated [N of examinations]} \]

Previous P & P accounts have not specified the details of the version of \(X\) theory they assume as pointed out in Pullum 1985. No P & P principles would have to be sacrificed by adopting GBUG’s constraint-based definitions of \(X\). However, maintaining the phrase-structure-based definitions of \(X\) theory would result in the sacrifice of crucial P & P analyses of small clauses, relative clauses, head modifier constructions, multiple complement constructions among others. I am aware of no structurally based P & P account that maintains an adjunct/complement/specifier distinction while providing analyses for all the examples described in this section. For example, all P & P accounts have difficulty accounting for small clause constructions and adjunct sisters of \(X\) while distinguishing adjunct from specifiers and \(X\)’s from \(X\)’s on structural grounds.

\(^{30}\) See Larson 1988 and Chomsky 1992 for a variant on this approach that is becoming popular in recent P & P literature.
GBUG based account of $\overline{X}$ theory provided in this chapter is superior to all such structural accounts on the basis of descriptive adequacy.

The problems cited above may lead some linguists to adopt analyses that would cause structural definitions to be problematic for fewer examples. Unfortunately, these "solutions" bring other problems in their wake. As long as there are some examples that defy structural definitions, it is unclear why anyone would still assume structural definitions. Under the Abney 1987 account of examples like (15), the phrases that begin with determiners are really determiner phrases, with determiners taking either $\overline{N}$ or $\overline{D}$ complements. This analysis would make (15) unproblematic for structural definitions of specifiers, but would raise serious difficulties for defining the term head, as discussed in Meyers 1995a and Section 4.3 in this book. The discussion of Examples (16) to (18) could lead to the rejection of binary branching and the non-non-branching-node analyses. The considerations that cause linguists to adopt these analyses should not be ignored if the only reason is to preserve structural definitions. Nonbranching nodes serve no descriptive function and are ignored for purposes of (most versions of) c-command, m-command, government, etc. The question of whether to adopt a binary branching analysis of any construction should depend on constituency tests and other theoretical considerations. It seems to me that any reason for maintaining structural definitions of phrase structure terminology is out-weighed by these other theoretical considerations.

3.8 Kayne’s Antisymmetry of Syntax

In the theory of phrase structure proposed in Kayne 1994 (Antisymmetry), each level of phrase structure can have at most one maximal projection.\footnote{Discussion of the Antisymmetry framework will be limited to the way heads, complements and adjuncts are defined.} The nonmaximal projection is the head and the maximal projection is either a specifier/adjunct or a complement, where the terms specifier and adjunct have been collapsed together. A nonterminal node $N$ can only be a maximal projection if $N$ is not the head of a superordinate phrase. In the Antisymmetry approach, each terminal head $h$ is immediately dominated by one nonterminal $H$. $H$ can have exactly one sister constituent, its complement. This complement, must be a maximal projection, a constituent that has at least three projections: the terminal and two nonterminal projections. $H$ and its complement are dominated by a node $H'$ of the same category as the head. Any sister of $H'$ is a specifier (or adjunct). The specifier must be a maximal projection. In Figure 40 (copied from Kayne 1994, p. 16), $R$ is a head projection from
r. $R$ has a complement $S$, which together form a constituent $P (= R')$ with specifier $M$. Kayne also allows adjunction to a head as in Figure 41 (copied from Kayne 1994, p.17). In this case, the head ($M$) has the same category as the parent node and the adjunct has a different category ($Q$).

The following factors are crucial for determining the head/complement/specifier status of two nodes $N$ and $N'$ with mother $M$: (1) Is $N$ the mother of a terminal or nonterminal? (2) Is $N'$ the mother of a terminal or non-terminal? and (3) Is $N$ or $N'$ the head of $M'$? The tree provides the answers to questions (1) and (2), but not the answer to question (3). The Antisymmetry approach indicates the mother daughter relation informally in Figures 40 and 41 by using the same capital letter for mother and daughter nodes. By informal, I mean that the head cannot be identified on the basis of facts that are represented in the tree model. The phrase *university buildings*, represented in Kayne's framework in Fig-
Figure 41 Kayne's Adjunction to Heads
Figure 42, is an instance of noun noun modification containing one head and one adjunct. The adjunct has the same category as the head of the phrase. If Kayne had a structural way of representing the head relation, it would be possible to tell that university is the adjunct. Similarly, the Figure 41 relies on information outside of the phrase structure tree model to distinguish the head from the adjunct.\textsuperscript{32}

A simple solution to this problem is to change the model. If head branches were labeled, then it would be possible to distinguish heads and adjuncts in the Antisymmetry model. However, this is clearly not a structurally-based solution. A structurally based solution that maintains the head/adjunct distinction does not seem possible.\textsuperscript{33}

3.9 Conclusions

GBUG

\textsuperscript{32}Choosing words as tree node labels as in some Minimalist approaches is not a way out. Although this may help for examples like university buildings, examples like buffalo buffalo (buffalo meat prepared by buffalo) would still be problematic.

\textsuperscript{33}In Stabler 1996, each nonterminal node in a tree is labeled with \textgreater{} or \textless{}, the opened side of the node label pointing to the head. These node labels are functions from nodes to nodes, just like arc labels. Therefore Stabler's approach is equivalent to posting arc labels as proposed here.
provides a formal model for representing P & P constituent relations, as well as for characterizing the relationship between constituent relations and other types of licensing relations. Constraints define levels of phrase structure. Generalizations between different licensing relations are captured in a subsumption hierarchy, summarized as Figure 34. The theory of phrase structure proposed here is relationally based, so that SPECIFIER, COMPLEMENT and ADJUNCT are viewed as non-derived syntactic relations, much like syntactic relations in Relational Grammar and subsequent frameworks.

The structural approaches that are usually assumed in P & P are inaccurate and imprecise because they lack the means to unambiguously represent basic constituent relations. Generalizations between types of licensing relations can only be expressed informally in these structurally based models because they lack the means to formally represent most licensing relations.

In this chapter (and throughout the book), I assume three classes of constituent relations: SPECIFIER, COMPLEMENT and ADJUNCT. Of these, the SPECIFIER class is the hardest to define. I am unaware of any attempt to show that those items assumed to be specifiers form a class, or to provide a notional definition of the term specifier in the way it is currently used in P & P theories. Unlike Kayne, I am inclined to make further distinctions instead of collapsing specifiers with adjuncts. For future research, I propose the following classes in lieu of a single SPECIFIER class: subjects (Ext-theta and Spec-Case licensees), degree licensors, quantifiers and possessives. It is easy to imagine how Figure 34 could be modified to accommodate this change. The specifier node in the graph would be pruned and replaced by its children. Then a new node would be created for the relation SUBJECT, with the relations EXT-THETA and SPEC-CASE as daughters.
Predicate Licensing

4.1 Introduction

This chapter presents a GBUG-based account of argument structure in which one constituent \( L \) (the predicate licensor) of each nonterminal \( N \) semantically selects its sister constituents (the predicate licensees) and determines all idiosyncratic word order properties of the predicate licensor relative to the predicate licensees. I show that the predicate licensor \( L \) justifies or ANCHORS the nonterminal \( N \). For every \( L \) and for each \( N \) anchored by \( L \), we propose one entry in our lexicon in the form of a FS partially describing \( N \). The complete lexical entry also includes linguistic constraints which are idiosyncratic to \( L \), including word order constraints (discussed below). Our usage of the terms \textit{lexicon} and \textit{lexical entry} differ from common usage in that we assume that \( L \) can be either a word or a phrase, whereas these terms are usually applied to words only.\(^1\) Stating that \( L \) is the \textit{anchor} of the nonterminal \( N \) refers to \( L \)'s special function as the constituent which justifies \( N \) in the sense that \( N \) is derived from a lexical entry for \( L \). The term \textit{predicate licensor} refers to the function of a constituent as the licensor in a predicate licensing relation. The predicate licensor's role in determining key semantic and word order properties of the phrase is given as support of a theory in which each nonterminal \( N \) contains exactly one predicate licensor \( L \) and \( L \) anchors \( N \).

4.2 Types of Predicate Licensing Relations

The predicate licensing relations considered in this chapter include: \textit{THETA, MODIFICATION, DEGREE}, and \textit{QUANTIFICATION}. Table 4.2 summarizes how the licensors and licensees of these predicate licensing relations correspond to licensors and licensees of constituent

\(^1\)Commonly, idiomatic phrases are given lexical entries. Here it is proposed that certain nonidiomatic phrases have lexical entries as well.
licensing relations. Heads of phrases are theta licensors (theta assigners) assigning theta roles to theta licensees (theta recipients). Heads are also modification licensees (modifiees), degree licensees and quantification licensees (quantifiees). Other predicate licensors may be adjuncts like modification licensors (modifiers), or specifiers like quantification licensors (quantifiers, including all determiners) and degree licensors (degree words).

The two types of THETA relations discussed here differ in that the theta licensee is a specifier of the THETA licensor in the EXT-THETA relation and a complement in the INT-THETA relation (cf. Williams 1980). EXT-THETA should be assumed to be approximately the same as the (initial) SUBJECT relation in RG, HPSG and other theories. The INT-THETA relation includes the (initial) object, indirect object and prepositional object relations of other relationally based theories.

4.3 The Anchor

Much like a functor of Categorial Grammar (CG), the anchor determines which constituents it can combine with and which type of nonterminal will result from such a combination (the "type" of constituent is assumed to include category, semantic features, etc.).

In GBUG, the combinatory features of an anchor A are stored in the lexicon, whether A is a word or a type of phrase. In contrast, the category of a word or phrase determines its combinatory properties in CG, e.g. an adjective modifier of a noun is of category N/N meaning that it combines with a noun to produce a noun. As diagrammed in Figure 43,

<table>
<thead>
<tr>
<th>Relation</th>
<th>Licensee</th>
<th>Licensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXT-THETA</td>
<td>Head</td>
<td>Specifier</td>
</tr>
<tr>
<td>INT-THETA</td>
<td>Head</td>
<td>Complement</td>
</tr>
<tr>
<td>MODIFICATION</td>
<td>Adjunct</td>
<td>Head</td>
</tr>
<tr>
<td>DEGREE/QUANTIFICATION</td>
<td>Specifier</td>
<td>Head</td>
</tr>
</tbody>
</table>

Table 4 Differences Among Predicate Licensor Relations

2The subject relation in LFG is similar to EXT-THETA for some constructions, but not others. For example, the subject of a passive bears the internal-theta in P & Pand the (initial) object relation in RG/HPSG, but is the subject in LFG's F-structure. The relation between the subject of the passive and object of the active is represented by means of a lexical rule in LFG.


4In theory one can factor out 1000 idiosyncratic differences between lexical items
the preposition on anchors a barbarP which can be completed by the
N complement the shelf. Figure 43 gives a representation of the word
on, the graph anchored by on, and the graph representing on the shelf.
The dashed lines indicate where the anchor and complement “fit” into
the graph anchored by on. The resulting phrase on the shelf anchors a
larger phrase, as shown in Figure 44. This graph can be completed by
the N it modifies resulting in a new N. Both the preposition on and
the P's anchored by on have entries in our lexicon which must be looked
up during a syntactic derivation of the phrase books on the shelf. The
conception of anchor proposed here is based on work in Lexicalized Tag
Adjoining Grammars (LTAGs) and Stratified Feature Grammar (SFG).
Many of the anchor-based analyses discussed in this chapter are based
on some SFG work by David Johnson, Lawrence Moss and myself (cf.

It is commonly assumed in P & P analyses that phrase structure is
projected from the heads of phrases. In CG terms this is equivalent
to assuming that each functor/anchor is an EXOCENTRIC category
X/Y, a category which combines with a category of type Y to produce a
category of type X, where X does not equal Y. For example, finite verb
phrases may be viewed as S/NP categories, categories which combine
with NPs to yield Ss. In P & P it is assumed that X/Y is the head of
X (X/Y and X are the same category, but the bar level of X/Y cannot
exceed that of X). A P & P finite T may be viewed as a T/N, an T lacking
a subject. The above analysis of modification by Ps assumes that these
Ps (and other types of phrases) are endocentric anchors, anchors of type
X/X which combine with a category of type X to yield a new X. Under
the analysis proposed here, the derivation of phrase structure is anchor-
driven, not head-driven. Anchors are only sometimes heads, following
previous work in Categorial Grammar, Stratified Feature Grammar and
Keenan 1974.

Some linguists have assumed that the role of a constituent C as

---

5 books is both an N and an N as defined above. The "one pronounalization" test
suggests that on the shelf modifies Ns rather than full Ns, e.g., The green books on
the shelf are bigger than the red ones.

6 Henceforth, I will use the term anchor to refer simultaneously to anchors and
functors.

7 Word order is not represented in the CG notation used here. X/Y is meant to
refer to any of the following: X/Y functors looking leftward for a Y, X/Y functors
looking rightward for a Y, or X/Y functors looking in either direction for a Y.
Graph Anchored by "on"

**Figure 43 A Word as an Anchor**
Graph Anchored by "on the shelf"

**Figure 44** A Phrase as an Anchor
an anchor of a nonterminal $N$ is evidence that $C$ is the head of $N$ (Hudson 1984, 1987, 1993). It would seem that this assumption is implicit in many accounts assuming analyses in which modifiers (adjectives), quantifiers (determiners) and degree words are special types of heads, e.g., Abney 1987 and Rothstein 1991. For example, the assumption that determiners are heads (the DP analysis) in Abney’s account allowed for a promising analysis of the POSSING construction in which the determiner *John’s* selects either an NP (1a) or a GERUNDP complement (1b). The crucial assumption for this analysis is that the determiner is the anchor selecting its sister constituent. Abney seems to assume that only heads can be anchors, an assumption not shared here. In Meyers 1995a (summarized below), it is shown that the determiner is the predicate licensor (anchor) and not the head of noun and gerund phrases. Similarly, it is shown that modifiers and degree words are anchors, not heads, in modifier + modifiee and degree word + head constructions. Abney’s account of the POSSING construction can thus be reformulated as a proposal that possessive determiners anchor $\overline{N}$s or $\overline{G}$s selecting their $\overline{N}/\overline{G}$ sisters, as in (1) (equivalent to Meyers 1995a, example (2)).

(1) a. Mary’s sandwich (DP or NP)
   b. Mary’s eating the sandwich (DP or GerundP)
   c. The sandwich (DP or NP)
   d. The big sandwich (DP or NP)
   e. The very big sandwich (DP or NP)
   f. The three biggest sandwiches (DP or NP)

The Abney 1987 account seems to analyze nominal constituents alternatively as $\overline{Degs}$, $\overline{Adj}$s, $\overline{Ds}$ or $\overline{Ns}$ depending on the anchor, as in (2) (from examples (1) and (3) in Meyers 1995a). Under a naive view of this DP analysis, verbs like *read* would seem to take complements with a wide range of categories. However, Abney would assume that verbs like *read* select for the THEMATIC head of the phrase, the noun, rather than the FUNCTIONAL head of the phrase, the anchor. It would appear that Abney’s functional heads are equivalent to GBUG predicate licensors (or anchors) and that Abney’s thematic heads are equivalent to GBUG constituent licensors (or heads). Under the GBUG account, constituent licensors determine the category of the phrase (the examples in (2) are all $\overline{Ns}$ and *read* always takes an $\overline{N}$ complement). Under Abney’s account, the functional head determines the category of the phrase (the examples in (2) belong to a variety of categories). What is at issue is which type of licensor determines the category of the phrase.\(^8\)

\(^8\)One might propose empty functional heads for all nominal phrases lacking adjec-
(2)  a. the very long books (DP or NP)  
b. very long books (DegP or NP)  
c. long books (AP or NP)  
d. books (NP)  
e. dirty water (AP or NP)  
f. primeval man (AP or NP)  

Most definitions of the term *head* (or equivalent) imply the head is the "distributional equivalent" of the whole phrase (Bloomfield 1933, p. 233). Abney's account does not explain how the functional head determines where a phrase can occur. In contrast, Definition 6 (repeated below as Definition 2) predicts that all phrases with a particular type of constituent licensor or head will have approximately the same distribution, as determined by selection and subcategorization. The category NP, accurately represents features of the phrases in (2) which will determine their common distribution. If functional heads are defined like our predicate licensors,⁹ it is not clear how they determine the distribution of the phrases they anchor and it is unclear why phrases anchored by determiners should be given the category DP.

**Definition 1** The CONSTITUENT LICENSOR of a phrase XP is the constituent that is obligatory, barring ellipsis, and that bears the features F of XP, such that F are subject to semantic selection and subcategorization restrictions imposed by constituents external to XP.

As noted by Zwicky 1985, Hudson 1987 and others, it is not entirely clear that nouns are "distributionally equivalent" to the phrases they head, given that the determiner is sometimes obligatory. The approximate restriction appears to be that determiners are obligatory with non-generic singular count nouns. This obligatoriness is explained in Section 3.4. Non-generic singular count nouns do not have any value for the feature *QUANT-VAL* (a head feature), and adjectives do not

---

⁹I am unaware of any definition of functional head in the P & P literature. Abney 1987 does not provide one.
either. Other types of nouns, as well as determiners, have either definite (DEF) or nondefinite (INDEF) values, where INDEF \(\subseteq\) DEF. As a result, a noun constituent whose determiner and/or head noun is marked with DEF or INDEF satisfy the \(\overline{N}\) constraint, i.e., they are complete \(\overline{N}\) constituents. Nominal constituents headed by non-generic singular count nouns with no determiners are not complete because their \(QUANT-VAL\) feature values are not subsumed by INDEF. In this way, the obligatoriness of determiners is explained. Many non-head constituents are also obligatory. Therefore the obligatoriness of determiners is by no means special. For example, the adjuncts \textit{last} and \textit{to please John's mother}, the complements \textit{his home} and \textit{for me to leave} and each instance of the specifier (subject) \textit{John} are all obligatory in (3).\(^{10}\)

(3) a. [Last night], John went to the movies
b. The building was built [to please John’s mother]
c. [John] laughed
d. John built [his home]
e. I would like [for you to leave]

While the above analysis seem to work well for \(\overline{N}\), Definition 6 may still be problematic for clausal constituents (Arnold Zwicky and John Richardson, personal communication). Verbs which select subjunctive clauses select both the complementizer and the morphological form of the verb, as in (4). Verbs like \textit{prevent} and \textit{refrain} select a complement consisting of a particular preposition (from and require that the prepositional object is a gerundal phrase (other verbal or nominal phrases cannot occur as their objects), as in (5).\(^{11}\) In both these cases, two sister constituents are selected for by a superordinate phrase. Thus my definition of constituent licensor does not identify a single item as the head. However, these constituents raise this same difficulty for all definitions of head that I know of (See Zwicky 1993 for discussion). It is therefore tempting to assume that these constructions violate \(\overline{\lambda}\) theory just like coordinate phrases (as discussed above). Not surprisingly, many linguistic theories assume that clausal complements are non-headed structures derived by phrase structure rules like \(S \Rightarrow NP VP\) and \(S' \Rightarrow COMP S\).

\(^{10}\)See Section 3.4 regarding the adverbial NP adjuncts like \textit{last night} and \textit{every day}. Grimshaw 1990 (pp. 132-133) notes that certain passives are ungrammatical without a passive \textit{by} phrase or a subject oriented adjunct.

\(^{11}\)The subjunctive case is discussed in Payne 1993. Both cases were previously noted to me by Mark Baltin in 1989 during my work with him as a research assistant studying \textit{The Lexical Representation of Non-Local Selectional Dependencies} (NSF Grant BNS8819857).
Pending further research, I assume (as a default) that the complementizers and prepositions are heads of these problematic constituents.

(4) a. I desire that he give me his trust
   b. *I desire that he gives me his trust

(5) a. I prevented him from leaving the room
   b. *I prevented him to leaving the room
   c. *I prevented him from leave the room
   d. *I prevented him from the departure

4.4 Selection Restrictions

4.4.1 Selection Restrictions as Evidence for Predicate Licensing Relations

One of the defining properties of predicate licensors is that they impose selection restrictions on their predicate licensee sisters. For example, the requirement that the referent of the object of cat be physically possible to eat (e.g., not an idea) and the requirement that the subject of cat be physically capable of eating (e.g., not a period of time) are evidence that the relations INT-THETA(cat, apple) and EXT-THETA(cat, Mary) are part of the syntactic description of the sentence Mary ate an apple. INT-THETA is a subtype of both the THETA relation and the COMPLEMENT relation. EXT-THETA is a subtype of the THETA and SPECIFIER relations. The requirement that the modifiee of pregnant be animate (not a color) and the requirement that the modifiee of dark be something which can have pigment (not a smell) are evidence that the relations Modification(pregnant, mare) and Modification(dark, blue) are part of the syntactic descriptions of the Ns the pregnant mare and dark blue. The meaning of each predicate licensing relation (e.g. the kind of information that the labels agent, patient, theme, etc. represent) are left to the field of natural language semantics. Specific labels which distinguish the meaning of the eater relation from, e.g., the believer relation are not provided in this account. Labels which distinguish modifiees of pregnant from modifiees of dark are not provided either.

Some previous P & P accounts assume that selection restrictions are completely predictable from particular theta roles, which are listed by name (agent, patient, theme, etc.). For example, Chomsky 1986b, p. 86, uses theta roles to represent the selectional properties which a verb imposes on its complements. One problem with these accounts is that it is not clear whether or not there are a finite number of different theta roles. If one takes the idea seriously that selection restrictions are completely
predictable from theta roles, the need for a possibly infinite number of theta roles becomes immediately obvious. On a simple minded account, the objects of *forecast* and *estimate* would appear to give the same theta roles to their objects (some subclass of *THEME* perhaps). However, a comparison of the examples in (6) demonstrate that the selection restrictions differ. In a theory in which selection restrictions are completely predictable from theta roles, even these very similar theta roles would have to be distinguished. One could imagine a different theta role for every verbal object. Another problem is that distinctions between types of theta roles is not always clear in particular cases, e.g., are *Mary* and *the dog* in patient or theme relations (7)? The problematic features of these accounts of theta roles would no doubt multiply in an analogous account of all predicate licensing relations.

(6) a. John forecast/*estimated the weather.
   b. John estimated/*forecast their expenses.

(7) a. Bernice pushed Mary onto the plane
   b. Bernice walked the dog

Other accounts assume that the content of theta roles has no status in syntax (Ravin 1976, Hoekstra 1984, Levin and Rappaport 1986). Under this latter assumption, it suffices to show that a theta relation exists and it is unnecessary for a syntactic description to specify the type of theta relation (agent, patient, theme, etc.). This section outlines a version of the latter approach, extended to include all types of predicate licensing relations. The existence of a selection restriction is taken as evidence for a predicate licensing relation, but subtypes of each class of predicate relations are not distinguished. The approach taken here, is to distinguish predicate licensing relations on syntactic grounds only, e.g., the constituent licensor/licensee status of the predicate licensors and licensees, the category of the predicate licensee, etc.\(^\text{12}\)

### 4.4.2 Defining Selection Restrictions

Following the approaches of McCawley 1968b, 1968a, 1970 (attributed to Fillmore) and Lakoff 1969, a selection restriction is a set of presuppositions associated with \(A\), a lexical item or phrase, about the type of lexical items or phrases which can co-occur with \(A\). An element \(B\) can only co-occur with \(A\) felicitously if \(B\) is compatible with these pre-

---

\(^{12}\) On syntactic grounds, it may be difficult to distinguish degree words like *very* and *disgustingly* (*very* rich, *disgustingly* rich) from adverbs that can modify adjectives and/or adverbs as in *legally blind, actively political and admirably well*. There is no available interpretation of *very actively political* in which *very* modifies the degree of political-ness as would be expected if *very* was a specifier, distinct from the modifier *actively*. 
suppositions and our assumptions about the world. For example, the adjective *pregnant* is associated with the presupposition that referents of nouns it can modify have the capacity to be pregnant. Our world knowledge tells us that only (female) animals can be pregnant. The Ns *the pregnant mare* and *the pregnant ophthalmologist* are compatible with our world knowledge. The Ns *the pregnant rock* and *the pregnant color* are incompatible with our world knowledge. *Color* and *rock* violate the selection restrictions associated with the adjective *pregnant*, but *mare* and *ophthalmologist* satisfy these restrictions.\(^{13}\)

Observations 1 and 2 are explained under the above account of selection restrictions. These informal observations may be used to determine whether \(A\) selects \(B\) or \(B\) selects \(A\). If \(A\) selects \(B\), \(A\) is the predicate licensor and \(B\) is the predicate licensee for a given predicate licensing relation. If \(B\) selects \(A\), the reverse holds.

**Observation 1** Given a predicate licensor \(P\) which imposes a selection restriction \(S\) on any item \(I\), if \(I\) is unspecified for the meaning trait \(T\) restricted by \(S\), then \(I\) is interpreted as if its value for \(T\) is compatible with \(S\), e.g., pregnant ophthalmologists are normally viewed as being female even though the noun *ophthalmologist* is unspecified for gender. This test is based on discussions in Katz and Postal 1964 (p. 83) and Katz 1972 which focused on the effect of selection restrictions on the meaning of pronouns. (See below for some examples using pronouns.)

**Observation 2** In the context of a belief, a joke, a riddle, a fantasy, a dream, a folktale, a science fiction story, or other contrived situation, phrases violating selection restrictions under normal assumptions are well-formed. In the terminology used here, a FORCED reading of a phrase relies on one of these fantastic or obscure settings to make sense. If \(A\) selects \(B\), then a forced reading of \(B\) will be possible in a semantically anomalous sentence, but not a forced reading of \(A\). For example, in the context of a science fiction book about a planet where inanimate objects and colors are alive, *the pregnant rock* and *the pregnant color* may be well-formed. What changes in this context are our assumptions about the meaning of *color* and *rock*, not our assumptions about

---

\(^{13}\)Semantic anomaly involves impossible, not improbable concepts, e.g., it is possible, but improbable that one might eat rocks. Also people vary as to what they believe is possible, e.g., regarding whether a non-human animal can think. To avoid confusion, examples of clear semantic anomaly will be used where possible, including violations of abstract/concrete and animate/non-animate noun restrictions, even though other examples may still be semantically anomalous, e.g., male/female violations.
Verb Class A  read, write, peruse, edit, . . .
Verb Class B  please, tease, tickle, annoy, . . .

Figure 45  Verbs Selecting Two Distinct Semantic Types of Complements

the meaning of pregnant. Therefore the adjective pregnant selects its modifiee. See McIntosh 1961, Ziff 1964 and Weinreich 1966 for analyses which assume that fantasy settings change the meaning of only one word of two word collocations.14

Assuming that theta role assigners (licensors) select theta role recipients (licensees), Observation 1 predicts that Ns which may or may not be interpreted as texts or animals, will be interpreted as text when given as a complement of verbs in Verb-class-A in Figure 45 and as animals when given as a complement of verbs in Verb-class-B, as evidenced in (8). Pronouns have unspecified referents. In (8a,b), the selection restrictions imposed by the verbs on their pronoun complements limit the possible referents of these pronouns. The pronouns in (8a) must refer to animals and the pronouns in (8b) must refer to texts. (8a,b) provide evidence that the verbs select their complements because the refinement of reference in those example can only be attributed to the verb. In (8c), the verbal unit “do something to” suggests some unspecified action, the complement of which fails to provide any additional specification, i.e., the thing done to the book or to the child is just as unspecified. Therefore, verbs select their complements, but complements do not select head verbs.

(8)  a. The child teased it/something.
     b. Gertrude read it/something.
     c. They did something to the book/the child.

In a \(\text{V}\) consisting of one verb \(V\) and one \(N\) complement \(N\), forced readings in the sense of Observation 2 allow our real world assumptions about \(N\) to change, but not our real world assumptions about \(V\). The sentences in (9) contain well-formed \(\text{V}\)'s and the sentences in (10) contain \(\text{V}\)'s which are incompatible with selection restrictions. This semantic anomaly is indicated by preceding the sentence with \(\sim\). In a fantasy world consistent with (10), our assumptions about the \(\text{V}\)'s change, not our assumptions about the verbs. In a world where written material is personified, it is possible for a book to be teased or annoyed. In a world where people and animals are made up of text, a child may be

14Ralph Grishman (personal communication) was very helpful in clarifying my views on this topic.
edited. We cannot find similar fantasy worlds where the meanings of the verbs change. For example, it is not the case that *The babysitter tickled the book* can mean that the babysitter read the book in some possible world. Therefore Observation 2 supports the hypothesis that verbs select their complements, but not the hypothesis that complements select head verbs.

(9) a. The play pleased/annoyed Gertrude/Cecil/the child/the young acrobat
   b. The babysitter tickled/teased Gertrude/Cecil/the child/the young acrobat
   c. The expert read/wrote/edited the book/the magazine/the screenplay

(10) a. ¬The play pleased/annoyed the book/the magazine/the screenplay
    b. ¬The babysitter tickled/teased the book/the magazine/the screenplay
    c. ¬The expert read/wrote/edited Gertrude/Cecil/the child/the young acrobat

In determining selection, we are only interested in changing views about the world so that a literal interpretation is possible. The only interpretations of (11) relevant to my approach are ones in which facts about the world change so that cars can be pregnant and machines are capable of eating. These forced readings of phrases which violate selection restrictions are distinct from poetic or metaphorical use of language. Metaphorically, any word in a phrase may be taken to mean something else. (11a) may be taken to be either a car which bulges in the middle or a human being who resembles a car in some way. Some metaphorically interpretations of sentences have entered common usage, e.g., (11b) means that my quarter went into the vending machine, but nothing was dispensed in return. These metaphorical readings are not relevant to determining selection restrictions.

(11) a. A pregnant car
    b. The vending machine ate my quarter

4.4.3 Examples of Selection by Predicate Licensors

Examples (12) through (15) illustrate that licensors of internal-theta, external-theta, modification and degree select their licensees, given Observations 1 and 2.

(12) a. Mary ate a cheese sandwich [INT-THETA AND EXT-THETA]
    b. Mary read the book [INT-THETA AND EXT-THETA]
c. That is a pregnant cow [MODIFICATION]
d. Gertrude read the text slowly/with a magnifying glass [MODIFICATION]
e. The politician was too/very greedy [DEGREE]
f. The politician’s explanation was accepted too/very quickly [DEGREE]

(13) a. Mary ate something [INT-THETA]
b. It ate the cheese sandwich [EXT-THETA]
c. It was pregnant [EXT-THETA]
d. The green thing [MODIFICATION]
e. The pregnant one [MODIFICATION]
f. Mary did that slowly/with a magnifying glass [MODIFICATION]
g. Mary was too/very something [DEGREE]
h. John danced too/very much that way [DEGREE]

(14) a. Mary did something to the cheese sandwich [INT-THETA/EXT-THETA]
b. Cecil found his keys that way [MODIFICATION]
c. They timed the race that way [MODIFICATION]
d. Cecil found his keys in some manner [MODIFICATION]
e. They timed the race in some manner [MODIFICATION]
f. John was, to some extent, angry [DEGREE]
g. John danced, to some extent, clumsily [DEGREE]

(15) a. ~Mary ate the color blue/the song. [INT-THETA]
b. ~The sneeze ate the breeze. [EXT-THETA]
c. ~The color was pregnant. [EXT-THETA]
d. ~You can’t drive the pregnant car [MODIFICATION]
e. ~Mary knew French slowly [MODIFICATION]
f. ~The goldfish was very/too dead/asleep [DEGREE]
g. *The politician’s explanation was accepted very/too completely [DEGREE]

Observation 1 states that the range of possible meanings of predicate licensee constituents with general meanings (pronouns it, one, something, somehow, complex verbs consisting of do plus a pronoun and sometimes a preposition; the adverbial NP that way; the PP in some manner, etc.) is narrowed by predicate licensors which select these constituents. These predicate licensees, as in (13), can be be interpreted as the corresponding licensees in (12), but not as the corresponding licensees in (15). This suggests that the predicate licensors in the above
examples (including 12) select their licensees. In (14), the predicate licensors have general meanings which are not narrowed in any way by their licensees. This suggests that predicate licensees do not generally select their licensors.

Observation 2 evidence is illustrated in (15). These anomalous examples would be well-formed in a world in which abstract things could be eaten, sneezes, colors and cars could eat and get pregnant, accessing one's knowledge could take time, death and sleep could be graded (without irony or referring to deepness of sleep). Note that ill-formed degree words plus adverb combinations are usually ungrammatical (e.g. 10g) just like the determiner examples discussed below (Section 4.4.5). Worlds in which the meanings of the predicate licensors change are not evoked by these examples. It seems that non-literal interpretations (metaphor, irony, etc.) provide the only way to change the meaning of the predicate licensors. Therefore the predicate licensors select their licensees.

4.4.4 Selection of Prepositional Objects

For a subset of $\mathcal{P}$s, it appears that the preposition does not select its object, but rather that the internal theta licensor (verb, adjective, noun) selecting the $\mathcal{P}$ selects the prepositional object as well. Observations 1 and 2 cannot distinguish whether the head governing the $\mathcal{P}$ or the preposition itself imposes a particular selection restriction, but only that the environment as a whole imposes the restrictions exemplified in (16). This analysis seems best suited for complement $\mathcal{P}$s such that the preposition is idiosyncratically selected, as in (17). For example, verbs of this type allow the prepositional object to be fronted in passive constructions as in (18). A straightforward account of passive is possible if it is assumed that the prepositional object is in some sense an object of the verb. Figure 46 is one possible GBUG account in which the adjective angry takes a $\mathcal{P}$ complement and assigns a theta role to the prepositional object. The preposition at licenses a COMP-CASE relation with the prepositional object as licensee. In this analysis, the $\mathcal{P}$ is a superficial complement of the adjective (not a theta role or case recipient).

(16)  a. Mary was angry [at it]
     b. Mary was angry [at Fran]
     c. ~She was angry [at the color blue]
     d. She spied [on them]
     e. She spied [on her friends]
     f. ~She spied [on ridiculous concepts]
     g. Mary examined the eggplant [on it].
     h. Mary examined the eggplant [on the table].
i. Mary examined the eggplant [on the sour taste].

(17) a. She spied [on her friends]
    b. They talked [about themselves]
    c. Fred was angry [at Bill]
    d. A gift [from Fred]

(18) a. He was spied on all the time
    b. They were always being talked about
    c. The evidence has been looked at

Analyses in which a head other than the preposition selects the prepositional object have been developed in a wide range of linguistic theories. In Case Grammar (Filmore 1968), it is assumed that the preposition merely acts like a case affix. In Relational Grammar (see various articles in Perlmutter 1984), the prepositional object bears a syntactic
relation to the main verb. In the related Arc Pair Grammar analysis (Johnson and Postal 1980, pp. 602–651), the preposition flags the NP: the NP is superficially part of the PP, but bears its fundamental relation to the verb. In some P & P work including George 1980 and Rouveret and Vergnaud 1980 (pp. 130–132), the preposition is inserted as a case marker. In other P & P work, the verb and preposition are reanalyzed as a complex verb (Riemsdijk 1978). Figure 46 represents an analysis along the same lines as these approaches.

The prepositions themselves seem to select their complement for adjunct \( \mathit{P} \)'s as in (19), \( \mathit{P} \)' complements that occur predicatively as in (20) and for a subset of the \( \mathit{P} \)' complements for verbs which select a semantic type of \( \mathit{P} \) as in (21)\(^{15}\). In each of these cases, the prepositional object can be predicted on the basis of the preposition, in contrast with the (17) examples. Since adjunct \( \mathit{P} \)'s select the heads they modify, it is not possible for the verbs to select the prepositional object. Evidence that the preposition selects its object in these cases comes from examples like (16i) in which a locational interpretation is forced on the prepositional object by the locative preposition \( \mathit{on} \). A wide range of prepositions can occur as predicate complements (there is no easy way to generalize). Unlike nonpredicative \( \mathit{P} \)' complements, the head does not force a particular meaning on the preposition, e.g., \( \mathit{in} \) may be temporal or locational as in (20j). Heads that select for prepositions by semantic types in cases like (21) contrast with the idiosyncratic nature of preposition selection in (17). These types behave in a fashion predictable by the preposition. These facts would be difficult to account for if it was assumed that some constituent other than the preposition selected the prepositional object in these cases.

(19)  
(a) Mary examined the eggplant [on the table]  
(b) Mary discussed that [in a silly voice]  
(c) Mary read [with a magnifying glass]  
(d) Mary left [on Thursday]  

(20)  
(a) Mary ended up/appeared [in the room]  
(b) John seems/acts [like a nice guy]  
(c) John seems [against the idea]  
(d) I felt [out of sorts]/[like a linguist]  
(e) I consider Mary [above suspicion]/[against the idea]  
(f) I consider the issue [of concern]  

\(^{15}\) See Grishman and Meyers 1996 for a proposal to use these distinctions in a dictionary of prepositional phrases. Some of the examples in (17) to (21) are taken from this proposal.
g. I caught/found him [in the bakery]
h. I declared the meeting [in recess]
i. I judged him [in the room]/[in contempt]
j. The meeting was [in the room]/[in June]

(21) a. John walked [to/away from the movies] [DIRECTIONAL PP]
b. John lived [on a farm]/[in Cleveland] [LOCATIONAL PP]
c. The meeting begins [at 5 o’clock]/[in five days] [TEMPORAL PP]
d. John behaved [in a peculiar way]/[like Fred] [MANNER PP]

4.4.5 Selection by Determiners

The selection restrictions between a determiner $D$ and its sister nominal constituent $N$ can be quite complex. For all $D$ and $N$ which form constituents, there is assumed to be a predicate licensing relation $QUANTIFICATION(D, N)$. $D$ selects $N$ in the sense that $D$ requires $N$ to have certain values of $QUANT-VAL$ (see Section 3.4), $INDIVIDUATION (+/- mass), NUMBER, AMOUNT$ (plurals whose quantities are specified are assumed to be $AMOUNT(TRUE)$), e.g., numbers, $several$, (unstressed) $some$, etc.) and other features. Once the determiner and its quantifier combine into a single constituent, the values of many of these features will combine via structure sharing. It is assumed that the selection constraints are imposed on quantifiers before such combination.

These requirements differ from other selection restrictions in that violations often result in ungrammaticality rather than semantic anomaly. Observation 2 cannot be applied when selection violations result in ungrammaticality. Also, Observation 1 cannot be applied because I could not find nouns that were unspecified for the traits that determiners select for. Pronouns like $it$ are not selected for by any determiner. Other common nouns with little semantic content like $thing$ are non-generic common nouns which are selected by determiners like $the, each, a$, but not by (for example) determiners like $much$ which require mass noun quantifiers.

In GBUG, the phenomenon of quantification is represented by means of two distinct licensing relations: (1) the predicate licensing relation (quantification licensing), discussed here, in which a determiner $D$ is the licensors and its adjacent $N$ is the licensee; and (2) the scope licensing relation (see Chapter 6) in which the determiner $D$ is the licensor and the phrase over which it has scope is the licensee. As in other predicate relations, the quantification licensor (quantifier) selects the quantification licensee (quantifier). It should be clear that when I discuss the quan-
tification relation in terms of quantifiers and quantifiees, that quantifier scope (scope licensing) is a separate topic.

Table 4.4.5 lists some determiners and FS logic formulations of the selectional restrictions that they impose on their quantifiees. Determiners which occur more than once in the table impose more than one of the constraints listed. Violations of these constraints may result in ungrammaticality (*) or semantic anomaly (~) as indicated in (22). With quantification licensing relations, semantic anomaly (as exemplified in (23)) can involve category violations (e.g., if is not a noun), non-mass nouns being interpreted as mass nouns (e.g., man), mass nouns being viewed as non-mass nouns (e.g., blood), and pronouns/pronouns (e.g., Neptune) (which are normally definite) being viewed as common nouns (which are indefinite). I propose that the quantifier selects its quantifiee in all these cases, even though the nature of the ill-formedness resulting from violations of these restrictions is not uniform. In fact some instances of pronouns being coerced into common nouns are not ungrammatical at all, e.g., The Adam Meyers in San Francisco is taller than the Adam Meyers in New York. Perhaps, it is the case that all pronouns have a common noun usage that is triggered by determiners like the.

(22) a. the (one) dog, the (three) dogs, *the the dog/dogs
   b. this dog, this one dog, *this the dog
   c. these dogs, these three dogs, *these the dogs
   d. each dog, *each one dog, *each a dog, *each the dog
   e. a dog, *a a dog, *a each dog, *a the dog, *a one dog
   f. every dog, every three dogs, *every the dog
   g. that dog, that one dog, *that the dog
   h. those dogs, those three dogs, *those the dogs
   i. such dogs, such a dog, *such three dogs, *such the dog/dogs
   j. any dogs, *any a dog, *any the dog/dogs, any one dog, any three dogs
   k. many a dog, many dogs, *many the/three dogs,
      l. some1 dog, some1 dogs, some1 water, *some1 the dog
   m. ~some2 book, some2 people, some2 water, *some2 the book
   n. ~more book, more books, *more the book
   o. ~much book, *much people, much water
   p. half a/the book, *half water
   q. both the men, both men, *both two/three/four men
   r. all (the) men, all the book, all metal, ~all book

16See Weinreich 1966, p. 435, for further examples of words being coerced by determiners into behaving like count and mass nouns.
17This is OK after a copula, e.g., The potion is half water (and half eye of newt).
<table>
<thead>
<tr>
<th>Quantifiers</th>
<th>Selection Restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Determiners</td>
<td></td>
</tr>
<tr>
<td>this/these, that/those, the,</td>
<td>Category(Nom)</td>
</tr>
<tr>
<td>every, such, each, any,</td>
<td>QUANT-VAL(¬Def)</td>
</tr>
<tr>
<td>stressed some (some1),</td>
<td></td>
</tr>
<tr>
<td>unstressed some (some2),</td>
<td></td>
</tr>
<tr>
<td>many, more, less</td>
<td></td>
</tr>
<tr>
<td>such, half</td>
<td>QUANT-VAL(Indef)</td>
</tr>
<tr>
<td>a, all numbers</td>
<td>QUANT-VAL(FALSE)</td>
</tr>
<tr>
<td>a, all numbers, some1, some2,</td>
<td>Specifier(FALSE)</td>
</tr>
<tr>
<td>many, more, less, much</td>
<td></td>
</tr>
<tr>
<td>much</td>
<td>Individuation(Mass)</td>
</tr>
<tr>
<td>a, every, each, many,</td>
<td>Individuation(¬Mass)</td>
</tr>
<tr>
<td>some1, half</td>
<td></td>
</tr>
<tr>
<td>a lot of, some2, more, less</td>
<td>Individuation(Mass)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>QUANT-VAL(Def)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>such, both</td>
<td>Amount(FALSE)</td>
</tr>
<tr>
<td>every, each, any</td>
<td>Number(Singular)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5** Constraints Quantifiers Impose on Quantifices
(23)  a. ~Scientists study the if (Weinreich 1966, p.463)
    b. ~I want to buy an angry
    c. ~The giant intellectual book worm from Mars ate too much book
    d. ~The breathing expert likes an air that has circulated properly before entering his nose.
    e. ~Most vampires prefer a blood they find themselves to a blood that is purchased from a blood bank.
    f. ~The cannibal ate too much man
    g. ~I know one she who owns a credenza.
    h. ~Each May 18, 1993 was the same, except for the very last May 18, 1993 when the earth finally exploded. (In a Sci-Fi universe in which time keeps repeating itself.)
    i. ~A new Neptune was found by the astronomer every day for a year.

As illustrated in (22), the determiners such, some1 (stressed some), some2 (unstressed some), every, half, both and the differ with respect to the quantifiers they allow. They all allow some type of Indef quantifier to follow, unlike a and numbers. Only half and both may be followed by the definite determiner the. Some1 and some2 cannot be followed by any determiner, but can be followed by Indef nominals headed by plural nouns or mass nouns respectively. The determiner much requires its quantifier to be headed by a mass noun. The determiners a lot of, more, less and some2 allow either mass or plural quantifiers. Note that selection for plural differs from requirements of number agreement in that these determiners are unmarked for plural or singular (mass nouns have singular agreement features), but still require plural quantifiers in some cases. The determiners a, every, each, many, half and some1 do not allow mass quantifiers. Such and both cannot be followed by numbers (which are all Amount(TURE)). every, each and any require their quantifier to be either singular or Amount(TURE). In the latter case they have a “group of” interpretation, e.g., every three dogs means “every group of three dogs”. The constraints in Table 4.4.5 are intended as approximations of what the actual selection constraints would have to be. The fact that these constraints permit some ill-formed examples suggests the need for additional constraints, some of which may not be selection restrictions.

Numbers (including one) and the determiner many can follow definite determiners the, this, these, that, those, etc. In my account, the value of QUANT-VAL is shared among the projections of a single noun (QUANT-VAL is a head feature). The Indef value for one unifies with
the Def value of the, making the one book a definite \( \overline{N} \). Other indefinite determiners could be expected to follow these definite determiners as well unless prevented by constraints like those in Table 4.4.5. However, examples like (24) are ungrammatical and are not prevented by these syntactic constraints. In particular, it would be very difficult to distinguish the determiners a and one semantically. Following Perlmutter 1970, we propose that the ill-formedness of (24) is due to a phonological constraint. The definite determiners precede a phonetically stressed position and a is obligatorily unstressed.\(^{18} \) Stressed some (some1) is definite, whereas unstressed some (some2) is indefinite. Some1 cannot occur in (24b) because the selects for a nondefinite quantifier and some2 is ruled out because the is followed by a stressed position and some2 is unstressed. Unlike the, such does not require its quantifier to be stressed. Thus the well-formedness of (24c) follows. (24d) is ruled out because some2 is an AMOUNT(TRUE) determiner.

(24) a. *The/this/that a (green) book  
b. *The some book/books  
c. Such a (green) book  
d. *Such some (green) books

4.4.6 Selection by and of Possessives

I propose that a possessive determiner, like other determiners, anchors an \( \overline{N} \) consisting of the determiner \( D \) and its quantifier \( N \). In addition to the selection restrictions associated with QUANTIFICATION\((D, N)\), other selection restrictions can exist between \( D \) and \( N \), including restrictions associated with the following relations: OWNERSHIP, a relation which I will assume includes authorship, inalienable possession, legal possession, etc.; and NOM-THETA, the THETA-like relation between a nominalization or picture noun and its possessive determiner.\(^{19} \) For NOM-THETA, \( N \) selects \( D \). For OWNERSHIP, the direction of selection is unclear. It is argued that on the simplest account, QUANTIFICATION is a predicate licensing relation and \( D \) is the anchor of the phrase. However, OWNERSHIP and NOM-THETA are...

---

\(^{18} \) Perlmutter 1970 proposes that a and one are unstressed and stressed forms of the same word. Although much of Perlmutter's account is preserved here, I assume that a and one are distinct words. For example, one is AMOUNT(TRUE), but a is AMOUNT(FALSE).

\(^{19} \) Catherine Macleod, Ralph Grishman, Leslie Barrett, Ruth Reeves and I are currently conducting research on mappings between verb arguments and the NOM-THETA arguments (possessive determiners, nominal adjuncts and complements) of the corresponding nominalization. We hope to create a dictionary of nominalizations (NOMLEX) which includes this information. This work is supported by the National Science Foundation under Grant No. IR1-9635280.
not predicate licensing relations. The fact that \( D \) selects some features of \( N \), and \( N \) may also select some features of \( D \) makes it necessary to distinguish the different semantic relations. Additionally, the assumption that, of these, only \( QUANTIFICATION \) is a predicate licensing relation means that selection is a necessary, not a sufficient condition for identifying predicate licensing relations.

As illustrated in Figure 4.4.6 and Table 4.4.6, \( \overline{N} \)s which contain possessive determiners, share the value of \( QUANT-VAL \) with the possessives. This follows from our previous discussion. Possessives like other \( \overline{N} \)s must obey the constraint \( CATEGORY : NOM \wedge QUANT-VAL : INDEF , where INDEF \subseteq DEF \). Like other determiners, possessives share their \( QUANT-VAL \) value with their quantifiee. The FS representation of the book’s cover in Figure 4.4.6 is derived as follows. The book’s, as represented in the top right of Figure 4.4.6 anchors the FS in the lower left. The book’s unifies with the value of \( QUANTIFIER \) and takes a quantifiee (head) of type \( Noun \) like \( cover \), resulting in the graph on the lower right.

The examples in (25) illustrate that selection for \( QUANT-VAL \) of the quantifiee is determined by the anchoring determiner of the possessive, e.g., both these and these women’s require their quantifiees to obey the constraint \( QUANT-VAL(\neg Def) \). Possessives lacking determiners behave much like the determiners the and a depending on whether they are definite or indefinite with respect to selection. Selection by possessives of features other than \( QUANT-VAL \) is a little tricky. No possessive \( N \) seems to allow definite quantifiees or obligatorily unstressed quantifiees, even if these quantifiees are allowed by the anchoring determiner. When mass nouns are possessives (they can possess qualities), their quantifiee cannot be \( Individuation(Mass) \), even if the anchoring determiner of the possessive allows mass quantifiees. Possessive determiners inherit many, but not all quantificational and selectional traits of their anchors.

(25) a. the woman’s (one) pet, the woman’s (three) pets, *the woman’s the pet, *the woman’s a pet
Figure 47: Structure sharing of QUANT-VAL in Possessives

Poss-Noun
Poss

Determiner the
Def

FS Representing "the book’s"

Poss-Noun
Poss

FS Anchored by possessive NP

Poss-Noun
Poss

FS Representing "the book’s cover"

Poss-Noun
Poss

FS Representing "cover"
b. a woman’s (one) dog, a woman’s three dogs, *a woman’s the dog, *a woman’s a dog
c. these women’s dogs, these women’s three dogs, *these women’s the dogs
d. every dog’s owner, every dog’s three concerns, *every dog’s the concerns
e. Mary’s (one) dog, Mary’s (three) dogs, *Mary’s the dog, *Mary’s a dog
f. such/many a dog’s owner, *such/many a book’s a cover, such/many a dog’s one favorite food
g. both the men’s coats, *both the men’s the coats
h. *some1/much water’s color, *some1/much water’s algae
i. some2 water’s color, some2 water’s algae

Williams (Williams 1982, p. 283) compares the theta relation \( \text{Theta}(V, S) \) between the verb \( V \) and its subject \( S \) to the relation \( R(N, P) \), where \( N \) is a nominalization which corresponds to \( V \) and \( P \) is the possessive determiner of the \( \text{NP} \) headed by \( N \). Williams notes that \( R(N, P) \) (or \( \text{NOM-THETA}(N, P) \)) varies to a much greater extent than \( \text{Theta}(V, S) \). The possessive can correspond to external or internal theta role recipients, including objects of prepositions within \( \text{NP} \) complements. In each example of (26), \( \text{NOM-THETA}(N, P) \) in each example of (26) corresponds to the theta relation \( \text{THETA}(V, N) \) listed in square brackets. For purposes of the \( \text{NOM-THETA}(N, P) \) relation, the head \( N \) selects the determiner \( D \), as evidenced by the following: (1) the possessives in (27) are semantically anomalous because they are not animate (or physical at the very least); and (2) the possessives in (28) are only analyzable as being animate. In contrast, the possessives do not impose selection restrictions related to the \( \text{NOM-THETA} \) relation.\(^{20}\)

\( \text{(26)} \)

\begin{enumerate}
\item a. The book’s interpretation [Ambiguous]
\item b. John’s interpretation of the book [External-Theta]
\item c. The book’s interpretation by the expert. [Internal-Theta]
\end{enumerate}

\(^{20}\) Note that if we were to assume that \( \text{NOM-THETA} \) was actually a type of \( \text{THETA} \), then a nominalization can have two \( \text{EXT-THETA} \) licensees when it occurs as a predicate nominal, e.g. (6). No \( P \) & \( P \) principle precludes this, outside of \( \text{APG} \) theory which does not allow multiple specifiers within a single phrase, a restriction which I have already shown to be to restrictive.

(i) I considered/judged the problem Fred’s obsession with money.

If \( \text{NOM-THETA} \) is really a type of \( \text{THETA} \), then in \( \text{RG} / \text{APG} \) terms, \( \text{NOM-THETA} \) would be interpreted as a variable over subject, object, etc. In (6), \text{obsession} would have two initial subjects, violating the stratal uniqueness theorem. If on the other hand, we assume that the possessive is not a real subject (there is no real theta relation), then this problem does not arise.
98 / Government and Binding Unification Grammar

d. John’s favorite interpretation was offered by Fred. [No Theta Relation]
e. John’s gift from Mary. [Internal-Theta of to]

(27) a. The time of day’s digestion of the food.
b. The color blue’s operation (on/by the doctor)
c. The flavor vanilla’s interpretation

(28) a. Its digestion of the food
b. Its operation on/by the scientist
c. Its interpretation is in question

Possessives in Ns headed by “picture nouns” behave much like nominalizations with respect to their alternation with the prepositions of and by, as in (29). Under the subject/author interpretations, I will assume that the NOM-THETA relation holds between picture nouns and their possessives, with the picture nouns selecting their possessives. The anomalous example favors a reanalysis of the word smell. The pronoun it must be interpreted as an cognizant entity\(^\text{21}\) (or instrument thereof) if it is to be interpreted as the painter of the portrait. I will assume that the NOM-THETA relation holds between picture nouns and their possessives.

(29) a. John’s picture/portrait/biography
b. John’s picture/portrait/biography of Mary
c. The picture/portrait/biography of John by Mary
d. John’s portrait by Van Gogh
e. Mary’s biography by Norman Mailer
f. The smell’s biography of/by John
g. Its portrait of/by John

OWNERSHIP(P,N) is another possible relation between the possessive P and the head N. In this case, the direction of selection is unclear. For some anomalous examples in (30), the head noun seems to change in meaning and for others, the possessive. In (30a), I imagine a lemon with human-like arms. In (30b), either Cecil owns gravity (the nature of gravity changes) or Cecil has his own gravity (the nature of people changes).\(^\text{22}\) The other examples vary with respect to whether the possessive and/or the head noun restricts the possible meanings of the other. I will not speculate any further, as it appears that identifying the

\(^{21}\) A person, an organization, a deity, a computer, an animal, etc.

\(^{22}\) Within the physics domain, all objects including people, exert gravity on other objects. However, within the domain of every day discourse, planets have gravity, but people do not.
“licensor” of the OWNERSHIP relation does not bear on predicate licensing. This is not an unusual property of the licensor. It is natural that licensor can take a licensor of its own. For example, if we change the licensor, we can have a licensor of itself. However, this is not the case for the licensor of itself. The licensor of itself cannot take a licensor of itself.

(30) a. The lemon’s arms [lemon changes]
   b. Cecil’s gravity [Either can change]
   c. The car’s missing part [part is restricted]
   d. Its idea [idea restricts the possessive]
   e. Its thing [Neither seems to restrict the other]

All possessives have the Quantification relation in common. If Quantification is assumed to be the only predicate licensing relation between a possessive determiner and the adjacent nominal constituent, the possessive is always the anchor of its N. A unified account of possessives follows. If, on the other hand, NOM-THETA, OWNERSHIP, and possibly other noun-possessive relations are assumed to have a role in determining the anchor of these Ns, we must assume a different account of possessives for each case. Under the simplest account, NOM-THETA and OWNERSHIP are not predicate licensing relations.

An analysis of nominalizations in which the possessive is the predicate licensor (and anchor) may strike some as counter intuitive because NOM-THETA is more important to understanding the meaning of the N and because of the obvious parallels with verbs and THETA relations. However, a uniform syntactic treatment of possessives requires that the possessive be the anchor.

4.5 Word Order

4.5.1 Introduction

This section demonstrates that the predicate licensor determines its relative word order to its predicate licensees. An examination of exceptional word orders reveals that varying the predicate licensor of a constituent leads to variations in word order, but varying the other constituents does not. In the examples discussed, changes in word order result from the presence of different modifiers, degree words, and theta role assigners, all of which are predicate licensors. Since heads are only sometimes predicate licensors, this section shows that word order cannot be determined solely on the basis of case properties and the head parameter, contrary to Travis 1983 and other previous P & P approaches to word order. I propose a new approach to parameter setting models in which

\footnotesize
23In Meyers 1994, I separated inalienable possession (INALIENABLE) from other types of possession (OWN). I claimed that possessives select their heads in the INALIENABLE relation. However, I no longer believe this is clear. I also assumed a different definition of predicate licensing which included NOM-THETA and OWNERSHIP.
word order constraints are parameterized for languages, parts of speech, lexical items, and possibly other levels of generalization. Both canonical and idiosyncratic word order is captured in a principled way. This section is based on Chapter 4 of Meyers 1994 and was previously presented at the 1995 Annual Meeting of the Linguistics Society of America (Meyers 1995b).

4.5.2 Word Order Parameter Settings
Following Johnson et al. 1993, I assume a word order parameter with three possible settings:

- Free Order
- Licensor First
- Licensor Last

For a given predicate relation: the Free Order setting imposes no constraints on word order, the Licensor First Setting requires the predicate licensor to precede the predicate licensee, and the Licensor Last Setting requires the predicate licensor to follow the predicate licensee. For purposes of parameter setting, I ignore passive, heavy NP shift and other constructions for which variation from basic word order is assumed to result from some operation on the lexical entry of the predicate licensor.\(^{24}\)

The setting of this parameter translates into a partial order on surface arcs, arcs representing surface constituent relations (see Section 3.6).\(^{25}\) A surface arc \(S\) is ordered by a parameter setting for a predicate licensing relation \(R\) if: (1) \(S\) is a \(R\)-licensor or \(R\)-licensee arc; or (2) \(S\) is a surface arc that structure shares with a \(R\)-licensor or \(R\)-licensee arc. This account of word order follows in the tradition of work which separates the mechanism for representing constituent structure from the mechanism which determines word order. Previous work on this approach includes Sanders 1967, 1974, Johnson 1974, Johnson and Postal 1980 and Gazdar et al. 1985.

Figure 4.5.2 represents a lexical entry for the \(\overline{P}\) in \(\textit{England}\) in which the anchor phrase modifies a \(\overline{V}\). The path \(\textsc{MODIFIER HEAD}\) is ordered before the path \(\textsc{MODIFIERMAC}\) because the \textsc{INT-THETA}

\(^{24}\)For example, internal-theta relations in passive lexical entries have the word order properties of canonical external-theta relations. It is assumed that this can be accounted for by a rule operating on the lexical entry for the verb.

\(^{25}\)The arcs are only partially ordered since: (1) the parameter settings include a free order option (no constraint on word order); and (2) the word order parameter only orders the anchor relative to its sister constituents — the word order parameter does not order one sisters of an anchor relative to another sister of an anchor. Although other word order constraints sometimes apply (e.g., case adjacency), it is not assumed that arcs need be totally ordered.
relation is set to Licensor First, the Head arc is an \textit{INT-THETA} licensor arc and \textit{MAC} is a (type of) \textit{INT-THETA} licensee arc. In Figure 4.5.2, the root \textit{MODIFIER} arc is ordered after the root \textit{HEAD} arc because the modification relation is set to Licensor Last. Each of these are surface constituent arcs. Figure 4.5.2 is a lexical entry for the finite verb \textit{laughs}. In this version of the VP-Internal-Subject analysis, the \textit{SPEC-CASE} arc of the INFL phrase, a surface arc, has the same value as (structure shares with) the \textit{EXT-THETA} arc of the $\downarrow$, a predicate licensee arc. The Licensor Last parameter setting results in the \textit{SPEC-CASE} arc being ordered before the \textit{HEAD INT-THETA HEAD} arc because: (1) the value of \textit{HEAD INT-THETA HEAD} is the licensor of the predicate relation \textit{EXT-THETA}; and (2) \textit{SPEC-CASE} is a surface arc which structure shares with the licensee of \textit{EXT-THETA}.\footnote{Note that the parameter setting mechanism discussed here works equally well for an empty category (ec) based version of the VP-Internal-Subject analysis e.g., Fukui and Speas 1986. Chains containing NP-traces (or other raising-type empty categories) would be given a treatment similar to that of sets of structure sharing in this account. The value of the surface arc would be the antecedent of an ec which is the value of some predicate licensor/licensee arc. The surface arc would be ordered on the basis of the predicate licensing relation which ordered the ec.}

\subsection*{4.5.3 The Predicate Licensor Determines Word Order}

For every predicate licensing relation, exceptional word order is lexically determined by the predicate licensor. This section provides examples from a number of different languages to illustrate this point for various type of predicate licensors.

Finnish is an SVO language which uses primarily postpositions (cf. Greenberg 1963). In the unmarked case verbal heads precede their internal theta licensees as in (31a), and most adpositional heads follow their internal theta licensees as in (31b). A few exceptional adpositions are prepositions as in (31c).\footnote{Examples were provided by Jussi Karlgren. Karlgren also notes that although SVO is the dominant order, other orders are possible, but marked, e.g., objects may precede their verbs in the marked case. In contrast, adpositions have fixed order with respect to their objects.} Finnish is a canonically an internal theta licensor first language, but adpositions are exceptional: they are canonically licensor last. There are a few exceptional adpositions like \textit{ilman}, which are licensor first. In each case, the theta assigning head determines all exceptional word order.

(31) a. Kasper lyö rumpua
    Kasper beats drum (part)
    ‘Kasper beats a/the drum’
FIGURE 48 Modifying Graph Anchored by in England
laughs
b. Pekka syö leipää auton luona
   Pekka eats bread car (gen) by
   ‘Pekka eats bread by a/the car’

c. Tulen New Yorkin ilman rahaa
   I arrive in New York without rahaa
   ‘I arrive in New York without money’

Modifier/Modifiee order in English varies considerably depending on
the modifier (the construction, category and lexical item).

(32) a. a city [in Maine]; the book [that I saw]
   b. I walked slowly; I slowly walked
   c. the blond gentleman; the blue marble
   d. president elect; eggplant parmiqiana
   e. the nearby lake; the lake nearby

In English, most of the adjectives with exceptional word order prop-
ties are obscure or belong to narrow semantic classes, e.g., legal, culinary,
etc. In Spanish, adjectives with opposite word order properties are more
common. In Spanish, adjectives like other modifiers canonically occur
predicate last, e.g., rojo and prenatal in (33a,b), but some adjectives
must occur predicate first, e.g., buen, bajo and gran in (33b,c,d). Others
like importante, malvado and cruel in (33e,f,g) can modify the noun
from either side.

(33) a. el libro rojo
   the book red
   ‘the red book’

   b. buen cuidado prenatal
       good care prenatal
       ‘good prenatal care’

\textsuperscript{28}Noun adjective order is often a possibility in Spanish, but it sounds marked, like
topicalization in English. Rojo may precede its modifiee in epithets without sounding
marked, el rojo amanecer (“red dawn”) and el rojo sangre (“red blood”). (Antonio
Moreno Sandoval, pc) In order to included special cases such as the language of
epithets, the word order order hierarchy would have to include: predicate relations,
part of speech, lexical items and special cases/senses of lexical items.

A complete account of Spanish word order should distinguish adjectives like rojo
which are marked when they precede the noun they modify from adjectives like grande
which never precede the noun. There are at least three ways of going about this: (1)
assume five word order parameter settings: free, licensor first, canonical licensor
first, licensor last (grande) and canonical licensor last (rojo); (2) mark both rojo
and grande as licensor last, but assume that the prenominal adjective construction
precludes grande by some constraint; (3) assume a theory of word order which states
what is allowed, not what is statistically likely or unlikely. In this case, rojo is a free
order adjective. The fact that it is marked in prenominal position is irrelevant.
c. al bajo costo
   at low cost
   ‘at low cost’

d. el gran evento
   the great event
   ‘the great event’

e. la importante noticia
   the important notice
   ‘the important notice’

e'. la noticia importante

f. la malvada/cruel criminala
   the wicked/cruel criminal
   ‘the wicked/cruel criminal’

f'. la criminala malvada/cruel

In English, most degree words are degree licensor first, but enough is licensor last.

(34)  a. too/very/quite angry; too/very/quite slowly
    b. angry enough; slowly enough

Demonstratives and numerals (Greenberg 1963 p. 86) need not have the same word orders relative to the noun, as evidenced by the Hebrew examples in (35). Hebrew demonstratives occur licensor last, as in (35a). Hebrew numbers canonically occur licensor first as in (35b), but echad occurs licensor last as in (35c). The determiner (numeral or demonstrative), not the noun determines word order. Variation within the same type of determiner (articles, numerals, etc.) seems rare in the world’s languages.29

(35)  a. ha-yom ha-ze
        def-day def-this
        ‘this day’

---

29 The Norwegian data in (i) and (ii) suggests that the article determines irregular word orders (the same phenomena occurs in Swedish, Danish Faroese and Icelandic). In most constructions, the definite article is a suffix and the indefinite article is a separate word that precedes the noun. However, this data is not conclusive evidence because definite determiners can be realized as separate words (den in Norwegian) which precede the noun when the NP contains a restrictive relative clause. See Fiorenta 1995 for further details about these constructions.

(i)   boken
     bok + en
     ‘the book’

(ii)  en bok
     a book
     ‘a book’
b. shtei yomim
two(gen) days
‘two days’
c. yom echad
day one
‘one day’

Further evidence that it is the predicate licensor that determines word order come from the examples in (36). In each example, two non-head predicate licensors predicate of the (projections of) the same head. Since the two predicate licensors have opposite word order properties, the head of the phrase cannot possibly be the element determining the word order of the phrase. I conclude that the predicate licensor determines word order.

(36) a. a great president elect (but a mediocre president)
   b. the disgusting eggplant parmigiana
   c. quite angry enough
   d. buen cuidado prenatal (= 33b)

The next subsection shows how idiosyncratic word order information can be accounted for in a parameter setting model.

4.5.4 A Word Order Hierarchy for English

Figure 4.5.4 is a default inheritance hierarchy,\textsuperscript{30} representing the canonical word order patterns for English predicate licensors. The word order parameter is set in the grammar and the lexicon for: (1) each language; (2) each type of predicate licensor which violates the word order parameter setting for that language; (3) each syntactic category which violates the parameter setting for that type of predicate licensor; and (4) each lexical entry that violates the word order parameter setting for that syntactic category. If the word order parameter is not set for a particular type of predicate licensor, syntactic category or lexical entry, the setting of its superclass is assumed as a default. In this way, both canonical and exceptional word orders are accounted for.

The Word Order Hierarchy states that English is predominantly a (predicate) licensor first language, and that, canonically, \textit{INT-THETA}, \textit{DEGREE} and \textit{QUANTIFICATION} relations follow this default. All such relations in (37) through (39) attest to this. The exceptional lexical

\textsuperscript{30}Default inheritance hierarchy are used in many areas of study involving knowledge representation. Default inheritance hierarchies have previously been used to associate linguistic rules to elements in a hierarchy based on parts of speech in Wilensky 1981, Flickenger et al. 1985, Watanabe and Johnson 1990, Johnson and Watanabe 1991 and Flickenger and Nerbonne 1992.
Predicate Licensing / 107

Predicate Licensing
item enough is a degree word which must be specifically marked licensor last, e.g., (40)

(37) a. Mary [saw John]
b. in the house
c. Matilda was [annoyed that Clyde ate the last cookie]

(38) a. You can’t be [too angry] with the fish merchants
b. the [very uncomfortable chair]
c. He was [quite hungry]

(39) a. The dog
b. five dogs
c. Every dog

(40) One can never get [angry enough] in these situations

The word order parameter is Licensor Last for the EXT-THETA relation in English; to my knowledge there are no exceptions to this setting, e.g., John is the EXT-THETA licensee of laugh in each example in (41). This word order parameter must be explicitly set in the grammar for English because THETA, the immediate superclass of EXT-THETA, has a Licensor First setting.

(41) a. John laughed
b. I saw [John laugh]
c. I knew [John to have laughed]

The Word Order Parameter must be set to Predicate Last for modification in English, since this setting is in conflict with the default. P’s, relative clauses, purposes clauses and other modifiers follow this default, as shown in (42). The Word Order Parameter must be set to: Free Order for adverb modifiers of verbs, as shown in (43); and Licensor First for adjectives and adverbial modifiers of adjectives, as in (44).

(42) a. the book [with the bright orange cover]
b. The astronaut jumped rope [on the moon]
c. the book [that I saw t]
d. I ran the race [Pro to prove that I am an athlete]

(43) a. I walked slowly; I slowly walked
b. I ate hungrily; I hungrily ate

(44) a. the slightly unusual woman
b. the light green book

The word order parameter must also specifically be set for the lexical entries for certain modifiers: (1) Licensor First for a few idiomatic P’s, as in (45); (2) Free Order for the adjective nearby, as in (46); and Licensor
Last for the adjectives *elect, general, public, parmigiana, almondine, across, above* and *agog* in the particular senses used in (47).

(45)   a. the out of shape athlete  
       b. the in the know politician  

(46)   the nearby lake; the lake nearby  

(47)   a. president/mayor/governor elect  
       b. secretary/post-master/surgeon general  
       c. notary public  
       d. chicken/eggplant parmigiana  
       e. beef/chicken/lamb almondine  
       f. with arms across (Websters Third New International)  
       g. the sky above  
       h. the eyes agog  

4.5.5 Apparent Counterexamples

Certain closed class items like clitics and indefinite pronouns need special treatment for word order under most accounts, including my own. These items are the only predicate licensees I am aware of which effect word order in English and Spanish. As shown above, most other closed class items, as well as, all open class items follow the generalization that word order variation is determined by the predicate licensor.

In English, compound indefinite pronouns, unlike other nouns, must precede the adjectives modified by them, as in (48) (David Johnson, personal communication). In these examples, the predicate licensee determines exceptional word order. Therefore these examples appear to be problematic for my account.

(48)   a. anyone/someone ill  
       b. anything/something good

A compound indefinite pronoun consists of a quantifier and a noun from the set *one, thing, body, place, where.*\(^{31}\) I assume that the pronouns in (48) are simultaneously quantification licensors and modification licensees. As quantifiers, they predicate of the \(N\) to the right and as modificies of adjectives, they would normally occur to the left of the adjective. Since the adjective is part of the quantified \(N\), the pronouns must simultaneously occur both before and after the adjective. This is impossible, so one predicate relation must have priority. I assume that the quantification relation takes priority because it is ‘higher’ in the graph, i.e., the quantification-licenssee arc predicate commands (see

\(^{31}\) *Where* is not actually a noun, although it functions like one when part of an indefinite pronoun, e.g., with respect to modification by adjectives.
Chapter 8 for a definition of PREDICATE COMMAND) the modification licensee arc in Figure 48. The fact that these indefinite nouns can only be modified by a single adjective (except for coordinated adjectives) is not surprising if it is assumed that the lexical entry for indefinite pronouns includes the position for the adjective modifier.

Spanish object pronouns (clitics) occur to the left of the verb assigning them an internal theta role, as in (49). The Internal-Theta relation is canonically Predicate First in Spanish. These pronouns appear to be a problem for my theory of word order. Previous P & P accounts of clitics do not derive surface positions by parameter setting. The surface positions of Spanish clitics would violate the head parameter (see below). P & P accounts of clitics (e.g., Burzio 1986) rely on case, theta and binding considerations among others. In short, clitics are treated as being more closely connected to the verb than other lexical items and their final surface positions depend on a variety of factors. There is a large
literature on clitics which I will not attempt to review here. I assume compatibility with previously proposed analyses. My theory of word order is no worse off than previous P & P analyses with respect to deriving the relative word order of clitics and verbs.

(49)  a. Los libros me gustan
       The books me are pleasing
       ‘I like the books’
   b. Usted me trae el almuerzo
       You me bring (third person singular) the lunch
       ‘You are bringing me lunch’
   c. Le fatiga el trabajo a mi hermano
       Him fatiques the work to my brother
       ‘The work fatigues my brother’

4.5.6 Comparison with Other P & P Accounts of Word Order

The Word Order Hierarchy contrasts with the two main ways parameters are set in previous P & P accounts. Stowell 1978, Travis 1983, and others claim that a parameter is set for an entire language. This does not permit exceptions. Wexler and Manzini 1985 argued that parameters are set for particular lexical items with respect to Binding Theory Domains of pronouns and anaphors. An approach in which parameters are set exclusively for lexical items would miss any typological settings of parameters for an entire language. A default inheritance hierarchy gives the best of both worlds, a systematic way of representing (a) canonical word order, e.g., “English is a predicate first language”; (b) systematic exceptions, e.g., “English modifiers are canonically modifier last”; “English adjectives are canonically modifier first”; and (c) “true” exceptions like the adjectives elect and parmigiana. Counter examples may be found if the parameter being modeled is linked to a different phenomenon than the linguist proposes. For example, elect and parmigiana are unproblematic for my account, but indefinite pronouns pose a potential problem, as discussed above.

Stowell 1978

and Travis 1983, among others, account for the order of heads with respect to their complements by means of the head parameter, which basically amounts to my word order parameter for the Internal-Theta relation with the following exceptions: (1) they do not allow a free order setting; and (2) they assume the parameter is set for the entire language, thus ruling out (for example) the Finnish data discussed above.

Travis 1983

argues that the relative word order of subjects and finite verbs (in
English) is determined by the interaction between the Head Parameter and the Case Adjacency Principle. She proposes that: (1) the verb must (immediately) follow INFL (the head of 7) because English is a head first language and (2) the subject must occur adjacent to INFL in order to get case. Therefore the subject must immediately precede INFL, i.e., the inflected verb.

In Travis’ account, it is unclear in what sense INFL precedes the verb when INFL manifests as a verbal affix. For example, in the sentence John laughs, the word laughs is arguably a combination of a verb morpheme and an inflectional morpheme. If the morphemes laugh and -s are considered to have any linear order at all, clearly -s follows laugh. Travis’s account provides no explanation of why the string *Laughs John is ungrammatical. Even if I assume (counterfactually), that there is some sense in which a verbal affix precedes the verb, this would not explain why the subject may follow the finite verb in Spanish, a predominately SVO language (like English), e.g., Example 50. If Travis’ account were correct, I would not expect these alternative word orders.

(50) a. Yo hablo español
   I speak Spanish
   ‘I speak Spanish’
   a’. *Hablo español yo
 b. El baila
   he dances
   ‘he dances’
 b’. Baila el
 c. Yo pienso
   I think
   ‘I think’
 c’. Pienso yo

Travis’ theory of word order also fails to account for the external-theta relation in SOV, VSO and VOS languages or in languages with no case adjacency requirement.

While I assume that case adjacency is not relevant to subject verb order in English, it may be relevant to the order of complements. For

---

32Travis’ account is based on a 1982 class lecture by Noam Chomsky.
33I assume that an affix A combines with a stem S in the lexicon and that there is no linear order relation between A and S that is relevant to syntax. I do however assume a hierarchical relation between phrases based on the distinction between certain affixes, e.g., INFL and verbal stems.
34One might propose movement analyses for these examples in order to ‘regularize’ Spanish word order. However, this move would make word order parameter settings un falsifiable, unless independent evidence could be found for a movement analyses.
example, I follow previous P & P accounts in assuming that if a verb takes one \( \mathbf{N} \) and one \( \mathbf{P} \) complement, the \( \mathbf{N} \) complement occurs first, barring considerations like heavy NP shift, as shown in (51). Plausibly, this is due to the requirement that the object of the verb be adjacent to the verb in order to get case.\(^{35}\)

(51)  a. Mary presented a book to Zelda.
      c. Mary presented to Zelda a golden book with pretty pictures on
         the outside and strange writing on the inside.

I also assume P & P accounts of variation from canonical order which result from some morphological or syntactic operation, e.g., passive, heavy NP shift, topicalization, the Koopman 1987 account of word order variation in languages like Vata, etc.

In Summary, my account provides a principled description of a wider range of word order phenomena than previous P & P accounts. I have shown that Travis’ account of the relative order of subjects and finite verbs is descriptively and explanatorily inadequate because: (1) Travis fails to motivate the assumption that finite inflectional suffixes precede verbs; (2) Travis fails to account for Spanish word order phenomena; and (3) ‘Travis’ account does not apply to SOV, VOS and VSO languages or languages without a case adjacency requirement.

### 4.5.7 Summary

Word order parameters are set for types of predicate licensing relations, syntactic categories of the predicate licensors and individual lexical items with respect to their roles as predicate licensors. It follows that:

1. Varying the predicate licensor of a constituent can lead to variations in word order, but varying the head does not, unless the head is the also the predicate licensor; and

2. In phrases with one head and more than one predicate licensor of the same type, the predicates may have different relative orders to the head. The head cannot determine word order in these phrases.

\(^{35}\)Paul Postal (p.c.) notes the following counter example:

i.  a. John gave back the book to Zelda.
    b. *John gave the book to Zelda back.
    c. *John gave back Zelda the book,
    d. John gave Zelda the book back.

This is problematic for all P & P accounts (including this one) because there is no satisfactory way to characterize these data in terms of case adjacency. If \( \text{give back} \) is a complex verb that assigns case, then the ungrammaticality of (c) is unexplained. If \( \text{give} \) is distinct from the particle \( \text{back} \) for purposes of case assignment, then the grammaticality of (d) is unexplained.
This position rests on an account of systematic exceptions to canonical settings of word order parameters. In contrast, previous P & P theories of parameter setting either: (1) set parameters for the entire language, thus not accounting for any exceptions; or (2) set parameters for individual lexical items, thus not accounting for canonical word order.

This analysis has the following implication for the structure and content of the lexicon: the word order parameter should be set for the predicate licensor of a phrase. Both canonical and exceptional word orders can be captured using a default inheritance hierarchy of word order parameter settings.

4.6 Alternations Among Predicate Relations

This section shows that when a given word or phrase X can be a licensor in more than one type of predicate licensing relation, X imposes the same selection restrictions on its licensee (Y) in each relation. These correspondences may be viewed as alternations, much like the dative alternation and other alternations among arguments of verbs. When productive, these alternations may be modeled as lexical redundancy rules, much like the way passive verbs are typically handled in (for example) Lexical Function Grammar.

Postcopular and prenominal positions for adjectives should be viewed as a productive modifier/external theta licensor alternation, given the following analysis. Copular verbs like be, seem and remain are assumed to be subject to subject raising verbs following Postal 1976 (p.158, note 18), Pullum and Wilson 1977, Stowell 1978, among others. Given C, a copular verb; S, the subject of C; and A, the adjectival complement of C, S is both the specifier of C and the external-theta licensee of A. A given adjective A imposes the same selection restrictions on some N regardless of whether the relation EXT-THETA(A,N) or the relation MODIFICATION(A,N) holds. The adjectives blue, long and angry select the same type of head nouns in both (52) and (53). Since most adjectives participate in this alternation, a lexical redundancy rule may be assumed. The exceptional adjectives like former or afar which only occur in one position or the other (e.g., (54)), would be specially marked in the lexicon. In early transformational grammar, this same correspondence was captured by transforming a relative clause containing a copula, e.g., the book that is blue to derive a noun phrase in which the adjective modifies the noun, e.g., the blue book. As this sort of transformation is not available in more recent versions of P & P, it is unclear how other P & P theories can capture the external-theta/modifier correspondence.

(52) a. The book is blue; ~The minute is blue
b. The day seems long; \(\sim\)Sincerity seems long  
c. The gorilla is angry; \(\sim\)The rock is angry

(53) a. The blue book; \(\sim\)The blue minute  
b. The long day; \(\sim\)The long sincerity  
c. The angry gorilla; \(\sim\)The angry rock

(54) a. The former mayor; *the mayor was former  
b. The leading lady; *The lady was leading  
c. The door was ajar; *The ajar door  
d. The victims were alive; *The alive victims

When a complement is extracted from a relative clause, the gap is assigned a theta role (putting aside the issue of whether a structure sharing or empty category analysis is appropriate.) The selection restrictions associated with this gap are imposed on the noun modified by the relative clause, as in (55) and (56). These relative constructions may be viewed as a vehicle for a theta license/modification license alternation.

(55) a. I read the book; \(\sim\)I read the frog  
b. I drank the juice; \(\sim\)I drank the steak

(56) a. The book that I read; \(\sim\)The frog that I read  
b. The juice that I drank; \(\sim\)The steak that I drank

When a modifier is extracted in a relative clause, the filler of the gap modifies the clause from which it is extracted, and the relative clause modifies the filler of the gap. This may be viewed as a modifier/modifiee alternation. The same selection restrictions must be satisfied by the clauses in (57) and (58). The (a) examples demonstrate that temporal modifiers representing specific points in time do not modify clauses that represent timeless facts.\(^{36}\) The (b) examples demonstrate that place modifiers exert this same restriction.

(57) a. I ate cheese (last) Thursday; \(\sim\)Two plus two equaled four (last) Thursday  
b. There was a pothole every three feet (along the road); \(\sim\)Two plus two equaled four every three feet (along the road)

(58) a. The Thursday that I ate cheese; \(\sim\)The Thursday that two plus two equaled four

\(^{36}\)The selection tests for predicate licensors predict that the gap should select its clause, but the whole relative clause (including the gap) should select the gap's filler. Both types of selection seem to occur. The relative clause does not allow proper nouns like Thursday. Thus (58a) would be ill-formed without the determiner the, which converts Thursday into a common noun. The selection restrictions between Thursday and the clause have already been discussed.
b. Every three feet (along the road) where there was a pothole;  
\sim Every three feet (along the road) where two plus two equaled four

The above data supports the position that a single lexical item (e.g., an adjective) may impose the same selection restrictions regardless of whether it is a modification licensor or a theta licensor, and further that a relative clause construction with a non-modifier gap, in effect, passes up to its modifiee the selection restriction imposed on its gap by a theta licensor contained in the relative clause. In a relative clause construction with a modifier gap \( G \), the relative clause (a modifier) is constrained by selection restrictions which are in effect imposed by its modifiee, the filler of \( G \). This filler and gap of these relative clauses represents a modifier/modifiee alternation. The gap is a modifier of some \( X \) (a modifiee) and the filler is a modifiee of that same \( X \) (a modifier).

This section discussed alternations between modification and theta relations, both types of predicate relations. Correspondences between the selection restrictions imposed by the predicate licensors seem to reflect meaning correspondences between phrases which instantiate these licensing relations. For example, the gap bears the same theta role to the subordinate verb, as the modifiee bears to the relative clause.

4.7 Summary

This chapter has motivated a distinction between heads (constituent licensors) and predicate licensors. The head determines the category of the phrase. The predicate licensor selects its sister constituents and determines word order properties. The view that a selection restriction on \( X \) is a presupposition about what \( X \) can be provided the basis for ways of testing for selection. Exceptional word order properties provided a way of testing which item determines word order. In both cases, it turned out that the predicate licensor, not the head, determined these properties. Predicate licensing relations were also used to formulate new approaches to parameter setting, selection by and of determiners and alternations between modification and theta licensors.

Much previous work in the P & P literature conflates heads and predicate licensors, in particular work based on Abney 1987. It would seem that these accounts would define functional head like predicate licensor and thematic head like constituent licensor. All functional head approaches fail to show that the functional head determines the category of the phrase. This chapter has provided evidence that these approaches should be rejected and a predicate licensor/head distinction should be adopted.
Agreement Licensing Relations

5.1 Introduction

In P & P theories, an abstract AGREEMENT (e.g., case) relation $A(X, Y)$ may hold between constituents $X$ and $Y$. $A(X, Y)$ is realized as a set of constraints on the cooccurrence of $X$ and $Y$ where morphological agreement, adjacency and other linguistic phenomena are the basis for these constraints. In this chapter, abstract agreement relations are modeled as licensing relations. The constraints representing “realizations” of these agreement relations are also discussed. Morphological agreement is modeled using a GBUG version of standard unification-based approaches to morphological agreement. Adjacency constraints are represented using an immediate precedence operator $\prec$.

An agreement licensing relation $A(X, Y)$ represents the satisfaction of a set of conditions for some predicate licensing relation $P(Z, Y)$ to be well-formed ($X$ and $Z$ may, but need not be the same constituent). P & P’s case filter presumes that theta relations are well-formed only if the theta recipient (or licensee) is also a case recipient (or licensee). It is assumed that all agreement relations are similarly motivated following recent P & P work which assumes that abstract case is a type of agreement relation (cf. Koopman 1983, Mahajan 1990, Chomsky 1991: p. 17, and much subsequent work in the Minimalist Program, e.g., Chomsky 1992: pp. 10–11). In the account presented here, the agreement relations between determiners or adjectives and their head nouns are taken as well-formedness conditions for the predicate licensing relations Quantification and Modification, holding between the determiners/adjectives and the head nouns. The two agreement licensing relations considered are:

- $CASE(X, Y) \leftarrow \text{P & P’s abstract case relation, where } Y \text{ is some } \overline{N}$
  or Gerund and $X$ is some case assigning item (e.g., verb, adposi-
tion or complementizer). $X$ is the head of a phrase containing $Y$.

The case relations COMP-CASE and SPEC-CASE are distinguished by whether $Y$ is in a complement or specifier relation with $X$. HEAD is the only case licensor feature. COMP-CASE and SPEC-CASE are case licensee features, corresponding to case licensing relations with the same names. The constituent licensor features COMPLEMENT and SPECIFIER respectively subsume COMP-CASE and SPEC-CASE.

- $N$-AGR($X,Y$) — an agreement licensing relation where $X$ is a determiner or adjective and $Y$ is some projection of the head noun of the $\overline{N}$ containing $X$ and $Y$. $N$-AGR-SPEC and $N$-AGR-ADJ each names an $N$-AGR - LICENSOR feature and the corresponding $N$-AGR relation. ADJUNCT $\subseteq$ N-AGR-ADJ and SPECIFIER $\subseteq$ N-AGR-SPEC. HEAD is the only $N$-AGR - LICENSEE feature.

This chapter takes a Government and Binding (GB) Theory approach to case (and agreement), in the following respects: (1) Case is assumed to be a SURFACE constituent relation in the sense of Section 3.6; and (2) Case licensees may be either complements or specifiers.

Drawing on evidence from Johnson and Lappin In Press, I argue against the Minimalist program approach which assumes that case relations hold at Logical Form (LF) and that all agreement licensees (including case) are either specifiers or adjuncts.

5.2 Agreement Filters

A function CONSTRAINT-ON is assumed that maps each agreement relation $A(X,Y)$ onto the predicate relation $P(Z,Y)$ that $A$ constrains. The AgreementFilter in Figure 52 represents the constraint between a given $A$ and the predicate relation $P = CONSTRAINT-ON(A)$. The interpretation of this constraint is as follows: An item $X$ can only be the value of a predicate licensee feature for predicate licensing relation $P$, if that same $X$ is also the value of an agreement licensee feature, for agreement licensing relation $A$, and CONSTRAINT-ON($A$) = $P$. Literally, the constraint says that given a path consisting of a predicate

---

1This assumption depends on a raising to object analysis of the $\overline{N}$ + infinitive complement of the believe class of verbs. Exceptional Case Marking analyses are barred because in such analyses, a verb assigns case to a non sister constituent. See Chapter 9 for further discussion.

2I have deviated from the convention in which the names of morphological cases (nominative, accusative, etc.) are used for abstract cases because that practice seems to obfuscate the abstract/morphological case distinction.
Agreement Licensing Relations / 119

\[ P = \text{Licensee}(X) \rightarrow P = \text{Licensee} A = \text{Licensee}^{-1} \]
\[ \land \text{CONSTRAINT-ON}(A) = P \]

Figure 52: The Agreement Filter

\[ \text{Theta Category}(\text{Nom}) \rightarrow \text{Theta Case}^{-1} \]

Figure 53: Case Filter

licensee arc \( A_p \), it must be possible to travel upward from the target of \( A_p \) along one agreement licensee arc \( A_A \) arriving at the source of \( A_A \). The upward traversal is indicated by the the \(-1\) superscript on the inverse path \textsc{agreement-licensee} \(^{-1}\) following previous work in Stratified Feature Grammar (SFG). The GBUG version of the case filter, repeated as Figure 53, is an instance of the agreement filter in which \textsc{constraint-on}(\text{case}) = \text{theta} \(^{3}\). If a given \( \overline{N} \) is the value of a \text{theta} arc, it must also be the value of a \text{case} arc. Both \( \overline{N} \)'s in Figure 54 satisfy the case filter. The \( \overline{N} \) John is the value of a \text{mac} arc, which is simultaneously a \text{comp-case} and \text{int-theta} arc; and the \( \overline{N} \) Mary is the value of one \text{spec-case} and one \text{ext-theta} arc. In both instances, it is possible to travel upward one case arc from the target of a theta arc. In the former case the case and theta arcs are one and the same; in the latter case they are two distinct arcs.

Figure 55 is a representation of a passive sentence. The chain (A-chain) from the \text{int-theta} position (arc) of the verb to the specifier (\text{spec-case}) position of the verbal inflection is represented as a set of structure-sharing arcs in a FS, rather than as a set of nodes in a tree, as in most P & P work. The case filter is a logical constraint on the possible set of arcs that can share an \( \overline{N} \) (or \( \overline{Gerund} \)) value. Figure 55, like Figure 54, satisfies the case filter since it is possible to traverse up one case arc (\text{spec-case}) from the target of the \text{int-theta} arc. GBUG’s case filter applies equally well to passive (and raising) and canonical (e.g., \text{mac}) instances of case assignment. \(^{4}\)

\(^{3}\)The individual relations \text{spec-case} and \text{comp-case} do not have distinct values of \text{constraint-on} in English. In English, subject to object raising (or Exceptional Case Marking) verbs like believe makes these more specific mappings untenable. The phrase \text{insects} in I believe \\text{insects} to carry diseases receives an external-theta role from carry and \text{comp-case} from believe. Believe does not assign \text{insects} any theta role. While in most instances \text{comp-case} is assigned to \text{int-theta} licensee, here \text{comp-case} is assigned to an \text{ext-theta} licensee. Usually \text{spec-case} is assigned to an \text{ext-theta} licensee. Mappings from specific case relations to specific theta relations are logically possible for other languages, particularly languages like Russian in which morphological case distinctions closely match theta distinctions.

\(^{4}\)Figure 53 does not seem to account for examples like There suddenly arrived three
Figure 54: Mary saw John
Figure 55: Mary was seen
This version of abstract agreement is based on Baker's visibility conditions (cf. Baker 1988, pp. 111-123). According to Baker, N's must meet some set of visibility conditions (which vary from language to language) in order to be "visible" for theta marking. Baker lists the possible visibility conditions including: (1) adjacency; (2) morphological case; (3) person/number/gender agreement; and others (e.g., noun incorporation). In the account presented here, the term (abstract) case is defined to include all visibility conditions as its realizations. As discussed below, N-AGR (and, in theory, other abstract agreement relations) have some of these same possible realizations.

In summary, the abstract agreement relation $A(X, Y)$ holds only if $R(X, Y)$, some realization of $A(X, Y)$ holds. For some $P = CONSTRAINT-ON(A)$, $P(Z, Y)$ is well-formed if $A(X, Y)$. Realizations of abstract agreement in the form of adjacency, morphological case and agreement are discussed in subsequent sections.

5.3 Adjacency

Adjacency constraints require that the licensor and licensee be in adjacent positions in a string, and in some languages, in a particular word order. For example, English case licensees must be adjacent and to the right of COMP-CASE licensors, as exemplified in (1). Figure 56 is a logical representation of English's COMP-CASE adjacency constraint. The symbol $\prec$ is a variation on the precedes symbol $\prec$, meaning "immediately precedes" rather than simply "precedes". $X \prec \prec Y$ if $X \prec Y$ and there is no $Z, Z \neq X$, such that $X \prec Z \prec Y$.

Words or larger constituents can be COMP-CASE licensors. In the MAC relation between simple transitives/prepositions and their objects, the COMP-CASE licensors are words ($X^0$s) (as in Figure 54). Examples of COMP-CASE relations in which larger constituents are possibly COMP-CASE licensors include MAC relations whose licensees are second objects of ditransitive verbs and objects of verb particle constructions in which the verb and particle are adjacent. In Figures 57 and 58, each licensee of COMP-CASE (or MAC) should be interpreted as occurring immediately adjacent and to the right of its head. This binary branching analysis of ditransitives, exemplified in Figure 58, is contro-

*men with funny hats on.* If *three men with funny hats on* bears a theta role, and *there* bears case, then the case filter must be revised so that the expletive argument relation is factored in. Under this view a theta-marked $\bar{N}$ linked to a case marked expletive "acts" like a case licensee with respect to the case filter by virtue of the expletive argument relation. An alternative view (which I prefer) is that both the case and theta role are assigned to *there* and *there* is referentially dependent on *three men with funny hats on*, which bears no case or theta role directly.
versial. In later chapters, I will assume a flatter analysis in which the verb is a sister to both following \( \bar{N} \) constituents. For this latter type of analysis, a more complex version of the adjacency constraint needs to be formulated.

(1) a. John read the book.
   b. *John read slowly the book.
   c. Mary believed Sally to have stolen the records.
   d. *Mary believed whole-heartedly Sally to have stolen the records.

In this account, no adjacency requirements are assumed for \( \text{SPEC-CASE} \) in English. Sentential adverbs like \text{probably} may occur between the inflected verb and its subject, as in (2). These adverbs modify the sentence and therefore must occur outside the \( \overrightarrow{\bar{V}} \), and within the \( \overrightarrow{\bar{T}} \). Assuming that INFL (or an INFL element in P & P theories that divide INFL) assigns SPEC-CASE, adjacency does not seem to be a requirement. One can imagine case-adjacency requirements between an \( \overrightarrow{T} \) and the subject. However, an adjacency requirement between two constituents \( A \) and \( B \) is vacuous if all possible intervening constituents can be attached by adjunction to \( A \) or \( B \). In other words, it must be possible to identify a string of words that is ruled out by this constraint or else the constraint has no motivation. Similarly, possessive \( \text{SPEC-CASE} \) licensees may be separated from their head nouns by any number of adjectives and other prenominal modifiers, as in (3).

(2) a. John probably read the book.

5If it can be shown that each sentence allows a finite number of sentential adverbial positions, an adjacency constraint may not be vacuous. In such an account, the ill-formedness of (1a) would be considered a case filter violation. The well-formedness of the two sentential adverbs in (1b) would also have to be accounted for. See for example, Jacobson 1964, Greenbaum 1969 and Huang 1975 for some illumination on the different classes of sentential adverbs.

(i.) a. *John unfortunately probably won’t read this letter
   b. John, however, probably won’t read this letter

Some adverbial modifiers of \( \overrightarrow{\bar{V}} \) can occur between subject and the inflected verb and/or auxiliary as in (1). These are unproblematic when INFL is instantiated as inflection on the verb if it is assumed that the modified \( \overrightarrow{\bar{V}} \) abstractly follows INFL. However this approach would fail to explain cases such as (1a, b).

(ii.) a. Mary suddenly could not be found anywhere
   b. We soon will leave town
   c. John slowly left the building
   d. Fred purposely dropped his wallet
Figure 57  Borimir messed up the plan
Figure 58: Frodo gave Borimir a dollar
b. They unfortunately won’t read this letter.

(3) a. Mary’s long-awaited phrenology test.
   b. Cecil’s big hairy legs
   c. Rome’s slow and tedious evacuation
   d. The bunny rabbit picture

In (2), INFL is the \textit{SPEC-CASE} licensor, but is not adjacent, nor even a sister to the surface subject. If \textit{SPEC-CASE} is indeed a surface relation as proposed, then the surface subject would have to be a \textit{SPEC-CASE} sister to INFL. (Licensors and licensees must be sisters and surface relations must be between surface constituents.) A principled solution to this problem is illustrated in Figures 59 and 60. In the FS anchored by the sentential adverb \textit{probably}, the value of \textit{HEAD SPEC-CASE-PRIME} is shared with the value of \textit{SPEC-CASE}. It is assumed that \textit{SPEC-CASE} \textit{⊈} \textit{SPEC-CASE-PRIME} and \textit{SPEC-CASE-PRIME} is a non-surface arc. \textit{SPEC-CASE-PRIME} \textit{∪} \textit{SPEC-CASE} = \textit{SPEC-CASE-PRIME} when the FS anchored by \textit{probably} combines with the FS for \textit{laughed} to yield the FS for \textit{probably laughed}. The FS for \textit{John} can then be added to yield \textit{John probably laughed}. In the chain of structure sharing \textit{SPEC-CASE} arcs in the FS for \textit{John probably laughed}, only the highest arc (the arc closest to the root) is a surface arc. The only surface \textit{SPEC-CASE} relation is \textit{SPEC-CASE([probably laughed], John)}.

An \textit{N-AGR} relation is proposed for each determiner and the adjacent nominal constituent, and for each adjective and the adjacent nominal constituent in examples like (4). An adjacency requirement posited for \textit{N-AGR} relations would not be compatible with intervening items occurring between the \textit{N-AGR} licensor and the head noun. In these examples, each \textit{N-AGR} licensor is also a predicate licensor, a quantifier or a modifier. If it can be shown that the predicate licensee is in fact the following bracketed constituent and not just the head noun, this provides a reason to believe that some sort of adjacency constraint holds. For example, (4a) describes a big member of the set of hairy spiders, and not a hairy member of the set of big spiders, the latter reading being barred by the adjacency constraint. In contrast, sentential adverbs that intervene between \textit{SPEC-CASE} licensors and licensees modify the whole sentence rather than some constituent that excludes the subject. For example, (2a) means “It is probably the case that John read the book” and there

\footnote{\textit{Probably} must precede its modifiee and \textit{John} must precede its \textit{EXT-THETA} licensor (the \textit{\textbullet}). These two precedence constraints are compatible with two possible word orders: \textit{John probably laughed} and \textit{Probably John laughed}. Both are well-formed and both are partially described by Figure 60.}
Figure 59 FSs anchored by John, probably and laughed
John probably laughed
is no acceptable interpretation in which probably read the book is predicated of John. Contrast (4d) and (4e). The latter requires a special intonation to be acceptable, similar to sentences like I gave to John the book. (4d) is comfortable without this special intonation, presumably because the adjective easy is adjacent to its modifiee man. Prenominal modifying adjectives can themselves be modified by the pred-first degree word very, but not the pred-last degree word enough as shown in (4g,h).

In the latter case, enough blocks adjacency between the adjective big and its modifiee. Similarly, pred-first, but not pred-last modifiers of numbers are acceptable when numbers function as determiners, as in (4i,j).

Based on this evidence, I posit an adjacency constraint for the N-AGR relation.

(4)  a. The [big hairy [spider]]
    b. Rome’s [slow [partial [evacuation]]]
    c. The [one [tiny [bug]]]
    d. An [easy man to please]
    e. An [easy to please] man
    f. The [big [strong [durable [plastic [bag]]]]]
    g. The [[very big] [strong [durable [bag]]]]
    h. *The [big enough] [strong [durable [bag]]]
    i. [[[Approximately three] [syntactic [problems]]]
    j. *[[Three in number] [syntactic [problems]]]

Chomsky 1981

(pp. 94–95) suggests that the case adjacency requirement is parameterized. Clearly, this can be generalized to an agreement adjacency parameter to cover all abstract agreement relations. In Chomsky’s account, the case adjacency parameter is set to positive in English and negative in many free word order languages. This seems to contradict the above analysis under which agreement adjacency requirements hold for COMP-CASE, but not SPEC-CASE in English. As discussed in Section 4.5, this book assumes that parameters are set for languages, licensing relations, parts of speech and lexical items using a default inheritance hierarchy. In English, COMP-CASE and N-AGR have a plus setting for the agreement adjacency parameter and SPEC-CASE has a minus setting.

5.4 Morphological Agreement

It is assumed that morphological agreement encompasses linguistic phenomena in which the grammatical cooccurrence of two constituents is de-
dependent on the morphological form (relative to a conjugation/declension paradigm) of one or both of these constituents. For abstract case, morphological case (MORPH-CASE) and person/number/gender agreement (MORPH-AGR) are treated differently. For both realizations of abstract case, only licensees that take a particular morphological form may occur with a particular case licensor. However (in the languages discussed here), a given case licensor is not itself marked for morphological case. There are not typically accusative and dative instances of the same verb that require accusative and dative licensees respectively, e.g., both eat and eats require their SPEC-CASE licensee to have a NOMINATIVE value for MORPH-AGR. Different morphological forms of a case licensor may require the licensee to have different MORPH-AGR properties. In this sense, each licensor “agrees” with its licensee. For example, the finite inflection on eat is different than the finite inflection on eats, the former compatible with third person singular Ns, the latter compatible with all other Ns. In most instances, MORPH-CASE and MORPH-AGR features are treated the same way when they are realizations of the N-AGR relation. The N-AGR licensor and licensee must agree with respect to both features. For example, Russian adjectives and the nouns they modify typically agree in case, number and gender. However, Section 5.5 describes one instance of N-AGR in which a determiner requires MORPH-CASE and MORPH-AGR feature values for its N-AGR licensees that are different from the MORPH-CASE and MORPH-AGR values of the constituent it anchors.

Various treatments of morphological agreement have the same basic notions at their core. Lexical items are assumed to “agree” if their morphology is compatible. FS unification provides one way of making this notion precise (See for example Shieber 1986). If an anchor X of a phrase P requires that some constituent C in P has certain morphological agreement features (and values), this information is specified in the lexical entry for X, a FS representation of P in which C is marked with the required morphological features. For example, the lexical entries for eat and eats in Figures 61 specify that their subjects have the appropriate MORPH-CASE and MORPH-AGR feature values. The verbs differ with respect to the value of MORPH-AGR. Both verb forms make the same requirement of their objects, i.e., that the object is in accusative case. In the proposed account, morphological agreement

---

8The generalization that both these verb forms require the same accusative case is captured in a hierarchical lexicon with multiple inheritance. Both verbs are simultaneously instances of the verb eat, which itself belongs to the set of transitive verbs. FSs representing the notion “transitive verb”, “specific features of eat”, “third person
is, by definition, a realization of agreement licensing relations (or abstract agreement) because morphological agreement explicitly constrains the possible values of agreement licensee features. Only FSs that satisfy morphological agreement conditions can unify with the value of the agreement licensee feature, because FSs representing morphological agreement restrictions are part of these feature values.

Shieber 1986

(pp. 26-27) provides a simple example of subject verb (i.e. subject INFL) agreement which I have modified for the GBUG approach. In Figure 62, the FS representing Other unifies with the value of the SPEC-CASE feature in the FS representing sleeps resulting in a FS representing Other sleeps. The FS for Other cannot unify with the value of SPEC-CASE in the FS for sleep as shown in Figure 63. The Other FS contains the feature value pair MORPH – AGR : 3-PERS-SING, and 3-PERS-SING does not unify with –3-PERS-SING, the value of the path SPEC-CASEMORPH – AGR in the sleep FS.9

Morphological case (MORPH-CASE) is included in the value of SPEC-CASE above. English finite verb inflection requires its SPEC-CASE value to have a NOMINATIVE value for MORPH-CASE, verbs and prepositions (in English) require their objects to have ACCUSATIVE values for MORPH-CASE.10 English pronouns him, her, me are accusative case; I, he, she are nominative and most other words are NOMINATIVE|ACCUSATIVE. The FS for she in Figure 64, but not the FS for her can unify with the value of SPEC-CASE in the FS for sleeps.

Most instances of morphological agreement realizations of N-AGR can be accounted for in much the same way. For example, Spanish determiners, adjectives and nouns agree in gender and number, as in (5). In Spanish, I assume that N-AGR-SPEC ⊆ QUANTIFIER and N-AGR-ADJ ⊆ MODIFIER.11 Figure 65 illustrates the merging of

\begin{itemize}
  \item singular verb', etc. are found in the lexicon. Members of each class is derived by unifying the FSs representing all superclasses.
  \item MORPH-CASE has a PARTITIVE case value for objects of of in the modifying nouns such that the prepositional object is the possessor (e.g., this book of mine/its/their). This case is homophonous with the possessive for most nouns, but distinct for the pronouns mine, ours, theirs, yours.
  \item I will gloss over the difference between adjectives and various other modifiers. In theory, I could differentiate between ADJ – MODIFIER (a N-AGR-ADJ feature) and MODIFIER (all other modifier features).
\end{itemize}
Figure 6. Entries for two morphological forms of eat.
FSs representing the masculine singular noun libro ‘book’, the masculine singular adjective rojo and the masculine singular determiner un. The FS for libro rojo is derived by unifying the FS anchored by libro with the dashed line portion of the FS anchored by rojo. The resulting FS unifies with the dashed line portion of the FS anchored by un, resulting in the FS un libro rojo. These unifications are restricted by the features and values in these dashed portions. Both un and rojo require the masculine singular heads. Thus libro and libro rojo could not combine with the feminine adjective roja or the feminine determiner una.

(5) a. un libro rojo
    a (MASC-SING) book (MASC-SING) red (MASC-SING)
    ‘a red book’

    b. las calles limpias
    the (FEM-PLUR) streets (FEM-PLUR) clean (FEM-PLUR)
    ‘the clean streets’

In Russian, N-AGR is typically realized as morphological agreement of case, gender and number between determiners, adjectives and nouns as exemplified in (6). This would seem to require a similar account to the Spanish data above, with the addition of MORPH-CASE agreement. However, an analysis of the perfectly grammatical (7) makes it clear that a more complex account is required. That is the topic of the next section.

(6) a. пяти хороших друзей
    five (DAT) good (PLURAL-DAT) friends (PLURAL-DAT)
    ‘(to) five good friends’
un libro rojo
b. trjawx bol’six avtomobil’ja
three (LOC) big (PLURAL-LOC) cars (PLURAL-LOC)
‘(in) three big cars’

(7) a. dva novyx studenta
   2 (MASC-NOM) new (GEN-PLUR) student (MASC-GEN-SING)
   ‘2 new students’

b. dve novyx/novye studentki
   2 new (GEN-PLUR/NOM-PLUR) student (FEM-GEN-SING)
   ‘2 new (female) students’

c. p’yat’ novyx studentov/studentok
   5 new (GEN-PLUR) students (MASC/FEM-GEN-PLUR)
   ‘5 new students’

5.5 **N-AGR in Russian**

It is possible for a FS anchored by Z with head X to have distinct morphological agreement features from X. Russian nominative plural nouns with numerals as determiners require such an analysis.\(^{12}\) Nominative plural numeral determiners anchor nominative plural nouns and require either genitive singular or genitive plural noun heads, depending on the numeral. The assumption that the agreement licensor (the determiner) is the anchor, but not the head (constituent licensor) of N provides an analysis of these constituents that maintains the Chapter 3 definition of head.

In Fioretti 1993, it is demonstrated that Russian agreement within and external to numerically quantified Ns cannot be accounted for while assuming a conventional unification-based analysis, in which each projection of the head and the maximal projection share all morphological agreement features.\(^{13}\) (7) consists of nominative plural Ns that are quantified by numerals and agree with plural forms of finite verbs. Depending on the final digit of the numeral, the head noun and all adjective modifiers have different number and case markings. The morphological properties of the anchoring numeral, any adjective modifiers, the head noun and the phrase as a whole are summarized as Table 5.5. Russian nominative Ns with numeral determiners ending in 1 are relatively unproblematic. The numeral determiner ending in 1 agrees with adjective modifiers and the

\(^{12}\)The analysis proposed here also applies to inanimate accusative Ns, since inanimate accusative Ns are marked with the same inflectional endings as nominative Ns.

\(^{13}\)This assumes that the noun is the head of the nominal phrase. A DP analysis could account for some, though not all, the data discussed in this section. However, a DP analysis precludes a principled definition of head as discussed in Section 4.3.
head noun in case, gender and number. \( \overline{N} \)s in cases other than nominative are also unproblematic in this way. However, nominative numbers ending in the digits 2, 3 or 4 take genitive singular head nouns; and nominative numbers ending in other digits take genitive plural head nouns. In \( \overline{N} \)s quantified by numbers ending in 2, 3 or 4, the masculine (singular) nouns co-occur with genitive plural adjectives and feminine (singular) nouns co-occur with either genitive plural or nominative plural adjectives. This complicated set of conditions cannot possibly be accounted for if the value of the features \textit{Morph} – \textit{Agr} and \textit{Morph} – \textit{Case} for each projection of a given head noun must be shared with the accompanying adjectives and determiners. The plural \( \overline{N} \)s in (7) all agree with plural finite verb inflection and occur in positions which require nominative case, even though the head nouns of (7a,b) are genitive singular and the head noun of (7c) is genitive plural. Secondly, internal to the \( \overline{N} \), the noun adjective co-occurrence patterns for (7a,b) are too complex. Corbett 1983 (pp. 215–240) and 1988 (pp. 28–29) observes similar phenomena in Polish, Serbo-Croat and Old Church Slavic as well as Russian. Fioretta observes similar problems in Finish, Icelandic, Norwegian, Swedish and other non-Slavic languages.

In this analysis, nominative numerals ending in digits from the set \( \{3, 6, 7, 8, 9, 0\} \) anchor a FS like Figure 66 (differing in the phonology feature). An \( \overline{N} \) anchored by one of these numerals is nominative plural, but the head must be genitive plural, as required by the relation \( N-AGR-SPEC(D, H) \), where \( D \) is the anchoring numeral and \( H \) is the head, consisting of the noun modified by zero or more adjectives. In Russian, \( N-AGR-SPEC\textit{Licensor} \subseteq \textit{Quantifier} \) and \( N-AGR-SPEC\textit{Licensee} \subseteq \textit{Head} \). The adjectives and nouns agree in case, number and gender, and thus lend themselves to the sort of \( N-AGR \) analysis discussed in the previous section. Timberlake 1988 takes a similar categorial grammar based approach to case phenomena in Lithuanian. In Timberlake’s approach these determiners could be viewed as belonging

<table>
<thead>
<tr>
<th>Final Digit</th>
<th>Adjective</th>
<th>Noun</th>
<th>( \overline{N} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Masc/Fem)</td>
<td>Nom-Sing</td>
<td>Nom-Sing</td>
<td>Nom-Sing</td>
</tr>
<tr>
<td>2, 3, 4 (Masc)</td>
<td>Gen-Plur</td>
<td>Gen-Sing</td>
<td>Nom-Plur</td>
</tr>
<tr>
<td>2, 3, 4 (Fem)</td>
<td>Gen-Plur</td>
<td>Gen-Sing</td>
<td>Nom-Plur</td>
</tr>
<tr>
<td>5, 6, 7, 8, 9, 0 (Masc/Fem)</td>
<td>Gen-Plur</td>
<td>Gen-Plur</td>
<td>Nom-Plur</td>
</tr>
</tbody>
</table>

Table 7 Agreement properties of nominative plural \( \overline{N} \)s anchored by numbers
to the category \(N\text{-Nom-Plur}/N\text{-Gen-Plur}\), a category that combines with a category \(N\text{-Gen-Plur}\) to yield a category \(N\text{-Nom-Plur}\).

Russian nominative numerals ending in the digits 2, 3 and 4 cannot be accounted for as above because the head noun \(N\), the determiner \(D\) and the one or more adjectives \(A_1, \ldots, A_k\) that can occur between \(N\) and \(D\) differ in their morphological agreement features. The analysis in Figure 68 is proposed for (7b), using the lexical entries in Figure 67. A non-surface \(N\text{-AGR-SPEC}\) arc is proposed for the lexical entry for each genitive singular noun, genitive plural adjective and nominative plural feminine adjective. The value of this \(N\text{-AGR-SPEC}\) arc must be a numeral ending in 2, 3 or 4. The \(N\text{-AGR-SPEC}\) arc of each adjective structure shares with the \(N\text{-AGR-SPEC}\) arc of its modifier. A numeral ending with 2, 3 or 4 anchors an nominal constituent like the one for *due* in Figure 67. In Russian, the feature QUANTIFIER is a surface feature subsumed by QUANTIFICATION licensor and \(N\text{-AGR-SPEC}\) (Licensor). The QUANTIFIER feature of the determiner structure shares with the \(N\text{-AGR-SPEC}\) feature of its quantifier. In other words, the determiner quantifier is the value of a head \(N\text{-AGR-SPEC}\) feature appearing at every level and placing morphological restrictions on the head noun and all its adjectival modifiers. The particular restrictions that these determiners place on adjectives and nouns are distinct. The proposed account resembles the Wasow et al. 1984 account of compositional idioms. A compositional idiom like *leave no stone unturned* can be altered in a number of ways that lead Wasow et al. 1984 to propose...
Idiom-specific lexical entries for each word in the idiom. Their analysis of leave no legal stone unturned includes that a special idiomatic sense of stone is modified by the adjective legal. The special entries for the words in the idioms, as well as the proposed special entries for nouns and adjectives, provide ways of capturing idiosyncratic properties of words and classes of words in special environments.

Chapter 3 defines the head of a phrase XP as the constituent that determines those features of XP which are subject to semantic selection and subcategorization. Chapter 4 provides some examples in which the predicate licensor is distinct from the head. This section provided one example in which the agreement licensor is distinct from the head. In the latter case, the non-head anchor determines agreement properties of the whole phrase, properties that effect the distribution of the phrase.

5.6 Agreement Relations: Surface or LF Relations

5.6.1 Introduction

In a GBugged model of Minimalist Program (MP) analyses, agreement licensing relations would be viewed as Logical Form (LF) relations, rather than surface relations as above. Case is assigned much the same way at LF across languages, although the steps leading up to LF case assignment may differ. Thus MP representations sometimes include arcs that do not represent observable phenomena, but rather exist to preserve aspects of a "universal" grammar. In the approach taken in this book, each licensing relation represents some observable relation within the language being modeled. From the latter point of view, the MP approach is ad hoc. Additionally, evidence from Johnson and Lappin in press shows that a surface interpretation of agreement relations yields a more efficient account.

5.6.2 Feature Checking in the Minimalist Program

In the MP, a set of lexical entries (lexical entries of the words in the input string) is mapped into a convergent derivation by repeated applications of two operations, MERGE and MOVE, where each MERGE/MOVE operation and the derivational process as a whole are constrained by various economy conditions. MERGE combines trees representing lexical entries (as proposed for unification above) and MOVE transforms one tree into another, moving a constituent from one position to a position where some of its features need to be "checked". There are two types of movement: covert movement, where the formal features FF of a constituent moves; and overt movement, where the whole constituent

---

14The Chomsky 1995 version is assumed.
Figure 67: FSSs anchored by dve, novyx and studentki.
FIGURE 68  dve novyx studentki
moves. Below, simple instances of movement is modeled with structure sharing. For other instances of movement, a structure building process is needed as well—that process is modeled as TAG adjunction. In the TAG adjunction cases, an additional projection of a node is added in the middle of a tree and the moved constituent becomes an adjunct or specifier under the new node. The final or interface levels are Logical Form (LF) and Phonetic Form (PF). An operation SPELL-OUT is applied to items when their phonological features are "eliminated" by MP's feature checking mechanism. SPELL-OUT feeds items into the level of PF. LF is the final syntactic level of the derivation.

Figure 69 contains two graphs which together outline a derivation of a DAG-based Minimalist LF for the simple sentence John saw Mary. The arcs next to the circles represent final LF positions and the arcs next to the squares represent SPELL-OUT positions. The graph on the left is derived by merging representations of John and Mary into the tree projected from (anchored by) the transitive verb see and then merging the derived tree with the tree projected from the functional head Tense. The tree on the right is derived by four movement operations, each modeled as an instance of structure sharing. For overt movement, only COMPL(ement), Head, Specifier, and Adjunct arcs dominate the "moved" node, e.g., John is dominated by two SPEC arcs. For covert movement, one FF arc dominates the node to represent that only formal features (FF) have moved. Additional structure is inserted by TAG adjunction for each of the three instances of covert movement (one creating a specifier position, the other two create adjunct positions).

In Chomsky 1995, compatible (unifiable) features cancel each other out when checked. Checked features are eliminated and cannot be accessed by further instances of merge or move. I propose a unification-based account, which will be kept informal (e.g., not represented in a DAG diagram) because Chomsky does not explicitly list the "checked" features. In this account, "eliminated" features do not percolate upward because they are not head features. Rather, they are features of the specifier or adjunct being checked. For example, the FS an-

---

15 See Joshi and Schabes 1991 for a definition of TAG adjunction. Stabler 1996 presents a formal model for some MP operations using TAG adjunction as proposed here. TAG adjunction for GBUG will be discussed in Chapter 6.

16 PF is not as clearly articulated in the MP as LF.

17 In order to keep this DAG small, full GBUG notation is not used. For ease of comparison with MP work, internal nodes are labeled.

18 In a more fully specified model, the X's would be complex FSs, rather than John, saw, Mary, light-cause-verb and T. Some subset of the features for each head would be a defined set of formal features. Only these features would be shared in instances of covert movement.
The structure after 3 instances of "merge".  

The Final Minimalist LF
chored by $T$ includes a nominative case feature ($MORPH-CASE: NOMINATIVE$) as part of the value of its specifier. The FS for $John$ is marked for $NOMINATIVE\mid ACCUSATIVE$, and its entry can therefore unify into the position specifier of $T$. If a feature value pair is to be percolated up to the next level, that feature must be shared with the root node of the anchored phrase. For example, the adjunction structure consisting of the light causative verb$^{19}$ and the adjoined verb $saw$ might share some of the features of the adjoined verb in order to license objective case at the next stage in the derivation (these shared features are head features). In this latter case, the shared objective case marking feature marks the dominating node as potentially anchoring an adjunction structure (inserted by TAG adjunction) with a morphologically accusative $\mathbf{N}$ in specifier ($SPEC\text{-}CASE$) position.

5.6.3 SPELL-OUT and S-structure

SPELL-OUT is triggered by the agreement of phonological features and results in PFs of convergent derivations. Additional (non-phonetic) feature agreement operations motivate other movement operations, movements in which only formal features are moved. SPELL-OUT positions in the MP are distinct from S-structure positions in previous P & P work in that they have no special status as a “level”. No special constraints apply at SPELL-OUT. There is no principle requiring that all SPELL-OUT positions must be derived prior to further derivational processes (MERGE/MOVE operations). SPELL-OUT positions are merely stepping stones on the way to final LF positions. In the MP, abstract case and agreement features are “checked” in the final LF positions, rather than S-structure/SPELL-OUT positions. SPELL-OUT positions are special only in that phonetic features are checked in those positions.

In Figure 69, abstract nominative case and abstract accusative case are assigned at the LF arcs dominating $John$ and $Mary$ (indicated with circles). The phonetic features of the NPs are checked at the SPELL-OUT positions (indicated with squares).$^{20}$ Constraints on morphological agreement features, adjacency (and the various other realizations of case under GB) all constrain PF via SPELL-OUT. In this sense, the SPELL-OUT positions are equivalent to GB S-structure positions. In contrast, the final LF positions of the moved NPs do not seem to model anything observable. They make it possible to claim that, case is universally as-

19The light causative verb is not an observable entity, but rather a constituent Chomsky assumes for theoretical reasons.
20Most MP accounts assume DP analyses. This NP/DP distinction is not relevant to the current discussion.
signed to specifier positions across languages. Analyses of languages with different word orders differ only in whether certain movements apply overtly (before SPELL-OUT) and covertly (after SPELL-OUT), which constituents move, and a few other details. I take issue with these “invisible” movements. Since they do not model any observable linguistic phenomena, the final LF positions that these movements assume are not falsifiable. They seem superfluous to a description of a given language and the assumed connection between languages that these additional positions are supposed to represent seems ad hoc. The additional positions provide just another way of representing that a particular piece of information is idiosyncratic to a language.

Johnson and Lappin In Press

show that the GB view of case results in a less computationally complex system than the MP view of case. Under the MP constraint PROCRASTINATE, covert movement (movement in which no phonetic features are moved) is “cheaper” than overt movement (which moves phonetic features). Covert movement takes place after SPELL OUT and overt movement takes place before. Overt movement of a constituent C to be checked by a head H is required only if this movement eliminates some categorial features on H that are not intrinsic to H, e.g., if H is a verb with nominal features, overt movement of a nominal may be required to eliminate these nominal features before LF. Procrastinate, in effect, says that overt movement is “marked” if it was not required in the first place. As with other of the MP’s economy conditions, several derivations must be tried and evaluated to see which is the most economical (e.g., which violates Procrastinate the least number of times).

Following Johnson and Lappin In Press (pp. 14–17), each English simple transitive clause has at least five possible derivations. The five derivations considered differ only in the movements of two elements: Vb (the light causative verb plus adjoined verb) and the object of the verb v.22 (8) (from Johnson and Lappin In Press) lists the differences between these derivations in terms of how and when these items move. The resulting LF trees are given in Figure 70. These trees are regular Minimalist trees rather than DAG variants as in Figure 69 (although arc labels are maintained for the sake of clarity): ts are traces, coindexed with the moved constituent (the antecedent) and FF(X) represents that

---

21 In GBUG, this would mean that SPEC-CASE is the only type of case. This does not seem to be a necessary assumption of the MP, as case is assigned to adjunct positions in some of the derivations that are considered. However, case is always assigned to specifiers in the most economic convergent derivations.

22 The number of possible derivations is significantly higher when other factors are considered.
only the formal features of X have moved. (A-E) in Figure 70 corresponds to (8a-e). The number of convergent derivations of a sentence increases exponentially as the number of clauses in the sentence increases, e.g., 25 derivations for a two clause sentence, 125 for three clauses, etc. (assuming all clauses in these sentences are simple transitives). Of these five derivations, (8c-e) will be ruled out once Procrastinate applies. The overt movement of Vb in (c) and (d), and the overt movement of OBJ in (d) and (e) ultimately cause Procrastinate violations. (a) and (b) contain no Procrastinate violations.

(8) a. $FF(Vb)$ adjoins to $T$ (covert) and $FF(OBJ)$ adjoins to $T'$ (covert).

b. $FF(OBJ)$ adjoins to $V'$ (covert) and $FF(Vb)$ adjoins to $T$ (covert).

c. $Vb$ adjoins to $T'$ (overt), $SUBJ$ moves to $SPEC$ of $TP$ (overt), and $FF(OBJ)$ adjoins to $T'$ (covert).

d. $OBJ$ moves to $SPEC_2$ of $V^M AX$ (overt), and $SUBJ$ moves to $SPEC$ of $T$ (overt).

e. $OBJ$ moves to $SPEC_2$ of $V^M AX$ (overt), $SUBJ$ moves to $SPEC$ of $T$ (overt), and $Vb$ adjoins to $T'$ (covert).

Johnson and Lappin propose a reformulation of Procrastinate so that movement is illicit if the resulting derivation $D$ contains a constituent $C$ such that $C$ contains phonological features and $C$’s categorial feature are not checked by a functional head in $D$. This version of Procrastinate constrains movement rather than derivations produced by movement. As a result, the total number of derivations is reduced in two ways. The new procrastinate rules out derivations for each clause separately, so that the number of derivations increases linearly rather than exponentially, e.g., 10 derivations for a two clause sentence, 15 for three clauses, etc. Secondly, since procrastinate constrains movement rather than derivations, procrastinate-violating derivations are not generated. The overt movement of $Vb$ in (c) and (d), and the overt movement of $OBJ$ in (d) and (e) are prevented. The result is only 2 derivations per transitive clause. This “local” version of Procrastinate makes it possible to define a point in a derivation which is, in effect, S-structure:

Given Procrastinate as a local constraint on movement, it is possible to define a point in a derivation which is, in effect, S-structure. Let a convergent derivation $D$ from a numeration $N$ to an $LF$ be an ordered k-tuple $< d_1, \ldots, d_k >$ of structures such that for each $d_i, d_{i+1}$ in $D$, \footnote{Figure 69 corresponds to derivation (A).}
$d_{i+1}$ is obtained from $d_i$ by an application of Merge or of Move, and $d_k$ is a well-formed LF. S-structure corresponds to S in 13.

13. S is the earliest $d_i$ in D such that (i) the items in N have been exhausted in $d_i$, (ii) $d_i$ is a single fully projected structure, and (iii) there are no non-intrinsic categorial features in any functional heads in $d_i$.

S is the first point in a derivation which is the output of all applications of Merge to items in N and all overt movement. It is the initial input to covert movement. While it is not an interface level as no interface conditions apply at S, it is a significant level of structure in a derivation. S non-arbitrarily factors a derivation into overt and covert components, and it constitutes the first structure in which all lexical items of a numeration have been projected into a single structure. It is also the last point in D that can yield a convergent PF. Moreover, in cases involving no phonologically null items, it is the unique point in D yielding a convergent PF. In a sense, then, the formulation of Procrastinate as a local condition on movement entails the re-appearance of S-structure in the grammar. (Johnson and Lappin In Press, p. 18)

The MP’s Smallest Derivation Principle (SDP) prefers derivations with one fewer instance of MOVE. This principle requires that all derivations are generated and then compared by this metric. As Johnson and Lappin point out, a more economic approach would be to rule out sub-optimal derivations before they are generated. Chomsky 1995 (p. 357) provides one possible empirical argument in favor of the SDP. Chomsky compares two convergent derivations for a transitive clause in a language with an obligatory subject in the SPEC of T and optional (overt) OBJ raising (e.g., Icelandic). Johnson and Lappin In Press (p. 21) summarize Chomsky’s two possible convergent derivations as (9). Chomsky claims that while both derivations involve exactly two violations of Procrastinate, (9a) is preferred because (9b) involves one additional instances of MOVE.

(9) a. SUBJ raises overtly to SPEC of T and OBJ raises overtly to SPEC$_2$ of V$^m$ax.

b. OBJ raises overtly to SPEC$_2$ of V$^m$ax. OBJ raises overtly from SPEC$_2$ of V$^m$ax to SPEC of T, and SUBJ adjoins covertly to $T'$.

Johnson and Lappin In Press

(p. 22) provide an alternative explanation. If case is viewed as a visibility condition for theta marking (as in GB), then the second movement of OBJ in (9b) would be illicit because an $\Box$ would not be able
to move to a case position, subsequent to case checking. This in effect reintroduces GB's case filter and GB's version of case. The GB approach prevents the suboptimal derivation from being generated. In contrast, MP's SDP (like Procrastinate) is a filter on already generated derivations. Under the MP, many ultimately discarded derivations must be generated which will ultimately fail. This is computationally expensive.

5.7 Summary

This chapter presents a GB-style view of abstract Agreement relations including abstract case and abstract N–AGR. Based on previous GB analyses involving the case filter (Figure 53), I have proposed a more general filter called the Agreement filter (Figure 52) that constrains a wider range of Predicate relations. In GB theory, morphological agreement and morphological case are phenomena that constrain the application of abstract case. However, the intricacies of handling these morphological phenomena have not previously been studied in GB theory. I have provided explicit means for representing morphological agreement and morphological case relations based on previous work in other unification-based theories.

The surface approach to abstract agreement is shown to be preferable to the LF approach taken in the Minimalist Program (MP). In the surface approach, abstract agreement is realized as a set of observable constraints on linguistic phenomena. The surface constraints studied can be modeled using FS logic constraints and FS unification. In contrast, Minimalist LF representations include structures, which do not model any observable phenomena. These structures provide a way of comparing how case is realized in a particular language to some universal grammar. In the grammar of the language being studied, however, these structures appear ad hoc. Additionally, deriving Minimalist LFs is significantly more computationally complex than deriving P & P representations that assume a surface approach to case.
Scope Licensing Relations

6.1 Introduction

In GBUG, the scope of each scope licensor $Q$, a quantifier, negative item, WH phrase, etc. is represented as a scope licensing relation \( SCOPE(Q, X) \), where $X$ is the constituent over which $Q$ takes scope. Scope licensing relations are often non-surface relations (covert movement under most P & P approaches), but may also be surface relations in the case of long distance dependencies (overt movement under most P & P approaches). See Huang 1981, 1982, Lasnik and Saito 1984, May 1985 and others regarding treating long distance dependencies and quantifier scope as related phenomena.

In particular, WH extraction (a surface phenomenon) is given an analysis similar to those given to WH phrases in \textit{in situ}. For example, (1b) is an informal LF representation of the interpretation of (1a) where \textit{which movie} takes wide scope, giving (1a) the meaning of (1c) (there is one movie in question). This contrasts with the reading in which \textit{every linguist} takes wide scope as represented by the LF in (1d) which gives (1a) the meaning of (1e) (there may be as many movies as linguists). In (1b), the WH element is a sister to the constituent over which it takes scope and is linked to its surface position by a filler/gap relation, where the gap is informally represented as $t_j$. A similar filler/gap relation is assumed to represent that the quantified expression \textit{every linguist} takes scope over the constituents other than \textit{which movie}. Example (2) represents a long distance dependency in virtually the same way. \textit{Which movie} fills the gap indicated by $t_j$ and $t_j$ is within the scope of the WH phrase. (2) has the same two interpretations as (1). However, there are two differences between these two examples. In (2) the WH phrase is in a surface position, as well as an LF position. In contrast, the filler in (1b) is in an LF position only. The gaps in (2) are both in a theta (or
D-structure) positions in contrast to the gaps in (1), which are both in surface positions.

(1)  
  a. Every linguist saw which movie?  
  b. \([\text{Whichmovie}]_i[\text{everylinguist}]_j[t_i[\text{saw} t_j]]\)  
  c. For which movie X, for every linguist Y, Y saw X.  
  d. \([\text{Everylinguist}]_i[t_i[\text{whichmovie}]_j[\text{saw} t_j]]\)  
  e. For every linguist Y, for which movie X, Y saw X.

(2)  
  a. Which movie did every linguist see?  
  b. \([\text{Whichmovie}]_i[\text{did}][\text{everylinguist}]_j[t_i[\text{sees} t_j]]\)  
  c. \([\text{Everylinguist}]_i[\text{did}][\text{whichmovie}]_j[t_i[\text{sees} t_j]]\)

In GBUG, structure sharing analyses may be given to these phenomena, as in Figures T1 and T2, the readings of (1) and (2) in which the WH phrase take wide scope. WH-Q and WH-SCOPE are scope licensor features, and HEAD is the scope licensee feature. While both WH-Q and WH-SCOPE represent LF relations, the former, but not the latter represents a surface relation as well. The scope of the quantified expression every linguist is represented in Figure T1 as the relation SPEC-SCOPE, with SPEC-SCOPE as the scope licensor feature and HEAD as the scope licensee feature.\(^1\)

6.2 Surface Scope Licensing Relations

A surface scope licensor SSL anchors a FS representing a \(\overrightarrow{C}\) with the SSL as a specifier and a scope licensor as the \(\overrightarrow{C}\) head. In Figure T2, the feature WH-Q is a surface scope licensor feature and SSL is the WH-N which movie.\(^2\) SPECIFIER \(\subseteq\) WH-Q, SCOPE-LICENSEOR \(\subseteq\) WH-Q and WH-Q \(\subseteq\) SURFACE.

The FS anchored by the SSL includes a long distance dependency with the SSL as the filler and the gap as a constituent within the \(\overrightarrow{C}\) head, the scope of the SSL. In GBUG, gaps of long distance dependencies may be represented using arcs labeled with some SLASH feature.\(^3\) MAC-SLASH is the SLASH feature representing the gap in Figure T2. The

\(^1\)As per the analysis in Chapter 4, the feature QUANTIFIER is a predicate licensor feature and the same source HEAD arc is its predicate licensee feature. Thus \(\overrightarrow{C}\) every predicates of \(\text{linguist}\) and the scope relation is indicated separately.

\(^2\)Which anchors a WH-N where the category noun subsumes the category WH-noun. The determiner which combines with its head noun to form a WH-N, a special type of N which anchors a WH Question, among other constructions.

\(^3\)This idea is based on the slash categories of GPSG, HPSG and related frameworks. The GBUG interpretation of slash categories is based on the use of slash r-signs in SFG. (See Johnson et al. 1986) The idea that trace analyses can be modeled using slash categories is based on the work of Nelson Correa (Correa 1987, 1988, 1991).
FIGURE 71 WH Scope Reading for Every linguist saw which movie?
FIGURE 72  Which movie did every linguist see?
name of a \textit{SLASH} feature is derived by concatenating some surface feature \(S\) and the the word \textit{slash}. The resulting slash feature is defined as a non-surface feature which is subsumed by \(S\) and is also subsumed by the feature \textit{SLASH}. \(\text{MAC} \subseteq \text{MAC-SLASH}\) and \(\text{SLASH} \subseteq \text{MAC-SLASH}\). Therefore \(\text{MAC} \cup \text{SLASH} = \text{MAC-SLASH}\).

In Figure 72, \textit{MAC-SLASH}, together with the same-source \textit{HEAD} arc, represents the set of licensing relations between the verb \textit{eat} and the WH phrase filling the gap. Thus the WH-phrase is a complement licensee, a case licensee, and a theta licensee, the verb being the licensor of each of these relations. The WH-phrase is subject to the selection and morphological agreement restrictions relating to the the case and theta relations. However, since \textit{MAC-SLASH} is not a surface feature, its value cannot be restricted by the word order constraints associated with these case and theta features. Word order is by definition a surface phenomenon. There is no sense in suggesting that a non-surface feature (or empty category) precedes or follows any other feature (or constituent). Therefore, a word order restriction cannot be violated by a non-surface relation.

The FSs in Figure 73 combine to form the FS in Figure 72 as follows. The FS labeled \textit{which movie} anchors the FS in the lower left corner of Figure 73. These two FSs combine by unifying the former under the \textit{WH-Q} arc of the latter, resulting in a FS which we shall call \(W\). WH noun phrases are instances of phrases which have lexical entries (as discussed in Chapter 4), and this is an example of lexical look-up. \(W\) is an extension of a normal FS in that it has two roots, an innovation developed in Johnson et al. 1993 for representing long distance dependencies. There is one normal root and one \textit{DANGLING} root. The dangling root is the source of exactly one arc, the \textit{DANGLING} arc. \(W\) requires that the value of the \textit{HEAD} arc contain a gap represented by some arc \(G\), such that the \textit{DANGLING} root unifies with the source of \(G\). \(R\), the FS on the right in Figure 73 can unify with the value of the \textit{HEAD} arc, and the incomplete \textit{MAC} arc under \(R\) can unify with the dangling arc in \(W\). The result is Figure 72.\footnote{Following P & P traditions, \textit{did} is assumed to be the head of \(R\) as well as the \(\overline{T}\) in \(R\). In the analysis presented here, auxiliaries anchor these constituents of category \textit{comp}, which can occur either independently as yes-no questions or as \textit{WH-Q} licensees, as in Figure 72.}

This account of long distance dependencies is further constrained in Section 6.4, where arcs representing “intermediate” positions are introduced.

\textsuperscript{Correa formalizes Government and Binding Theory trace analyses in a manner similar to slash categories, using an Attribute Grammar model.}
**FS anchored by which movie**  
**Possible Head of WH Question**

Figure 73 Which movie + did every linguist see?
6.3 Nonsurface Scope Licensing Relations

In GBUG, a nonsurface scope licensor $NSL$ anchors a FS which is TAG-adjointed to an existing FS $H$ representing a maximal projection that already contains $NSL$, where (a version of) TAG-adjunction is defined below. The FS derived this way represents the scope of the $NSL$, where scope is an LF relation. $NSL$s are quantifiers, in situ WH operators, negative items and other scope operators. $H$ is the constituent over which a given $NSL$ takes scope.

Figure 71 is a final LF representation of the sentence *Every linguist saw which movie?* in which the WH phrase gets wide scope. Figure 71 is derived from the FSs in Figure 74. The surface representation of the sentence in Figure 74 unifies with the value of the $HEAD$ arc in the $SPEC$-$SCOPE$ FS and then the resulting FS unifies with the value of the $HEAD$ arc in the $ADJ$-$SCOPE$ FS. The dangling arcs are labeled $SURFACE$, a feature which may unify with any surface arc. In deriving Figure 71, the dangling arc from the $SPEC$-$SCOPE$ FS unifies with $SPEC$-$CASE$ and the dangling arc from the $ADJ$-$SCOPE$ FS unifies with $MAC$. In the derived FS, the scope of the value of $MAC$ (the WH-phrase) includes the scope-licensor arc dominating the value of $SPEC$-$CASE$ (*every linguist*). Thus the WH phrase has wider scope. The LF in which *every linguist* gets wide scope could be derived by reversing which surface arc unifies with which dangling arc.

The following constraints on scope licensees are met by the scope licensees in the above example. Following Chomsky 1986a, it is assumed that a scope-license: (a) cannot have a category of $INFL$; and (b) cannot be the value of any $THETA$ feature unless the theta licensor is of category $INFL$. Figure 75 is a FS logic version of this constraint on scope licensees. Additionally, in order to prevent too many scope licensing relations in a given phrase, constraints on the value of the head.

---

5In previous sentences, it is assumed that surface sentences are $I$. These examples follow the convention that they are in fact $\hat{I}$’s. No theoretical position is assumed on this matter. In previous sections, the complementizer position was irrelevant to the discussion at hand. Here, the conventional analysis is assumed as a default.

6$Surface$ is assumed to be a disjunction of all surface features in the grammar. This is of course too broad. For example, pronoun modifiers of nouns cannot extract, as evidenced by the ungrammaticality of (6), where the type of man is being questioned. If this sentence were well-formed, the answer might be (6).

(i.) *What did Mary see the t man?*

(ii.) Mary saw the ice cream man.

7The assumption is that $INFL$ is defective, in that its complement is not an argument and its maximal projection is not a permissible adjunction site.
Every linguist saw which movie

**Surface Representation of**

**Spec-Scope Feature Structure**

**Adj-Scope Feature Structure**

**FIGURE 74** FSs needed for deriving a WH Scope Reading
\[ \text{SCOPE} \subseteq \text{SPEC-SCOPE} \] and \[ \text{SCOPE} \subseteq \text{ADJ-SCOPE} \]

In the above wh in-situ analysis, the scope licensee FSSs unify under the scope licensor FSSs. This appears to be similar to analyses in previous chapters in which anchors and nonanchors are merged to form a phrase, e.g., when a complement unifies under the FS anchored by the head of a phrase. However, if it is assumed LF configurations are derived from surface representations, it will sometimes be necessary to insert the scope FSSs in the middle of existing graphs. TAG-adjunction, an operation borrowed from the Tree Adjoining Grammar formalism (cf. Abeille et al. 1988 and Joshi and Schabes 1991) provides the formal mechanism needed to achieve this. Figure 77 provides a simple example of TAG-adjunction based on the \textit{ADJ-SCOPE} graph in Figure 74. The auxiliary tree in Figure 77 is a labeled tree version of the \textit{ADJ-SCOPE} graph from Figure 74 without the dangling arc. The elementary tree is a simplified labeled tree representation of the verb phrase from the sentence \textit{I asked which movie every linguist saw}, in which \textit{every linguist} takes wide scope (there may be more than one movie). The derived tree is formed by replacing the CP node in the elementary tree with the auxiliary tree, such that: (a) the root of the auxiliary tree is the target of the arc (branch) which previously lead to the CP in the elementary tree; and (b) the foot of the auxiliary tree is the source of all branches for which the elementary tree CP was the source.

Figure 78 extends TAG-adjunction to FSSs (cf. Vijay-Shanker and Schabes 1988).\footnote{Vijay-Shanker and Schabes 1988 combines FS and TAG formalisms. That paper includes a FS version of TAG-adjunction.} The derived graph (derived FSS) is formed in a similar manner as the derived tree in Figure 77, except that tag-adjunction must be adapted for structure sharing. In the derived graph, the root of the auxiliary graph is the target of all arcs which, in the elementary graph, had the CP as their target. Additionally, the dangling arc is unified with a FS in the
which movie every linguist saw
elementary tree. Figure 79 is a more complete version of the derived graph from Figure 78.

It is assumed that every FS anchored by a nonsurface scope licensor (e.g., the FS anchored by everyone) licenses a scope adjunction operation. This information is listed in the lexicon. Scope adjunction operations apply at some point in the derivation after a FS is unified into its surface position. To keep things simple, it is assumed for now that a (surface) FS for the entire utterance is derived first and then all scope adjunction operations are applied to the resulting FS. Given this assumption, a variant of TAG-adjunction is required to derive representations of quantifier scope. See Schieber and Schabes 1990 for another TAG based approach to representing scope relations.

6.4 Intermediate Positions

Both surface and nonsurface scope licensing relations are constrained by a number of factors pertaining to some item X. In long distance dependencies, X is the gap. For nonsurface scope licensing relations, X is the surface position of the scope licensor. The first issue is whether or not X is an internal-theta licensee. If X is an internal-theta licensee and the internal-theta licensor is not of category INFL, constraints are minimal (X is lexically governed). Otherwise, there must be a series of “close enough” intermediate positions linking X and the scope licensor position (so X can be antecedent governed). In GBUG the intermediate positions are modeled as types of SPEC-SCOPE and ADJ-SCOPE licensor arcs. Each S, the source of an intermediate arc dominates X and the source of the (initial) scope licensor arc dominates S. These intermediate positions are required for long distance dependencies of adjuncts or subjects, as in (3) and for the wide scope readings of the bracketed phrases in (4). Figures 80 and 81 are final FS analyses of WH extraction and quantifier scope relations, each of which require one intermediate position, derivable by TAG-adjunction using the SPEC-SCOPE and ADJ-SCOPE Licensor FSs from Section 6.3. Additional intermediate positions would be necessary for more embedded clauses.

(3)  a. How do you think (Mary said + John assumed etc.) I won the big prize t?
    b. Who do you think (Mary said + John assumed etc.) t won?

(4)  a. Every linguist thinks (Mary said + John assumed etc.) John laughed [some place]?
    b. Every linguist thinks (Mary said + John assumed etc.) [some contestant] won?
which movie every linguist saw

Auxiliary Graph

Elementary Graph

Derived Graph

FIGURE 78 TAG-Adjunction for Feature Structures
Figure 79: every linguist taking wide scope in complement of ask
Who do you think won?

**Figure 80** Wh extraction with one intermediate position?
Every linguist thinks some contestant won
In the analyses presented, scope licensing relations represent two distinct phenomena: (1) they mark scope; and (2) they mark intermediate positions. These two phenomena are not mutually exclusive, e.g., the phrase *some contestant* in Figure 81 has scope over two constituents the *C* complement of *think* and the matrix sentence. Stating that a phrase takes scope over both of these constituents is equivalent to stating that it takes scope over the matrix clause since the subordinate clause is part of the matrix. It therefore seems reasonable to use the same relation to represent both intermediate positions and non-surface scope licensing relations. Additionally, the same constraints are assumed on inserting these FSs by TAG-adjunction as in the previous section.

An account of the locality constraints on the positions of fillers, gaps, scope licensors and intermediate positions has been omitted due to space considerations. See Chapter 8, for the vocabulary required for composing such constraints. See Joshi and Vijay-Shanker 1989 and Kroch 1989 for other TAG treatments of long distance dependencies.

### 6.5 Discussion

This chapter adds mechanisms to (Meyers 1994) to accommodate LF phenomena. An attempt has been made to maintain the flavor of P & P analyses while adopting some mechanisms (e.g., slash categories and TAG-adjunction) from other linguistic frameworks. This formalization is one step towards comparing P & P approaches to long distance dependencies and scope phenomena to approaches in other frameworks. Although slash categories and trace accounts seem to have much in common, it is unclear to me how HPSG accounts (cf. Pollard and Sag 1994) using Cooper storage (cf. Cooper 1983) compare with P & P representations of quantifier scope.

Internal to P & P, the above formalization may provide a basis for evaluating some recent work in the Minimalist Program. The previous chapter argued that the LF agreement relations assumed in the Minimalist Program are ad hoc because these relations do not represent any observable linguistic phenomena, e.g., these LF arcs do not represent scope. In order to show that LF agreement (licensing) relations exist, it would seem necessary to show not only that they represent some observable linguistic phenomenon *P*, but also show that *P* and scope (licensing) relations form a natural class. Chomsky 1995 (pp. 311-312) suggests that the same mechanisms (ATTRACT/MOVE) can be used to characterize both phenomena. Given that it is unclear what phenomenon LF agreement relations represent, it is hard to determine whether this is a justified assumption. Under the approach advocated here, it is neces-
sary for LF agreement relations to be defined in terms of some observable phenomena, prior to merging them with scope relations.
Coindexing, Empty Categories and Structure Sharing

7.1 Introduction

I propose a version of P & P theory that uses both empty category (traditionally used in P & P) and structure sharing analyses (used in graph-based linguistic frameworks) to model various filler/gap constructions. First, GBUG coindexing relations are introduced in order that empty category analyses can be properly formulated. Then various filler/gap phenomena are assessed with respect to whether GBUG empty category or structure sharing analyses are preferred. The concerns that are used to make these assessments are then applied to both HPSG structure sharing accounts and previous P & P versions of empty category accounts.

A coindexing relation $CR(X, Y)$ links an empty category (ec), pronominal or anaphor $X$ to its antecedent $Y$. Different types of coindexing relations are distinguished. A $SAME-INDEX$ relation represents that a pronominal/ec has the same (referential) index as its antecedent, i.e. the same entity is represented by $X$ and $Y$. A $SHARED-INDEX$ relation represents that the index of the pronominal/ec $X$ includes the index of the antecedent $Y$. Example 1 exemplifies these possibilities using informal notation. Mary and the reflexive pronoun herself refer to the same individual: the $SAME-INDEX$ relation holds between Mary and herself. Mary and John belong to a group of people, represented by the pronoun their, who have extraterrestrial ancestors. Two $SHARED-INDEX$ relations hold: one between Mary and their and another one between John and their.

(1) a. Mary$_i$ saw herself [SAME-INDEX]
    b. Mary$_i$ asked John$_j$ about their$_{i,j}$ extraterrestrial ancestry
       [SHARED-INDEX]

In previous chapters, structure sharing analyses have been assumed
for phenomena traditionally analyzed with empty categories in the P & P literature. This chapter provides principled reasons for choosing empty category or structure sharing analyses. Empty category analyses are most suitable for representing phenomena where a gap can be shown to have distinct properties from its filler, apart from those captured by licensing relations. For example, each of the antecedents in (2) have distinct properties from the coindexed PRO. A GBUG structure sharing analysis is difficult to formulate since John and Mary do not form a constituent. Structure sharing analyses require fewer stipulations than empty category approaches to model phenomena in which a gap and filler need not be distinguished from each other, e.g., the object equi sentence (3) ("equi" is also known as "control").¹

A view of structure sharing is assumed in which whole constituents are shared, following previous work in APG. This contrasts with subsequent HPSG analyses in which only part of the FSs representing constituents are shared: local features in the case of long distance dependencies, and the value of the feature INDEX in the case of equi and raising. As discussed below, the HPSG view of structure sharing may be available in a subset of cases for which GBUG requires an empty category analysis. However, many problematic cases resist HPSG structure sharing analyses.

(2) Mary talked John into ei j leaving together
(3) Sally told Mary ei to go

7.2 A GBUG Representation of Coindexing
In GBUG, each coindexing relation is represented as a pair of arcs with the same target as in Figure 82. The source of the arc labeled ANTECEDENT is the root of a graph representing the antecedent. The source of the arc labeled with an INDEX-TYPE label (SAME-INDEX, SHARED-INDEX, etc.) is the root of the coindexed anaphor, pronoun, empty category, etc. The shared value of the index-type and ANTECEDENT arcs is a uniquely assigned integer, effectively differentiating indices from each other.²

¹The gaps in (2) and (3) are represented informally as an ε, coindexed with a filler. These representations are intended as neutral with respect to a structure sharing or empty category analysis of the filler/gap relation. In the P & P literature ε is PRO in both cases.

²In Meyers 1994, it was assumed that the value of these arcs was NIL, rather than a unique index. That analysis could be maintained, provided that the values of antecedent and index-type arcs are constrained so that: (1) No non-NIL value could be unified with these values; and (2) An INDEX-TYPE arc could not structure share with another INDEX-TYPE arc. Other constraints may depend on INDEX-TYPE,
Figures 83 and 84 are FS representations of (1). The SAME-INDEX and ANTECEDENT arcs in Figure 83 represent that Mary and herself are coindexed, and therefore share the same reference. The SHARED-INDEX and ANTECEDENT arcs in Figure 84 represent that the index for John and the index for Mary are included in the index for their. Each of the two ANTECEDENT arcs is separately paired with the SHARED-INDEX arc to represent these two distinct SHARED-INDEX relations.\(^3\)

It is assumed that every lexical entry includes one antecedent arc whose value is a unique index (an integer). A different index is assigned each time the lexical entry is accessed. Each pronoun/anaphor/empty category includes some INDEX-TYPE arc with a value of ANY-INDEX, a special atom (or node label) which can only unify with unique indices.

This approach is based on HPSG’s representation of coindexing (cf. Pollard and Sag 1994), which is diagramed as Figure 85. The value of features labeled INDEX includes agreement properties (person, gender, number). The INDEX arc of a pronoun/reflexive and its antecedent have the same value. Therefore, these items must agree in person, number and gender.

HPSG

e.g., SAME-INDEX arcs can structure share with at most one ANTECEDENT arc.

\(^3\)In Meyers 1994, I used labels like SAME-REFERENCE and OVERLAP-REFERENCE, instead of the current labels SAME-INDEX and SHARED-INDEX. The inclusion of the word REFERENCE would seem to suggest that empty categories have reference (R. Fiengo, p.c.). This implication was unintentional.

The approach to indexing I assume here is limited to relations between anaphors/pronouns/empty categories and their antecedents. Other types of coindexing from the literature are modeled differently in GBUG, e.g., subject verb agreement (Chomsky 1981), predication (Williams 1980), etc.
FIGURE 83  Mary saw herself
Picture 84. Mary asked John about their extraterrestrial ancestry.
coindexing and GBUG coindexing differ crucially in that only GBUG allows multiple index types. HPSG would have to be altered for coindexing relations other than \textit{SAME-INDEX} to be represented. The assumption that an item’s \textit{INDEX} includes agreement features is only correct for the \textit{SAME-INDEX} relation. For example, \textit{John} and \textit{Mary} are singular antecedents of the plural pronoun \textit{their} in (1b) above. The sharing of indices would also be problematic since \textit{John} and \textit{Mary} must each be coindexed with \textit{their}, but must not be coindexed with each other. The constraints on agreement which follow from the HPSG account of \textit{SAME-INDEX} must be assumed to be part of the Binding Theory (cf. Chapter 8) in the GBUG account.

The \textit{SAME-INDEX}/\textit{SHARED-INDEX} distinction is inspired by Chomsky and Lasnik 1993. They make the following claim for interpreting indices on pronouns and anaphors:

Suppose NP and α are coindexed. Then
i. if α is an anaphor, it is coreferential with NP;
ii. if α is a pronoun, it overlaps in reference with NP.
(Chomsky and Lasnik 1993, p.99)

Rather than leaving co-indexing up to interpretation, GBUG explicitly states the type of indexing relation because: (1) \textit{SHARED-INDEX} relations are only available for items that are plural (e.g., plural pronouns) or items that are unspecified for number, e.g., PRO (see below). Chomsky and Lasnik’s claim that it applies to all pronouns is incorrect;
(2) **SHARED-INDEX** and **SAME-INDEX** do not exhaust the set of possible indexing relations. Other indexing relations are discussed below.⁴

### 7.3 Advantages of Structure Sharing in GBUG

Following APG (cf. Johnson and Postal 1980), certain linguistic phenomena including passive, raising and equi (obligatory control) are modeled by analyses in which two arcs share the same target. The shared constituent plays multiple roles in that linguistic analyses. For example, in Figure 86, *Mary* receives theta roles from *leave* and *want*, receives case from the matrix verbal inflection (*INFL*), and is the specifier of infinitival *to*. It follows logically, that the *N Mary* must meet all the restrictions (selection, morphological case, etc.) associated with each of these licensing relations. In contrast, assuming the analysis in Figure 87, special rules are required to explain why *Mary* must meet the restrictions associated with the non-surface positions, e.g., why *Mary* must satisfy the selection restrictions associated with the position of *PRO*. In other words, much of what must be stipulated in the theories of control and raising follows from a structure sharing analysis without being stated explicitly. The structure sharing account explicitly represents there is only one constituent being considered.⁵

A comparison of Figures 88 and 89 yields a similar conclusion. The fact that *whom* must be compatible with accusative morphological case follows without stipulation in Figure 88. In contrast, the analysis Figure 89 requires principles on the coindexing relation to achieve this. A structure sharing analysis is preferred on grounds of simplicity. Other considerations may lead to an empty category approach, as discussed below.

In Culy 1994, a structure sharing approach is shown to be superior to an empty category approach for two obligatory control constructions in Donno So (DS), a Niger Congo language. Some of Culy’s examples from this language are given as (4) to (6). Although Culy assumes Lexical Function Grammar (LFG), his findings are independent of the framework assumed.

---

⁴Chomsky and Lænke also discuss a contra-indexing relation which may be stated in GBUG as a constraint of the form \(- (INDEX \rightarrow TYPE = ANTECEDENT)\) for a given pair of constituents. The interpretation is that a given pair of constituents are definitely not coreferential. This area of research will not be explored here.

⁵The **SAME-INDEX** relations in Figure 87 differs from the **SAME-INDEX** relation in Figure 83. In the latter case, *Mary* and *herself* are distinct items which share an index. In the former case, there are only theory-internal reasons for positing separate constituents.
FIGURE 86 Mary wanted to leave
Figure 87  Mary wanted to leave
FIGURE 88 Whom did I see?
**Figure 89** Whom did I see?
(4) a. Mi iye [irinɔ yogo wazu] namam
    I tomorrow blacksmith-df see-inf want-lsg
    ‘Today I want to see the blacksmith tomorrow’
b. Annanɔn yogo wazu mɔnɔ (mi) woolo eebeezem
    old man-df-obj tomorrow see-inf before I colanut buy-fut-lsg
    ‘Before seeing the old man tomorrow, I’ll buy some cola nuts’

(5) a. [s Yogo mi irinɔ] namam
    tomorrow I blacksmith-df see-inf want-lsg
    ‘I want to see the blacksmith tomorrow’
b. [s Yogo irinɔ Omar wazu] nama
    tomorrow blacksmith-df Omar see-inf want-3sg
    ‘Omar wants to see the blacksmith tomorrow’
c. Annanɔn yogo mi wazu mɔnɔ woolo eebeezem
    old man-df-obj tomorrow I see-inf before I colanut buy-fut-lsg
    ‘Before seeing the old man tomorrow, I’ll buy some cola nuts’

(6) a. *Mi [s Yogo mi irinɔ] wazu
    I tomorrow blacksmith-df see-inf want-lsg
    ‘I want to see the blacksmith tomorrow’
b. *Annanɔn yogo mi wazu mɔnɔ mi woolo eebeezem
    old man-df-obj tomorrow I see-inf before I colanut buy-fut-lsg
    ‘Before seeing the old man tomorrow, I’ll buy some cola nuts’

In (4a), (5a,b) and (6a), the matrix subject controls the subject of the complement of a subject control verb. In (4a), (5a,b) and (6a), the matrix subject controls the subject of an adjunct. For both of these constructions, either the controller (4) or the controllee (5) can be overt, but not both (6). On Culy’s account (but using GBUG terminology), the controller and controllee are the same constituent, simultaneously the subject/external-theta recipient of both the matrix and subordinate clauses. The surface syntax of DS permits this subject to be realized in either position, but not both. The fact that the overtness of the controller and controllee in DS are interdependent makes this a promising analysis.

Culy claims that (5) poses a problem for P & P theories because the controller c-commands the controllee and the controllee is an R-expression. This violates Binding Principle C if an empty category analysis is assumed, but does not violate any principle under a structure sharing analysis.

In Rizzi 1982 (e.g., pp. 140-142), some Italian examples similar to (5a,b) are given an analysis in which the subject is in a chain with a “dummy” pro subject (analyzed like English expletives).6 This analysis

---

6Carmen Picallo (personal communication) alerted me to this possibility.
of (5a,b) is given as (7a,b). It avoids the principle C violation because the
cotrollee is not assumed to be the overt  N. However, if this pro-subject
analysis is extended to (5c), as in (7c), the matrix subject and the subject
of a subordinate adjunct clause would form an expletive argument chain.
This chain is ill-formed because the expletive does not c-command the
argument  N.

(7) a. pro j [s Yogo m i PRO j irin wazu] namam
b. pro j [s Yogo irin Omar PRO j wazu] nama
c. Annan n yogo m i PRO j wazu m  nc pro j woo ebezzam

Structure sharing approaches can capture linguistic generalizations
which follow from the fact that a single constituent is being shared. In
contrast, empty category approaches must stipulate that the empty
category “acts” like its antecedent with respect to selection restrictions,
agreement, etc. The Culy data is explained better by a structure sharing
approach than an empty category approach for two reasons: (1) Surface
syntax is compatible with an approach which does not distinguish filler
and gap. The surface constraint on where the constituent was realized
would be less strict than in English (for example). In P & P empty
category approaches, the controller and controllee are distinct for pur-
poses of surface syntax. To account for the DS data, a constraint of the
form ¬(surface(filler) ∧ surface(gap)) would be needed in addition to
whatever constraints were needed in the structure sharing approach. (2)
P & P assumes a Binding Theory which is violated by the DS data. This
suggests that an empty category approach is possible for the DS data,
only if the Binding Theory is revised.

For a given structure sharing analysis, there is probably an empty
category analysis which is nearly equivalent on descriptive grounds (but
may require more stipulations). The following sections show that the
reverse is not true. The use of empty categories is more powerful than
the use of structure sharing. There are some filler/gap phenomena which
are not properly accounted for by the GBUG structure sharing approach.

7.4 EC analyses in GBUG

This section discusses filler/gap constructions in which the filler and gap
are distinguished due to at least one of the characteristics listed in Ta-
ble 7.4. It is argued that a GBUG empty category analysis is more viable
than a structure sharing analysis for constructions with these charac-
teristics.

The instances of the empty category PRO in (8) partially corresponds
to two antecedents (Characteristic 1). The GBUG SHARED-INDEX
relation links PRO to each of these antecedents, as in Figure 84 above.
1. One gap corresponds partially to more than one filler (an empty category with split antecedents)
2. One filler corresponds partially to more than one gap
3. The gap has a “sloppy” interpretation
4. A type of gap is barred from positions which also bar pronouns (Postal 1994)
5. A gap requires case, although its filler does not (Postal 1994)
6. A gap occupies a position which its filler cannot
7. A gap has no syntactically realized filler (in the same sentence)
8. The filler and gap have different morphological requirements

**Table 8** Characteristics of Empty Category Constructions

The PRO subject (external-theta licensee) represents a set of people, including, but not limited to the two antecedents. It is unclear how a structure sharing account could represent these data: two distinct constituents John and Mary would have to be treated as one. Other antecedents would not be accounted for (e.g., suppose PRO was bound by a group including John and Mary, but the other members of the group are not mentioned in this sentence.) PRO in these same set of environments can occur without any same-sentence antecedent as in (9). Since the antecedent of these instances of PRO may be contained in separate sentences, a structure sharing account would have to allow Ns from distinct sentences to be shared. This becomes further problematic, if the filler of the gap is not stated, but understood from context. One can imagine a view in which the unexpressed subject was left unexpressed, the FS for the sentence simply containing a partial description of what that N could be, leaving further elaboration up to other processes, just as a lexical entry for a verb contains an underspecification of its subject. This unexpressed subject would behave a lot like a pronoun given some facts discussed in later sections. An empty category approach could capture this by classifying the gap as a particular type of empty category. The empty category-less approach, it would seem, would not distinguish this unfilled position from any other.

(8) a. Johni asked Maryj if it was a good idea PROi,j to tease each other
    b. Maryi talked Johnj into PROi,j leaving together [= (2)]

---

7 I am ignoring some complexities here. PRO, as used in this book, should be broken down into a number of sub-types. As noted in Leacock 1991 (also Postal, pc), the PRO in (9), behaves much like the pronoun one, e.g., it can bind to the pronouns one or oneself, but no other pronouns.
(9)  a. It was a good idea PRO to tease each other
    b. PRO leaving together was discussed

Following Postal 1970 (pp. 476–478), Fiengo 1974, 1980, Williams 1980, Koster 1984, and others, the various constructions often assumed to contain PRO are not uniform. This has led some of those cited to regard the obligatorily controlled (equi) variety of PRO as an anaphor-like empty category (like NP-trace) and the other varieties as pronominal-like (like “little” pro). The two properties mentioned here are only a subset of those previously cited in the literature which distinguish these two classes of PRO. However, these are the properties which seem relevant to its empty category status. In the obligatory control constructions (equi), PRO must have a single antecedent which is indistinguishable from its controller (filler). Therefore a structure sharing analysis provides an elegant analysis. In other constructions with PRO, as in (8) and (9), PRO may have Characteristics 1 and/or 8 from Table 7.4. On this analysis, there is only one PRO, a pronoun-like empty category. PRO can take split antecedents and need not be bound. Equi constructions are given structure sharing analyses—no PRO is assumed.

As pointed out by Paul Postal (p.c.), it is possible for a single filler to partially correspond to multiple gaps, as in (11) (Characteristic 2). If the phrase which two girls is assumed to occupy the positions of both gaps, an incorrect analysis of the sentence would result since neither girl is both kissed and hugged. Figure (90), the GBUG analysis of (11a) uses two SAME-INDEX relations to represent the filler/gap relations. It is assumed that ANTECEDENT \( \subseteq \) PARTIAL-ANTECEDENT1, ANTECEDENT \( \subseteq \) PARTIAL-ANTECEDENT2, etc.\(^8\)

(11)  a. [Which two girls]_{i,j} did respectively Frank kiss t_i and Mike hug t_j? (Postal’s example)
    b. [The pair of documents]_{i,j} which i, j respectively Mary burned t_i and Glinda shredded t_j?

\(^8\)Some features, including COMPLEMENT, PARTIAL-ANTECEDENT, and the features they subsume, seem to violate the functionality constraint discussed in CHAPTER 2. This problem can be circumvented assuming there are a finite number of distinct complements, partial-antecedents, etc., in which case they can be enumerated (COMPLEMENT1, COMPLEMENT2, etc.). For example, I am aware of no English word that can take more than three complement phrases—Bet and trade can take three complements, e.g. (9). Although prepositional complements of motion verbs iterate as in (10), these cases seem to be instances of coordination.

(i.)  a. I will bet [you] [one dollar] [that I will win]
    b. I will trade [you] [this book] [for that one]

(10)  They walked from their house, over the bridge, down the path, through the woods, to Grandmother’s house.
Secondly, it is assumed that there is a constraint \((\text{PARTIAL-ANTECEDENT} = \text{SAME-INDEX})\). In combination with the previously mentioned constraints, this prevents coindexing relations other than \((\text{SAME-INDEX})\) from having partial antecedents. Without this constraint, (for example) \((\text{SHARED-INDEX})\) could have partial antecedents. This is possible in theory, but ruled out in practice as evidenced by the ill-formedness of the coindexing in (12a). The only available reading for this sentence is the \((\text{SAME-INDEX})\) with partial antecedents as in (12b).

(12) a. *Which two girls \(_{i,j}\) respectively told which two boys \(_{k,l}\) that it would be nice PRO\(_{i,k}\) to work together and PRO\(_{j,l}\) to play together.

b. Which two girls \(_{i,j}\) respectively told which two boys \(_{k}\) that it would be nice PRO\(_{i,k}\) to work together and PRO\(_{j,k}\) to play together.

In Figure 90, the coindexing relations between the matrix INFL and the INFL of the coordinates, is assumed to be a “sloppy” coindexing relation as discussed below.

Gaps of ellipsis, gapping, right node raising and other constructions are all of the same type as their fillers, but the gaps and fillers are token-distinct — some examples are found in (13). There are two distinct instances of eating events described in (13a): Sam is the eater in one event, and I am in the other. Type-identity is represented as a \((\text{SLOPPY-INDEX})\) coindexing relation in Figure 91 (Characteristic 3). A GBUG structure sharing approach would incorrectly represent that the two events were token identical. Therefore an empty category analysis is preferred.\(^9\)

(13) a. Sam [ate green eggs and ham] \(_{i}\) and so did I \(_{i}\).

b. Sam [ate green eggs and ham] \(_{i}\) and I might have \(_{i}\) also.

c. Cecil ordered ice cream and Fred \(_{i}\) pancakes.

d. I probably ate \(_{i}\) and Sam definitely ate \(_{i}\) [five green eggs and one half a pound of green ham] \(_{i,j}\).

\(^9\)Although, the term sloppy characterizes the filler/gap coindexing relation here, this term is usually used for readings of pronominal elements within the elided phrase. For example, (12b) is a “sloppy” reading of the pronoun his in (12a): the pronoun has two antecedents, John for the \(\overrightarrow{\text{loves his mother}}\) and Sam for the gap filled by the \(\overleftarrow{\text{loves}}\). (12c,d) are two possible strict readings in which his has the same antecedent for both filler and gap.

(i.) a. John loves his mother and Sam does too

b. John loves John’s mother and Sam loves Sam’s mother

c. John and Sam both love John’s mother

d. John and Sam both love Fred’s mother
Sam eats eggs and I might also.
Many important details of these “sloppy” constructions have been omitted due to space limitations. The literature on these constructions suggests that a pure empty category approach is insufficient. Some approaches (Lappin and McCord 1990; Fiengo and May 1994; Lappin and McCord 1990; Lappin and Shih 1996) reconstruct the antecedent at the sight of the gap in order to account for the interaction between pronominal elements, their antecedents, and quantifiers. This reconstruction could be modeled as an operation applied to Figure 91 to produce a new FS containing the reconstruction, just as TAG-adjunction is used for deriving scope relations in the previous chapter. Other approaches (Dalrymple et al. 1991) use lambda extraction to achieve a similar effect. Modeling these approaches in GBUG would require adding lambda extraction to our FS logic. Details are left for future research.

Like the PRO examples, some instances of ellipsis allow, the filler of the gap to occur in a separate sentence or inferred from context (Characteristic 7), (cf. Macleod et al. 1996b). (14a-b) are coordinated sentences, the latter of which contains a gap filled by the former. (14c) is an example of ellipsis which can be filled by previous sentences or from context. As with PRO, it is difficult to see how these could be handled with structure sharing. (14d) (from Macleod et al. 1996b) presents a further problem because the filler let me give Papa blood cannot occupy the position of the gap (Characteristic 6), as evidenced by the ungrammaticality of (14e). The sentence is grammatical if infinitival to is inserted before let. However, (14f) cannot be fixed quite as easily. The filler of the gap would be the infinitival clause in (14g), the meaning of which is implied, but does not appear in surface syntax.10

(14) a. Mary asked John to leave and he agreed —
b. Mary thinks that linguistics is fun and John agrees —
c. He agreed —
d. Let me give Papa blood. The doctor agreed —
e. *The doctor agreed let me give Papa blood.
f. He asked to leave. Mary agreed —
g. Mary agreed to let him leave.

Postal (Postal 1994 and elsewhere) has identified a number of antipronominal contexts, syntactic environments which bar weak definite pronouns (normally unstressed). (15) gives some examples of an-

10 These examples pose problems for reconstruction-based accounts as well, unless some Generative-Semantics-like approach is assumed in which the infinitival complement of ask contains an implicit be allowed to. Until recently, such an approach would be disallowed in P & P theory. However, recent work in Minimalism (Chomsky 1995) seems to allow such devices, e.g., a metalinguistic causative verb is assumed to be projected from each transitive verb.
tipronominal contexts. Parasitic gaps, topicalization gaps, right node raising gaps and others are barred from antipronominal contexts, as shown in (16) (Characteristic 4). Under the view that these gaps are “pronominal” empty categories, antipronominal contexts bar all pronominal constituents, whether or not they are realized overtly. There appears to be no clear way of generalizing in this way for a GBUG structure sharing approach. If gaps and fillers are indistinguishable, there is no reason why a gap cannot appear in a position where its filler can appear.

(15) a. There are pickles/*them in that jar   
   b. They named the ship Abigail/*it   
   c. He hit himself on the finger/*it

(16) a. *What did he imagine t on the ceiling before observing that there were _ on the table?   
   b. *What did they name their ship t after naming their cat _?   
   c. *Which finger did he hit himself on t after breaking _?   
   d. *The book, there is _ on the table   
   e. *Abigail, they named the ship _?   
   f. *The foot, he hit himself on

(Postal 1994) notes that gaps filled by that clauses sometimes have properties of noun phrases (these gaps also cannot occur in antipronominal contexts). The gaps require case marking prepositions (Characteristic 5), as in (17a-b). Not only do the fillers not require these prepositions, they are ungrammatical with them, as in (17c) (Characteristic 6). A GBUG structure sharing account could not differentiate the filler from the gap in this respect. Under the empty category approach assumed here, the filler is a that clause and the gap a pronominal type of empty category (an N which requires case).

(17) a. That Sonia attend the interview, I couldn’t insist on _ (Postal 1994, p. 70)   
   b. *That Sonia attend the interview, I couldn’t insist _ (Postal 1994, p. 70)   
   c. *I couldn’t insist on that Sonia attend the interview

German free relatives raise a problem (characteristic 8) for structure sharing because morphological case is shared. In (18) from Ingrin 1990, the relative pronouns was simultaneously satisfies the accusative requirement of gegeessen and the nominative requirement of the gap following was. An empty category approach would allow the gap to bear a different case than the filler. Was can satisfy either case requirement, but not both simultaneously in a unification-based approach. Thus for purposes of case agreement, the gaps morphological features would be identical to
the lexical morphological features of the lexical entry of was. Agreement
with filler and gap would represent distinct agreement checks via unifi-
cation (cf. Ingria 1990 for more similar data and a different approach.)

(18) Ich habe gegessen was noch übrig war
     I have eaten (acc) what (nom or acc) still left was (nom)
     'I ate what was left.'

This section has shown that a simple account of structure sharing
cannot handle a range of filler/gap phenomena. The next section shows
that a more elaborate version of structure sharing is limited in much the
same way.

7.5 Structure Sharing in HPSG

In the versions of HPSG presented in Pollard and Sag 1994, Sag 1996
(and others), part of the FS representing a filler is shared with the
gap, rather than the whole constituent. The particular portion of the
FS shared depends on the filler/gap construction. This section shows
that HPSG structure sharing analyses face all of the problems for struc-
ture sharing posed in Section 7.4. Some minor revisions to HPSG are
suggested which enable HPSG structure sharing to handle a few (but not
all) problematic cases.

With respect to long distance dependency constructions, examples
(11), (16), (17) and (18) pose the same difficulties for HPSG as for GBUG.
Each HPSG long distance dependency analysis includes one instance of
structure sharing between the “local” features of some filler and the lo-
cal features of the gap. Local features are all the features representing
the constituent except for those special features used to keep track of
long distance dependencies. e.g., SLASH. Since all the features which
have parallels in GBUG analyses are shared, the same problems arise.\textsuperscript{11}
Details are omitted for the sake of brevity.

I am also unaware of any structure sharing account which handle
the sloppy coindexing properties of VP ellipsis, gapping and right node
raising constructions, e.g., (13). Blevins 1995 proposed a structure shar-
ing account of gapping in which the FS representing the gapped verb
was shared. In the question/answer period, he claimed that only the
syntactic properties were shared. It is unclear how this idea would be
formalized.\textsuperscript{12}

\textsuperscript{11}For relative clauses, the relative pronoun is assumed to fill the gap. The head
modified by the relative shares the index of the relative pronoun. In the case of WH
questions, a feature QUEUE is used to identify the WH filler. QUEUE percolates up from
a given wh word to the wh phrase (cf. Ginzburg 1992). Above, I use WH values of
CAT in much the same way, e.g., it which anchors a phrase of category WH – noun.

\textsuperscript{12}This was a short talk, so details were omitted. One version of this idea might
For raising and equi constructions, no gap is assumed. The HPSG analysis for these constructions bears some similarities to the HPSG analysis of pronominal coindexing. For all verbs, state of affairs relations (equivalent to theta roles) are assigned to indices of constituents rather than the constituents themselves. Each referential constituent has a feature INDEX whose value is assumed to represent the referential index of that constituent (similar to Discourse Representation Theory). Indices (the values of INDEX features) are also values of the feature SUBJ (the subject index) and COMPS (the list of complement indices). Raising and equi verbs require that the value of their SUBJ (subject equi, subject raising) or COMPS feature (the controller) is also the value of the SUBJ feature of their complement. For both raising and equi verbs, a lower predicate bears a state of affairs relation to the subject. The controller of equi verbs, but not raising verbs are also in state of affairs relations with the main verb. Figure 92 is a simplified version of an HPSG equi analysis. The infinitival phrase (assumed to be a VP) has no constituent position for the subject, just the index value of the SUBJ feature. The bearer of the shared index is marked as both a wanter and a lever. The HPSG analysis of equi and raising has the same consequences as the GBUG analysis above. For example, selection restrictions associated with state of affairs relations must be met by bearers of index values of state of affairs features. A constituent whose index is a value of multiple state of affairs features must obey the selection restrictions associated with each relation.

Like the "controlled" subjects of raising and equi constructions, pronouns also share indexes with their antecedents. These indices include agreement features of the \( N_s \). Unfortunately, this approach lacks a way to accurately represent types of pronominal coreference other than those of the SAME-INDEX variety. A revised approach to indexing is needed to show that they in (19a) takes split antecedents or that himself in (19b) is coindexed with both John and Bill separately, via a sloppy coindexing relation (RNR gap/filler relation is a necessary part of this analysis).

Agreement between coindexed items must also be dealt with differently: it is unclear how the plural agreement features of they should be matched against the two singular features of its antecedents in (19a). The corresponding examples with gaps pose exactly the same problem, as in (20). Given that the subjects of infinitives are treated as indices and not "invisible" constituents, accounts must be able to treat these indices involve the assumption that the value of the RELN feature (equivalent to theta role assigner) is shared for the verbs/verb phrases in (13).
Figure 92 An HPSG analysis of Mary wanted to leave
like pronominals with respect to their ability to take split antecedents, and their ability to occur without fillers. Since pronominal coindexing and nonfinite subject coindexing are already handled in similar ways, it seems like the solution can be unified under an alternative treatment of indexing, perhaps like the GBUG approach above. Topicalization gaps and other gaps which act like pronouns (and therefore like $\Lambda$s) could be analyzed using shared indices rather than shared constituents (cf. example (17)). The antipronominal contexts would bar certain “pronominal” indices (indices of pronouns and pronominal gaps) rather than pronouns. However, examples like (17) would still be problematic since case is not part of the index.\footnote{Making case part of an index would raise other problems as discussed, e.g., see example (18)}

(19) a. John$_i$ asked Mary$_j$ whether they$_{ii,j}$ had a deal
    b. John$_i$ wanted to and Bill$_j$ actually did buy himself$_{ii,j}$ a car

(20) a. John$_{ii}$ talked to Mary$_j$ about PRO$_{ii,j}$ making a deal
    b. John wanted to ___$_{ii}$ and Bill actually did ___$_{ii,j}$ leaves$_{ii,j}$

In summary, all of the problems discussed in the previous section carry over to HPSG, except perhaps for cases where HPSG does not assume a gap, i.e., subjects of nonfinite clauses (and various predicates). In such cases, it seems that whatever theory binds pronominal indices could apply to these indices as well. Many of the problematic examples could also be helped by assuming that those gaps take the form of pronominal indices rather than “missing” constituents. This potential solution depends on a revised approach of HPSG coindexing.

7.6 What are GBUG Empty Categories?

7.6.1 Two Approaches

In GBUG, empty categories are part of FSs anchored by other items. The empty categories have neither independent lexical entries, nor phonetic representations. For example, Figure 93 is a lexical entry for the gerund eating. PRO is a property of the lexical entry, not an independent item. Similarly, each $\text{N}$ anchors a topicalization construction, which includes a FS representing a coreferential empty category at the target of a dangling arc. This topicalization construction can combine with a full sentence that is missing some specifier or complement to form a complete sentence as in Figure 94. The mechanisms described in the previous chapter for filling dangling arcs of long distance dependencies apply (some details are omitted.) The topicalization gap is a property of the topicalization construction anchored by the $\text{N}$, not an independent item.
Eating pasta is great

FIGURE 93 PRO in the lexical entry for the gerund eating
Mary, I know

Figure 94: Forming a topicalization from a topicalized entry
Previous P & P approaches (Chomsky 1973 and subsequent work) treat empty categories like special words that are either base generated like “normal” words or are “inserted” by transformations. If base generated empty categories are possible realizations of null strings, an untenable theory results. Where $\epsilon$ is the empty string and $X$ is an empty category, a phrase structure rule of the form $\epsilon \Rightarrow X$ allows an infinite number of $X$s to be generated anywhere in any sentence. $^{14}$ The correct derivation for a sentence may never be found, because an infinite number of spurious derivations may be tried first. The set of analyses generated for the utterance John laughed include phrase markers in which one million empty categories precede John, two million empty categories precede John, etc. If base-generated empty categories are properties of other lexical items, as proposed here, no such problem arises. $^{15}$

Under most current P & P approaches, transformations (move-$\alpha$) can move anything anywhere, the result being filtered by well-formedness conditions (Chomsky and Lasnik 1977). If base-generated empty categories are allowed to move, then an infinite number of empty categories may move to an infinite number of positions. Under GBUG this combinatorial explosion may be avoided. The effect of movement in long distance dependency constructions is obtained through the use of dangling arcs and tag adjunction. These operations are like lexically anchored transformations.

7.6.2 Empty Categories as Independent Constituents

There have been various attempts to show that certain empty categories are phonologically or psychologically independent constituents. It turns out that that structure sharing analyses and empty category approaches are equally compatible with every one of these findings (cf. Sag and Fodor 1994, Fordor 1993).

Human sentence processing studies (e.g., Bever and McElree 1988, MacDonald 1989, Nicol and Swinney 1989, Hickok and Nicol 1993), for example, show that information about the filler is “activated” at gap sites. This evidence is equally compatible with syntactic models in which the gap is modeled as an empty category or a “shared” instance of the filler.

In Miyagawa 1989, it is argued that passive and unaccusative subjects act like active unergative objects in sentences like (21) (Miyagawa’s ex-

---

$^{14}$For example, suppose empty categories are introduced by $X$ rules of this form.

$^{15}$Correa 1987, 1988, 1991 circumvents this problem by introducing empty categories with Attribute Grammar (AG) rules rather than ordinary phrase structure rules. These rules insert empty categories into phrase markers if any empty syntactic position meets a set of criteria. Empty categories are not “looked for” in a string of words, but rather are created on the basis of lexical items found in the input string. In this respect, Correa’s approach is like the one presented here.
amples (66), (66), (81) and (67)). Quantifiers in object positions modify either “real” objects (21a) or NP-traces in object position: the trace of the passive in (21b) and the trace of the unaccusative in (21c). Sentences lacking NP-traces cannot allow similar quantification of the subject, as evidenced by (21d). Structure sharing and empty category approaches are equal compatible with these observations. They both model the fact that the subject of passives and unaccusatives “act” like they are in the position of the gap.\footnote{A gap–free HPSG account explains the data equally well. The HPSG index $I$ in the \textit{COMPS} list of the verb is the value of the \textit{REST IND INDEX} path of the quantifier’s lexical entry (equivalent to marking the $\overline{N}$ a GBUG quantifier).}

(21) a. Boku wa yuumei na gakuya ni 3-nin atta
   I TOP famous scholars DAT 3-CL met
   ‘I met three famous scholars’

b. Yuube, kuruma ga dorobo ni 2-dai musum-are-ta
   last night cars NOM thief by 2-CL steal-PASS-past
   ‘Last night, two cars were stolen by a thief’

c. Gakusei ga ofisu ni 2-ri kita
   students NOM office to 2-CL came
   ‘Two students came to the office’

d. *Tomodati ga Shinjuku de Tanaka-sensei ni 2-ri atta
   friends NOM Shinjuku in Prof. Tanaka DAT 2-CL met
   ‘Two friends met Professor Tanaka in Shinjuku’

7.6.3 \textit{Wanna} Contraction

Aoun and Lightfoot 1984 and other have argued that contraction of \textit{want} and \textit{to} to form \textit{wanna} can be blocked by the presence of a case-marked trace, with the further stipulation that \textit{want} must govern \textit{to} in order for contraction to take place.\footnote{As noted in Meyers 1994, Aoun and Lightfoot’s definition of \textit{government} is non–standard and may cause problems for other P & P analyses, e.g., the PRO-theorem. Other definitions of government would prevent all well-formed instances of \textit{wanna} contraction (cf. Postal and Pullum 1986, pp. 105–106). For the sake of discussion, I will ignore these problems and assume that the government requirement limits \textit{wanna} contraction to instances in which \textit{to} heads the complement of \textit{want}.}

This analysis of \textit{wanna} contraction extends to contraction of various verbs and semi-modals including \textit{have}, \textit{got}, \textit{going}, \textit{ought}, \textit{use} with infinitival \textit{to}, forming \textit{hafta}, \textit{gotta}, \textit{gonna}, \textit{oughta} and \textit{usta}.

The idea is that case, a phonological feature, provides traces with some abstract phonological or psychological presense (cf. Chomsky 1980b, pp. 159–160). Thus a case-marked trace blocks contraction, just like any phonologically present item would. This account permits contraction
(22), but blocks contraction in (23). A case marked trace intervenes between want and to in (23a), blocking contraction in (23b). The trace in (22a) does not block contraction because it is not case-marked. In examples (23c–h) to is not governed by want. The to infinitive is the subject of the lower clause in (23c–d) (the external-theta licensee of the complement clause and the comp-case licensee of the matrix clause). to is assumed not to be governed by want due to intervening nodes, e.g., the infinitive is dominated by some ➞ node. The adjacent want and to in (23e–f) are in different subordinate clauses: want is in a relative clause and to heads a matrix complement. On the “in order to” reading of (23g–h), the to clause is an adjunct rather than a complement of want.

(22)  
a. They want to leave?  
 b. They wanna leave

(23)  
a. Who do you want t to leave?  
 b. *Who do you wanna leave?  
 c. *I don’t want [to flagellate oneself in public] to become standard practice in this monastery (Postal and Pullum 1982, p. 124, 3b)  
 d. *I don’t wanna flagellate oneself in public to become standard practice in this monastery (Postal and Pullum 1982, p. 124, 3c)  
 e. I don’t want anyone [who continues to want] to stop wanting (Postal and Pullum 1982, p. 125, 7a)  
 f. *I don’t want anyone [who continues to wanna] stop wanting (Postal and Pullum 1982, p. 125, 7b)  
 g. One must want (in order) to become an effective overconsumer (Postal and Pullum 1982, p. 126, 9a)  
 h. One must wanna become an effective overconsumer (no “in order to” reading) (Postal and Pullum 1982, p. 126, 9b)

On the alternative approach, a verb can only undergo wanna contraction if it shares its subject with its complement (subject raising and subject equi verbs) (Postal and Pullum 1982 p. 130). In the Meyers 1994 version of this approach, a phonology-less lexical entry for to unifies with a phonology-less lexical entry for want to form the phonology-less lexical entry for wanna. In Figure 95, the lexical entries for want and to unify to form the lexical entry for wanna if you ignore the phonology and if you also assume that INFL and Verb unify to INFL. The phonologica l merging of want and to is handled separately. If such unification is a necessary condition for wanna contraction, it is clear that only subject raising and subject equi verbs can participate. The

\footnote{According to Pullum 1982, to is a type of verb. Assuming that is correct, the second stipulation is unnecessary.}
SPECIFIER arc of to unifies with either the EXT-THETA arc of subject equi (control) verbs or the SPECIFIER arc of subject raising verbs. The INT-THETA SPECIFIER path of to unifies with the identical arc in raising and control verbs.

In Meyers 1994, I showed that the subject sharing approach to wanna contraction also applies to contractions of modals and the auxiliary have into should’ve, could’ve, etc. Figure 96 illustrates this. The lexical entry for have is very similar to the lexical entry for to and can readily combine with the entry for a modal like should just as want and to combine in Figure 95. Modals with subject control properties, e.g., ought can also combine with have, as can ambiguous modals like can. Just like wanna contraction, have contraction is limited to items which share the subjects of their complements.

A debate between case-marked-trace advocates (Chomsky and Lasnik 1978, Chomsky 1980b, 1981, Jaeggli 1980, Aoun and Lightfoot 1984) and shared-subject advocates (Postal and Pullum 1978, 1979, 1982, 1986, Carden 1983) ended with the formulation of the case-marked-trace approach noted above. Aoun and Lightfoot 1984 (p.472, note 8) offer the only criticism of the subject sharing approach I am aware of – they assert that the structure sharing approach is more of a descriptive generalization than an analysis because it is far too complex to meet general requirements. It is unclear what this means. Perhaps this reflects the difficulty of representing the subject sharing generalization in terms of previous versions of P & P theory. In this sense, the generalization is not “too complex” for GBUG.

The crucial flaw with the case-marked trace position is its universality. All case marked traces should block all types of contraction. However, as shown in Carden 1983, case marked traces fail to block contraction of is and the preceding word in examples like (24) (Carden’s examples (25) and (26)).

(24) a. Who t is going
   b. Who do you think t is gonna to win?
   c. Who’s going
   d. Who do you think’s gonna to win?

Under the case marked trace approach, want and to must be adjacent for wanna contraction to take place, although the definition of adjacency must be adjusted for across-the-board phenomena as in (25). (Postal and Pullum 1982, p. 126, 10, 11). Unfortunately, this definition

19Following Lakoff 1970 and Pullum and Wilson 1977, the is likely to reading of can is subject raising and the is able to reading is subject control.
Lexical Entry for wanna

Lexical Entry for to
Figure 36: A Unification based version of have Contraction

Lexical Entry for **should**

Lexical Entry for **have**

Lexical Entry for **should’ve**
of adjacency does not hold for other contraction phenomena, as shown in Carden 1983. His examples (18a) and (19a) are repeated as (26).

(25)  
a. I want to dance and to sing  
b. *I wanna dance and to sing  
c. I don’t need or want to hear about it  
d. *I don’t need or wanna hear about it  

(26)  
a. I’m ready and will leave at once  
b. Either Tom or Dick’s got it  

7.6.4 Summary

I advocate an approach in which empty categories are properties of lexical entries, rather than independent word-like items. Under this approach, unlike its alternative, derivations for simple sentences terminate.

Psycholinguistic and linguistic studies that might differentiate these approaches based on human data have all been inconclusive. The most popular study of this type concerned two generalizations seeking to describe wanna contraction: the case marked trace approach and the subject sharing approach. Only the case marked trace approach presupposes that empty categories are independent word-like items. Although both generalizations describe the wanna data presented here, the subject sharing approach is more general. The subject sharing analysis applies to at least two types of contraction, while the case-marked trace analysis describes exactly one type of contraction. Furthermore, the case marked trace approach makes predictions about other types of contraction which do not hold.

7.7 Possible Misconceptions about Binding Theory

Above, structure sharing analyses are posited for equi, passive and raising constructions. This section addresses some possible misconceptions regarding empty category based analyses of these constructions.

Binding theory serves to constrain the possible antecedents a particular pronominal may be bound by. In (27) himself may be bound to either John or Bill. Under the empty category analyses of (28), the NP-traces must be bound by the surface subjects, and not the other Ns preceding the traces (objects and prepositional objects). Binding Theory restricts the possible antecedents, but does not determine which N must be the antecedent. Principle A is not sufficiently strict in its application to NP-traces because NP-traces have a precisely defined antecedent, unlike true anaphors. It is unclear what is gained by assuming that NP-traces are constrained by binding theory.
(27) a. John sold Bill a picture of himself
    b. John bid against Bill for a picture of himself
(28) a. A toy was given the little girl t
    b. John was believed by Bill t to be a linguist
    c. John seemed to Bill t to be a linguist

On the view (discussed above) that the obligatorily controlled PRO of equi constructions is an anaphoric empty category, the same objection arises. Particular lexical entries determine whether that PRO is bound by the matrix subject (29a) or object (29b). Binding theory is not needed to make this determination. There are ambiguous cases, e.g., (29c), but only for verbs with both subject and object control lexical entries.

(29) a. John promised Bill PRO to leave
    b. John told Bill PRO to leave
    c. John asked/badgered Mary PRO to leave

7.8 Summary

I compare empty category and structure sharing devices in terms of capacity to model specific linguistic phenomena. Although structure sharing analyses require fewer stipulations than empty category analyses, structure sharing is too constrained for representing linguistic phenomena in which the gap and filler need to be distinguished. Considerations for choosing empty category or structure sharing analyses were discussed. Alternative versions of empty category and structure sharing analyses were also investigated.
8

Command Relations and Binding Theories

8.1 What are GBUG Command Relations?

Relations between nodes or arcs, called COMMAND RELATIONS, are used in P & P, APG, LFG and HPSG (among other theories) to define constraints on coreference and disjoint reference (Binding Theory), to define constraints on empty categories and to define other constraints and relations. This chapter provides general means for representing command relations between items in FSs and formulates a version of binding theory. This binding theory is compared to previous P & P, and HPSG binding theories. Additionally, an LFG constraint on nonobligatory control is translated into GBUG terminology.

As set forth in Table 9, arc $X$ $\alpha$ COMMANDS arc $Y$ if the relative positions of $X$ and $Y$ in a GBUG FS meet a set of conditions, where $\alpha$ is a variable over sets of arcs$^2$ and instances of $\alpha$ command differ by the set of arcs $\alpha$ over which they are defined. An arc $X$ c-commands (constituent commands) an arc $Y$ if $\alpha$ is the set of constituent licensor and licensee arcs (all licensor/licensee arcs); arc $X$ predicate commands arc $Y$ if $\alpha$ is the set of predicate licensor and licensee arcs; and arc $X$ surface commands arc $Y$ if $\alpha$ is the set of surface arcs. Other possible command relations include: theta command, scope command, agreement command and case command. Command relations originate with Langacker (cf. Langacker 1969). C-command, the command relation most commonly employed in P & P theories is usually attributed to Reinhart (cf. Reinhart 1976). Command relations that are defined between arcs,

$^1$For an account of the mathematical properties of command relations defined on constituent structure trees, see Kracht 1993

$^2$In Myers 1994, I used the term X-command instead of $\alpha$ command.
The Dominates Relation:
A. A node $a$ dominates a node $b$ if there exists a path with source $a$ and target $b$
B. A node $a$ dominates an arc $B$ if there exists a path $P$ with source $a$, such that the target of $B$ is the target of $P$
C. An arc $A$ dominates a node $b$ if the source of $A$ dominates $b$
D. An arc $A$ dominates an arc $B$ if the target of $A$ dominates $B$

Maximal Projection: A maximal projection is an X which is not the value of any Head-Proj arc (e.g., the X head of a small clause is not a maximal projection).\(^3\)

$\alpha$ Branching Node: An $\alpha$ branching node is a node which is the source of more than one arc of type $\alpha$.

$\alpha$ Command: Arc $A$ $\alpha$ COMMANDS arc $B$ if:
A. Arcs $A$ and $B$ are of type $\alpha$
B. Every $\alpha$ branching node that dominates $A$ also dominates $B$
C. $A$ does not dominate $B$
D. $B$ does not dominate $A$

Maximal $\alpha$ Command: Arc $A$ maximal $\alpha$ COMMANDS arc $B$ if:
A. Arcs $A$ and $B$ are of type $\alpha$
B. Every $\alpha$ branching root of a maximal projection that dominates $A$ also dominates $B$
C. $A$ does not dominate $B$
D. $B$ does not dominate $A$

$\alpha$ GOVERNMENT: $A$ $\alpha$ GOVERNS $B$ if $A$ Maximal $\alpha$ commands $B$ and $B$ Maximal $\alpha$ commands $A$.

| Table 9 | Definition of Command Relations and Related Terms |

rather than nodes, originate with Johnson and Postal ($\text{arc-command}$ is defined in Johnson and Postal 1980, p.257).

Following work in APG and LFG, GBUG command relations are defined on arcs instead of nodes. In previous versions of P & P, command relations are defined on nodes in trees. Let’s compare some arc-based and node-based versions of the binding theory outlined in Table 10, where domains $\text{Dom}_A$, $\text{Dom}_B$ and $\text{Dom}_C$ are defined in later sections. The definition of the word $\text{bind}$ differs for arc-based and node-based versions of binding theory as in Definitions 1 and 2. The word $\text{free}$ means “not bound”.\(^4\) Both the arc-based and node-based versions of binding theory

---

\(^3\)Here $\alpha$ command includes both arc-based and node-based command relations. In the rest of this chapter, only arc-based command relations are included under $\alpha$ command.
Principle A: An anaphor must be bound within a defined domain $\text{Dom}_A$.

Principle B: A pronoun must be free within a defined domain $\text{Dom}_B$.

Principle C: An r-expression (a non-pronoun/non-anaphor $\overline{N}$) must be free within a defined domain $\text{Dom}_C$.

Table 10 A Generic Binding Theory

entail that: (1) an anaphor coindexed with $P$ must be $\alpha$ commanded by $P$; and (2) an r-expression that is coindexed with a phrase $P$ cannot be $\alpha$ commanded by $P$. It will be shown that the arc-based, but not the node-based version of binding theory properly characterizes the binding relations in Figure 97. This reflects that the node-based version of binding theory assumes a phrase structure tree, rather than a DAG model. Failure of the node-based version of binding theory to apply correctly here is a limitation and not a defect, since the node-based theory was not devised with DAGs in mind.\textsuperscript{5}

Definition 7 A node $a \alpha \text{BINDS}_{\text{node}} b$ if $a$ and $b$ are coindexed and $a \alpha \text{commands}_{\text{node}} b$.

Definition 8 A (constituent rooted at) node $a \alpha \text{BINDS}_{\text{arc}} (a \text{constituent rooted at})$ a node $b$ if: nodes $a$ and $b$ are coindexed; and arc $A$ with target $a \alpha \text{commands}_{\text{arc}}$ arc $B$ with target $b$.

Consider an arc-based binding theory defined in terms of predicate command. In Figure 97, every predicate branching node dominating the matrix $\text{EXT-THETA}$ arc dominates the $\text{MAC1}$ arc immediately above $\text{himself}$, but the reverse does not hold. Thus $\text{HEAD INT-THETA EXT-THETA}$ predicate commands $\text{HEAD INT-THETA HEAD MAC1}$ where the paths are short-hand for the final arcs in those paths. Principle A of the binding theory requires $\text{John}$ to bind $\text{himself}$ because $\text{John}$ is the target of the matrix $\text{EXT-THETA}$ arc, the only arc (in the domain $\text{Dom}_A$ of $\text{himself}$) which predicate commands the predicate licensor/licensee arc ($\text{MAC1}$) with $\text{himself}$ as the target. Principle C of Binding Theory does not rule out this coindexing since no predicate licensor/licensee arc with $\text{himself}$ as a target predicate commands an arc with $\text{John}$ as the target.

Now consider versions of binding theory based on the following node-based command relations:

\textsuperscript{5}In HPSG style analyses, where indices are shared rather than constituents, node-based command relations may suffice.
John promised himself to rest.
C-Command\textsubscript{node}: A node $x$ c-commands\textsubscript{node} a node $y$ if every branching node dominating $x$ dominates $y$, and neither $x$ nor $y$ dominates the other.

Predicate Command\textsubscript{node}: A node $x$ predicate commands\textsubscript{node} a node $y$ if every predicate branching node dominating $x$ dominates $y$, and neither $x$ nor $y$ dominates the other.

In Figure 97, *himself* both c-commands\textsubscript{node} and predicate commands\textsubscript{node} *John*, violating Principle C of the node-based binding theories. *John* also c-commands\textsubscript{node}/predicate commands\textsubscript{node} *himself*, a fact that is consistent with Binding Theory (Principle A). Node based command relations are designed for tree-based models, where only sister constituents can mutually c-command each other. In DAGs, mutual (node-based) c-command is not limited to sisters, due to structure sharing (*John* in Figure 97). In Figure 98, a tree version of the sentence in Figure 97, *John* c-commands\textsubscript{node} *himself*, but *himself* does not c-command\textsubscript{node} *John*, as desired.

8.2 Binding Theory Domains in GBUG

This section proposes the binding theory in Table 11, an instance of the theory in Table 10 in which domains $Dom_A$, $Dom_B$ and $Dom_C$ for a given $X$ are its $\alpha$ governing category as per Definition 3. This binding theory makes similar predictions to the binding theory in Chomsky 1986b, once various fillers are reconstructed in the position of their gaps, although, this reconstruction is unnecessary given a GBUG structure sharing approach.

**Definition 9** The $\alpha$ GOVERNING CATEGORY of $X$ is the smallest maximal projection containing an $N$ specifier $S$ and some item $G$, where $S$ does not contain $X$ and $G$ $\alpha$ governs $X$.

Specifically, lets assume that $\alpha$ command is predicate command.\(^6\) (1) includes a range of phenomena that previous P & P binding theories as well as the proposed binding theory account for. Same indices indicates that binding theory permits these items to be coindexed. Different indices means that binding theory prohibits these items from being coindexed. No indices means that binding theory does not address whether these items are coindexed. In (1b,f) the predicate governing categories for *himself* and *them* are $\overline{N}$s with possessive $\overline{N}$ specifiers. In (1c,g), the

---

\(^6\)One alternative would be to assume that $\alpha$ command is theta command. The examples presented in this section do not choose between these two theories. The crucial difference may be that some adjuncts (e.g., modifiers) and specifiers (e.g., quantifiers) of a head predicate command, but do not theta command complements of that head. I cannot construct any examples for which this fact makes a difference.
**Principle A:** An anaphor $X$ must be bound within $G$, the $\alpha$ governing category of the smallest projection dominating $X$ that does not violate the following constraint:

The $i$ within $i$ condition: A constituent $C$ cannot be coindexed with $C'$ if the root of $C$ dominates the root of $C'$ ($C'$ is part of $C$).

**Principle B:** A pronoun $X$ must be free within the $\alpha$ governing category of $X$

**Principle C:** An $r$-expression $X$ must be free (within the $\alpha$ governing category of the head of $X$’s chain)

**Table 11** GBUG Binding Theory

Predicate governing categories are the if clauses. For the remaining examples, the matrix clause is the predicate governing category. In (1c), binding theory allows an antecedent to occur in the matrix clause to avoid the following *i within i* violation (the subject of the if clause contains the anaphor). Examples (1d,h) contain pro-forms that are in passivized objects. Binding theory makes the correct predictions provided that the theta arc dominating the object of the *by* phrase predicate commands the internal-theta arc, i.e., if the verb assigns a theta role to the object of *by* (see Section 4.4.4).

(1) a. John$_i$ gave Bill$_j$ a picture of himself$_{i\omega j}$
   b. John bought Bill’s$_i$ picture of himself$_i$
   c. They$_i$ wondered if pictures of each other$_i$ were in the gallery
   d. A picture of herself$_i$ was stolen by Mary$_i$
   e. [John and Mary]$_i$ saw a picture of them$_j$
   f. [John and Mary] saw Fred’s picture of them
   g. [John and Mary]$_i$ wondered if pictures of them were in the gallery
   h. Pictures of her$_i$ were stolen by Mary$_j$

The $r$-expressions in (1) are all free. The binding theory above states in parentheses that an $r$-expression must be free within the $\alpha$ governing category of the head of its chain. Chomsky (e.g., Chomsky 1986b) assumes that gaps of long distance dependencies (WH-traces) are $r$-expressions. Given this assumption, then WH-traces must be free within the $\alpha$ governing category of their antecedents. In the GBUG accounts above, WH extraction may be handled by structure sharing. In that case the arc relevant to binding theory is the highest predicate licensor/licensee arc with the extracted element as its target. If all long distance dependencies are handled by structure sharing, then the parentheses are unnecessary. In at least some cases, the previous chapter
shows that gaps of long distance should be viewed as empty categories and that some of these empty categories are pronominal. Due to these complexities, it is difficult to determine exactly how (and whether) binding theory should treat WH-traces and other empty categories. If binding theory only applies to overt elements, Principle C need only read “An r-expression must be free”.

Another complexity which binding theories need to account for is that different languages require different binding theory domains (DOM$_A$, DOM$_B$ and DOM$_C$) and different pro-forms in the same language may require different domains (see for example Wexler and Manzini 1985). This restriction is addressed above by using the variables DOM$_A$, DOM$_B$ and DOM$_C$ for binding theory domains and assuming that these variables are set by language specific and even lexical considerations. I believe that a complete binding theory would have to incorporate a default inheritance hierarchy, much like the account of word order constraints in Chapter 4.

8.3 LFG’s F-command and GBUG’s Predicate-command
Bresnan (Bresnan 1982a, p.386) provides the following definition of f-command:

For any occurrences of the functions $\alpha$, $\beta$ in an F-structure $F$, $\alpha$ f-commands $\beta$ if and only if $\alpha$ does not contain $\beta$ and every F-structure of $F$ that contains $\alpha$ contains $\beta$.

Functions in this definition refer to grammatical relations like subject, object, etc. These are represented graphically as arc labels in an F-structure, an LFG FS representation of predicate argument structure (in contrast with C-structure which represents surface constituent structure).

Predicate command ignores all licensor/licensee arcs that do not represent some predicate argument relation, just as f-command is defined on a structure that represents purely a predicate argument structure. Therefore, it might be expected that analyses in terms of predicate command and f-command can be easily translated from one to the other.

Figure 99 is an f-structure for the sentence *People who know John discuss working too hard* from Bresnan 1982a (p. 387). Bresnan 1982a requires that all possible antecedents of PRO f-command it (the Universal Condition on Anaphoric Control). *People who know John*, but not *John* may be coreferential with the PRO subject of *working too hard* because the Matrix SUBJECT arc f-commands the SUBJECT arc of the matrix object. *(SUBJ f-commands OBJ SUBJ)* Figure 100 is the GBUG representation of the same sentence. Given the assump-
Figure 99 F-structure for *People who know John discuss working too hard*
FIGURE 100 GBUG FS for People who know John discuss working too hard
- \textit{EXT-THETA} and \textit{Subject} are the same relation. \textit{Object} is a type of \textit{INT-THETA} relation.
- The category \( S \) and the category \( I \) are the same category.
- It is possible to control for the following structural difference between LFG and P & P analyses: LFG assumes that predicates (heads), subjects and objects form a single constituent at F-structure; /PnP/ assumes that \( \overline{N} \)s are \textit{EXT-THETA} licensors and \( X^0 \)s are \textit{INT-THETA} licensees.\(^7\)

\begin{table}[h]
\centering
\caption{Mappings between LFG and GBUG predicate relations}
\end{table}

The parallel predicate command based constraint holds as well. \textit{People who know John}, but not \textit{John} may be coreferential with the \textit{PRO} subject of \textit{working too hard} because the Matrix \textit{EXT-THETA} arc predicate commands the \textit{EXT-THETA} arc dominating \textit{PRO}. (\textit{HEAD INT-THETA} \textit{EXT-THETA} predicate commands \textit{HEAD INT-THETA HEAD MAC EXT-THETA}).

This is just one example in which basically the same analysis may be stated either in terms of \textit{f}-command or predicate command. The salient differences between these theories of predicate argument structure seem to be structural, e.g., are subjects and objects sisters at predicate argument structure as in LFG or is the subject predicated of the \( \overline{V} \) (the verb plus complements) as in P & P?\(^8\)

\subsection{8.4 HPSG's o-command and GBUG's predicate command}

HPSG's binding theory is based on the relations o-command and local o-command, defined in Table 13.\(^9\) These command relations take

\footnote{\textit{Marantz 1984} (pp. 48-51) argues that the verb and all its complements assign the external theta role, due to different external theta roles for \( \overline{N} \)s headed by the same verb, but with different complements, as in the following examples:
(i) throw a party; throw a fit; take a bus to New York
(ii) throw a baseball; take a book off the shelf
Each of the phrases in (8) requires a sentient subject (agent), whereas (8) requires an animate subject. A human (e.g., Fred) can be the subject of the verbs in (8) or (8). A machine (a robot arm) can be a well-formed subject for the verbs in (8), but would be anomalous with the verbs in (8).}

\footnote{The following clarifications are needed for those not fluent in HPSG. \textit{Synsem} objects are syntacto-semantic FS representations of words or phrases (not phonetic representations, not discourse representations, etc.). For HPSG, all \( N \)s which are not expletives are \textit{referential}. Fillers and gaps in HPSG accounts of long distance dependencies (see the previous chapter) have shared "local" features, \textit{synsem} features other than those used to keep track of long distance dependencies (slash features). Thus
• local o-command

Let Y and Z be synsem objects with distinct LOCAL values, Y referential. Then Y locally o-commands Z just in case either:
i Y is less oblique than Z; or
ii Y locally o-commands some X that subcategorizes for Z.
(Pollard and Sag 1994, p.278)

• o-command

Let Y and Z be synsem objects with distinct LOCAL values, Y referential. Then Y o-commands Z just in case either:
i Y is less oblique than Z; or
ii Y o-commands some X that subcategorizes for Z; or
iii Y o-commands some X which is a projection of Z (i.e., the HEAD values of X and Z are token-identical).
(Pollard and Sag 1994, p.279)

Table 13 HPSG’s o-command and local o-command

into account the hierarchical structure of HPSG FSs as well as an ordering of grammatical relations called the obliqueness hierarchy, given as (2) (Pollard and Sag 1994, p. 24). The obliqueness hierarchy is based on the relational hierarchy or noun phrase accessibility hierarchy (Perlmutter and Postal 1974 and Keenan and Comrie 1977)\(^\text{10}\), which has previously been used by Ross 1974 and Johnson 1977 in Relational Grammar accounts of reflexivization.\(^\text{11}\)

\[(2) \quad SUBJECT < PRIMARY OBJECT < SECONDARY OBJECT < OBIQUE PP < VERBAL/PREDICATIVE COMPLEMENTS\]

Strictly speaking, the elements in the obliqueness hierarchy do not exist in P & P theory. However, equivalent terms or combinations of terms exist, sufficient that a GBUG binding theory based on predicate com-

---

\(^{\text{10}}\)Kim does not o-command the topicalization gap in (9) (from Pollard and Sag 1994, p.157) because Kim and the gap do not have distinct local features.

\(^{\text{i.}}\)Kim, Sandy loves ___

\(^{\text{11}}\)The primary object in HPSG is assumed to be the first NP complement. The second NP complement in a ditransitive structure is the secondary object. No equivalent of Relational Grammar’s direct and indirect objects are assumed.

\(^{\text{12}}\)As Pollard and Sag 1994 (pp. 275–77) note, Jackendoff 1972 (among others), seek to constrain coreference constraints on reflexives by means of the following (thematic) hierarchy of types of theta roles:

Agent < Location/Source/Goal < Theme

One can imagine a predicate command based approach that incorporates Jackendoff’s thematic hierarchy and assumes a flat constituent structure like HPSG and LFG.
o-bound/α-free. Y (locally) o-binds Z just in case Y and Z are coindexed and Y (locally) α-commands Z. If Z is not (locally) α-bound, then it is said to be (locally) o-free.

Principle A. A locally o-commanded anaphor must be locally o-bound.
Principle B. A personal pronoun must be locally α-free.
Principle C. A non-pronoun must be α-free.

TABLE 14 HPSG’s Binding Theory (Pollard and Sag 1994, p. 254)

command can make substantially the same predictions as an HPSG binding theory based on α-command and local α-command. For example, a subject of a verb V o-commands, f-commands and predicate commands the object of V. Although, the object of V f-commands V’s subject, it neither o-commands nor predicate commands V’s subject. This difference between subjects and objects is captured by the obliqueness hierarchy for HPSG’s o-command and by the assumed hierarchical structure of P & P constituents. 12

HPSG’s Binding Theory is given as Table 14. This theory correctly predicts the well-formedness or ill-formedness of the following examples from Pollard and Sag 1994 (pp. 245, 254 and 256) Subjects locally o-command objects of the same verb. Thus the pronoun like him in (3a) cannot be coreferential with the subject John, but himself must be coreferential with John in (3e). The determiner of the object is α-commanded, but not locally o-commanded by the subject. Thus their in 3c is not locally o-commanded by the children and coindexing is permitted. The anaphors in (3d,e,f) are not locally o-commanded by any noun phrases. Thus Pollard and Sag do not require them to be o-bound. Rather, they suggest that coreference in these cases are determined by discourse factors.

(3)

a *Johni likes himi
   b Johni likes himselfi
   c [The children]i like theiri friends
   d The childreni like [each other’s]i friends.
   e Why are [John and Mary]i letting the honey drip on each other’si feet? (Chomsky 1973, p.261)
   f Johni was going to get even with Mary. That picture of himselfi in the paper would annoy her, as would the other stunts he planned.

12 As noted above, LFG captures this asymmetry in the Lexical Rule of Functional Control.
The predicate command based version of binding theory described above would be equivalent to HPSG’s binding theory if: (a) the i-within-i condition is ignored; and (b) the following portion of the obliqueness hierarchy was taken into account:

\textit{PRIMARY OBJECT} > \textit{SECONDARY OBJECT} > \textit{OBlique}

Pollard and Sag specifically argue against the i-within-i condition, citing counterexample such as (4) from Pollard and Sag 1994 (pp. 245, 263). This and other examples lead them to the conclusion that such anaphors are not locally o-commanded and therefore need not be locally o-bound.

(4) John suggested that [tiny guilt frame pictures of [each other]]; would make ideal gifts for [the twins].

The fact that primary objects precede secondary objects and oblique objects in the hierarchy is used to capture the well-formedness judgements for coindexing in examples like (5) from Pollard and Sag 1994, pp. 255-256). Following previous P & P approaches, command relations can capture this same relation. If it is assumed that binding theory influences constituent structure, then a structure such as Figure 101 (when incorporated into a larger FS) would yield the intended precedence for some command relation. Similar structures have been proposed in Chomsky 1975, 1992, Larson 1988, Kayne 1994 and others. These authors have argued for binary branching constituent structures generally.

(5) a. *Mary described Bill to him
   b. *He sold the slave; him
   c. Mary described Bill to himself
   d. ?John sold the slave; himself

A preferred option in my view, would be to require that a node a binds a node b if a predicate commands b and there is no lexically based linear precedence constraint which orders b before a. The linear precedence constraints associated with the canonical verb entry are examined, rather than ones derived by some rule (e.g., passive). Thus if a verb takes two complements a and b, and orders a before b (see Section 4.5), b could not bind a, even if b precedes a due to passive or extraction. Thus the asymmetry in (6) is accounted for.

(6) a. ?The slave was sold himself.
   b. *Each other were sold the slaves.

8.5 Summary
This chapter formulated GBUG command relations and a GBUG-based binding theory. I compared this binding theory to previous ones in a number of frameworks. It turns out that LFG’s f-command, HPSG’s
o-command and GBUG’s predicate command are very similar relations. However, the data structures on which they are applied differ. One difference is simply terminology, e.g., the term subject is used similarly to the term ext-theta. The main difference appears to be whether asymmetries between grammatical relations are captured by hierarchical orderings of these relations or by the graph configurations in which these grammatical relations are encoded. Word order constraints may also be used in place of some hierarchical orderings.
Believe -type Verbs

9.1 Introduction

The N's following the main verb in examples like (1) are assumed to receive case from the main verb and to receive a theta role from the lower predicate. In GBUG, COMP-CASE is a type of complement relation and EXT-THETA is a type of specifier relation. Thus the N is assumed to be a constituent of both clauses. This makes the GBUG analysis of these verbs a version of the raising to object analysis (RO) (cf. Rosenbaum 1967; Postal 1974, among others), rather than the exceptional case marking (ECM) analysis (Chomsky 1980a, 1981, among others) or the related raising to SPEC of AGR-O analysis (RSA) (cf. Chomsky 1992) popular in many P & P theories.

(1) a. John believes ketchup to be a vegetable
    b. Cecil considered ketchup a vegetable
    c. Mildred made Herman angry
    d. John saw Irving stealing the documents

Under the ECM analysis, the matrix verb assigns case to the specifier of the subordinate clause, but that specifier is not assumed to be a constituent of the higher clause. This is problematic for GBUG because it contradicts the assumption that all licensing relations are types of constituent relations. In the RSA analysis, the N is a specifier of AGR-O in the matrix clause at LF where it receives case. However, it is not a complement of the matrix clause, as in the RO analysis and it is not a surface (or SPELL-OUT) constituent of the higher clause. The RSA analysis is problematic because LF positions are not clearly defined in the Minimalist Program (see Section 5.6). In GBUG, each constituent node must be the value of a particular licensor or licensee feature which represents some observable linguistic phenomenon. Minimalist LF Specifier posi-
tions do not pass this test. It therefore appears that these alternatives to the RO analysis are not compatible with GBUG based theories. Evidence is presented below from Postal and Pullum 1988 and Dougherty and Leacock 1993 that supports the RO analysis.

9.2 Evidence from Postal and Pullum 1988

Government and Binding Theory assumes the projection principle, which includes the requirement that all complements of a given head are theta marked by that head. As noted by Postal and Pullum 1988, this not only prohibits RO analyses, but also prohibits expletives from occurring in subcategorized positions, because one of the defining characteristics of expletives is that they do not bear theta roles (cf. Chomsky 1986b, pp. 131–144). Postal and Pullum show that contrary to this prediction, expletives do in fact occur in complement positions, as in (2), all taken from Postal and Pullum 1988. ECM analyses are unavailable for these examples. If Postal and Pullum’s arguments are correct, then non-theta-marked complements are possible and the RO analysis is available. In that case, the ECM analysis becomes undesirable, because it complicates the grammar by allowing a head to case mark the specifier of its complement. Postal and Pullum show that the significant difference between the RO and ECM approaches is that the former is based on a theory which bans subcategorized expletives and the latter is based on a theory which does not.

(2) a. I take it that you will pay
b. He never gave it a thought that Bolsheviks are human beings.
c. John would hate it for him to win.
d. They never mentioned it to the candidate that the job was poorly paid.
e. They hold it against me that I am an extraterrestrial.
f. We can prevent it, I assure you, from becoming known that we are here.
g. John will see to it that you have a reception.
h. He made a point of honor of it, I recall, to respect agreements he had negotiated.
i. I figured/made/reasoned it out.

1Other objections are based on positions which are no longer held. In particular, movement rules which do not change constituent order were once banned (Chomsky 1972), but are now allowed under most versions of P & Ptheory (e.g. Chomsky 1986a), as evidenced by examples like

(i) [Who, [I, made this coffee]] (Postal and Pullum 1988, no. 17)
j. You two are going to battle it out when things go wrong between you.
k. dish it out; does it; fight it out; give it to NP; have it out with NP; beat it; blow it; get it together; make it; make it snappy; wing it, etc.

One can imagine ECM analyses of a subset of (2) in which it is a specifier of the following constituent. Pullum and Postal reject such a possibility on grounds that sentences like (2d.e.f) because it is separated from its head by an intervening \( \overline{F} \). Barring discontinuous constituency, it is hard to support the claim that it belongs to that constituent on the surface.\(^2\)

Before Postal and Pullum 1988, some linguists have attempted to eliminate sentences like (2) from consideration. Postal and Pullum argue specifically against such claims. Williams 1980 (p. 222) claims that sentences like (3a) occur too sporadically. Chomsky 1981 (pp. 147–148) claims that sentences like (3b) are too idiosyncratic. As Postal and Pullum argue, it is irrelevant whether or not a particular construction is sporadic or idiosyncratic because informal measures of statistical frequency have nothing to do with whether or not a grammar handles a particular construction. (Actually, expletive it in complement position is not really sporadic given the wide range of examples Postal and Pullum provide.) Postal and Pullum suggest that Chomsky and Williams are attempting to dismiss problematic data.

(3) a. John saw to/regretted it that Bill had a good time.
    b. They forced it to rain (by seeding the clouds.)

Chomsky also claims that (3b) is *derivatvely generated*, a term previously applied to semantic anomaly (Chomsky 1965: p. 227; 1970), but the intended sense of *derivatvely generated* is left obscure in this context.

The fact that it does not front in passive is also given (Chomsky 1981) as a reason why (3b) is *derivatvely generated*. Once again it is unclear why passivization tests for being *derivatvely generated*. In fact many expletive complements do passivize, as in (4) from (Postal and Pullum 1988, p. 656).

\(^2\)See Stroik 1990 for arguments supporting a small clause analysis of the examples in (2) with that clauses. Stroik claims that the \( \overline{C} \) moves adjacent to the expletive in those examples. This argument is problematic because \( \overline{X} \) movement and small clauses headed by complementizers are rarely found in the literature. Stroik cites a few cases of \( \overline{X} \) movement from the literature, all of which I find questionable. Also, Stroik does not try to account for the other examples. See Meyers 1994 for further discussion.
(4) a. It was never mentioned to the candidate that the job was poorly paid.
   b. It is (always) held against me that I am an extraterrestrial.
   c. It was figured out (that everyone in the office was an extraterrestrial.)
   d. It has to be fought out.
   e. If it can be prevented from raining.....

Postal and Pullum’s analysis is generally supported by Lasnik and Saito 1991. Lasnik and Saito, however, claim that the following empirical problem exists, citing examples (5) and (6):

All of these are fully acceptable, and in all of them the it is reasonably regarded as an expletive. And if pleonastic objects are allowed, Postal and Pullum argue, there is no principled basis for rejecting raising to object position. This reasoning seems sound, but the actual patterning of facts is surprisingly discordant. In particular, there is very little correlation between verbs that take expletive objects and those that take infinitival complements with overt subjects, ... Thus while Postal and Pullum’s examples provide a conceptual basis for allowing raising to object, they provide little if any empirical basis. (Lasnik and Saito 1991, pp. 340-341)

Examples with Verbs taking Expletive Complements, but not infinitival raising to object complements:

(5) a. ??I dislike him to be so cruel.
   b. ??I didn’t suspect you to have failed
   c. *I regret them not to have hired Masconi
   d. *I resent you not to have called me
   e. *I don’t mind him to have done that

Examples with verbs that take infinitival raising to object complements, but not expletive complements:

(6) a. I believe (??it) that John left
   b. I will prove (??it) that Mary is the culprit
   c. They have found (??it) that there is a prime number greater than 17
   d. I will show (??it) that the Coordinate Structure Constraint is valid

However, I found that some verbs with RO infinitival complements do allow expletive complements when negated, or with modals as shown in (7). Therefore, the empirical problem noted by Lasnik and Saito doesn’t exist.

(7) a. I hardly believed it that John left
b. I didn’t believe it for a minute that John left

c. I can’t prove it that Mary is the culprit

d. I hate it that Mary is turning into a politician

In summary, Postal and Pullum show that the only viable analyses of the examples discussed above is one in which expletive it occupies a complement position and that this fact raises problems for the assumption that all complements are theta marked. Once this idea is abandoned all obstacles to an RO analysis are removed. Exceptional case marking involves a “unique” process in which an element case marks the specifier of its complement, rather than the complement itself. Since other “raising” type analyses exists, it seems that there is more independent support for an RO analysis than an ECM analysis.

9.3 Raising to Object in The Minimalist Program

The Minimalist Program dispenses with the projection principle (cf. Chomsky 1992), and therefore complements are not required to bear theta roles. However, Chomsky raises a new objection to the RO analysis: it violates a requirement that items only move to specifier (and adjunct) positions, positions external to the “targeted phrase marker”. However, given the syntactic configuration Chomsky assumes for multiple complements, it appears that some version of the RO analysis is actually available.

Chomsky assumes the structure in Figure 102 for verbs with multiple complements, except that I have taken the liberty of representing verb movement as structure sharing (Chomsky 1992, p. 19). Figure 103 is a similar analysis for believe type verbs. Although the “object” position turns out to be a specifier in these structures, this is still a raising to object analysis, since (apparently) Minimalist Program objects can be specifiers. Therefore Chomsky’s objection to the raising to object analysis does not hold.

Arguments for a raising to (surface) object analysis in the Minimalist Program are presented in Lasnik and Saito 1991. They note that the \textit{few} students is higher in the phrase structure tree in (8b) than (8a) in order to explain why \textit{few} can have either wide or narrow scope in (8a), but only wide scope in (8b). Similarly, the relative height of \textit{none} and \textit{any} are at issue in (8c,d,e) and whether the anaphor \textit{each other} is in the same clause as its antecedent is at issue in (8f,g,h)

(8) a. The FBI proved that few students were spies (wide/narrow)

b. The FBI proved few students to be spies (wide only)

c. The DA accused none of the defendants during any of the trials
John put the book on the shelf

**Figure 102** A Minimalist Binary Branching Analysis of Multiple Complements
Mary believes John to be a linguist

Figure 103: A Minimalist Raising-to-Object Structure
d. "The DA proved [none of the defendants to be guilty] during any of the trials

  e. "The DA proved [that none of the defendants were guilty] during any of the trials

  f. "The DA proved [the defendants to be guilty] during each other’s trials

  g. "The DA accused the defendants during each other’s trials

  h. "The DA proved [that the defendants were guilty] during each other’s trials

Lasnik and Saito show that the differences in structural height between the objects of the believe type verbs and the subjects of the tensed complements are S-structure differences rather than LF differences due to the interaction between wh extraction, quantifiers and binding theory. Following Chomsky 1981, they claim that (9) (their 42 and 43) shows that Principle C of Binding Theory must apply at S-structure, not D-structure or LF. Following Barss 1986 and Lasnik and Saito 1992, they claim that (10) (their 47 and 48) shows that anaphors must be bound at S-structure and not at LF. Lasnik and Saito also show that the contrast in (11) can only be due to S-structure facts since nothing rules out LF raising of the negative element Neg of unlikely which would allow (11b) to be well-formed since Neg would c-command any. Thus Lasnik and Saito conclude that the above quantifier scope and binding distinctions show that the raising to object at S-structure approach is preferred over the raising to AGR-O at LF approach.

(9) a. Which book that John; read did he; like (not at D-structure)

    b. *He; liked every book that John; read (not at LF)

(10) a. John; wonders which picture of himself; Mary showed to Susan

    b. *John; wonders who showed which picture of himself; to Susan;

(11) a. It is unlikely that anyone will address the rally

    b. *Anyone is unlikely to address the rally

Therefore, it appears that there is good reason to posit a raising to object analysis in the Minimalist Program.

9.4 Constraints on Conjoined Complements

Examples (12) from Dougherty and Leacock 1993 (their 15a, 16, 18) are problematic for an ECM analysis. The verb persuade is an object control verb and the verb expect is a believe type verb. Under an ECM analysis, these coordinated verbs subcategorize for structures which are not compatible with each other. Therefore it is difficult to assign a single constituent structure to the phrases following the pairs of coordinated
verbs. This leads Dougherty and Leacock to argue that there can be no S-structure for these sentences. In contrast, we show that an RO analysis explains this data.

(12)  a. John both persuaded and expected the doctor to visit Bill.
     b. Who did John persuade and expect the doctor to visit?
     c. The doctor was persuaded and expected to visit Bill.

If we assume some constraint like (13) (cf. Ross 1967, Williams 1978), then the RO analysis diagrammed as Figure 104 obeys this constraint. The only difference is that the two verbs are in different complement relations with the \( N \) following the verbs. *Expect* only assigns complement-case to this \( N \), whereas *persuade* assigns a theta role as well.

(13)  Coordinated heads must take the same set of complement constituents

It is not be possible to construct an ECM analysis which obeys (13) for examples like (12). Figure 104 models an ECM analysis if we assume that *COMP-CASE* is not a type of *COMPLEMENT* relation. Therefore, it would be assumed that *persuade* takes two complements
and expect takes one, in violation of (13). Given that (13) is valid, an RO analysis is possible, but an ECM analysis is not.

Dougherty and Leacock show that other properties of conjuncts, like control properties may conflict if natural language semantics allow, as in the following examples (Dougherty and Leacock 1993 29a and 57b):

(14) a. John could neither promise nor persuade himself to study harder.

b. John promised and expected himself to study harder

Under our analysis, promise, persuade, and expect all take \( \overline{N} \) and infinitival complements. They obey constraint (13). The difference between these verbs is the relation of subject of the infinitive to the matrix clause. In (14a), the subject of promise and the object of persuade are the controllers of the infinitival complement. Notice however that the sentences in (14) are problematic for a GBUG structure sharing account of control. It does not appear to matter in these cases whether the anaphor coindexed with the subject or the subject itself is the controller of the infinitive, as might be expected if coordination was sensitive to whether the control/raising properties of coordinated verbs were the same. An HPSG-style structure sharing account works here however, since the indices of the anaphors, the subjects and the infinitives will be shared in these cases.

In summary, Dougherty and Leacock’s coordination examples provide evidence both for an RO analysis of believe type verbs and for an HPSG-style account of structure sharing in obligatory control constructions.

9.5 Summary

The assumption that all licensing relations are types of constituent relations has the implication that the raising to object analysis of believe type verbs is compatible with GBUG analyses, but other popular P & P analyses are not. Evidence from a variety of sources is presented above which suggests that the raising to object analysis is the correct analysis for these verbs.
Concluding Remarks

The leading idea of this book is that a phrase structure tree model is inadequate for P & P theories and that the GBUG model is more adequate. Many popular linguistic frameworks also use FS models, and partially for this reason, FS formalization is itself an active area of research. GBUG makes it easier to incorporate ideas into P & P from these other FS-based theories since theories using similar models are easier to compare than theories that do not. The GBUG model has also benefited from a great deal of research into the mathematical properties of FSs, DAGs and FS logics.

I formulated a new P & P framework, borrowing from a wide variety of previous linguistic frameworks. Unlike many previous formalizations of P & P theories (e.g., Correa 1987, Stabler 1993), I changed the theory I was formalizing. My GBUG-based theory is based on licensing relations and \( \overline{X} \) constraints, whereas previous P & P theories are based on levels of representation and conventional \( \overline{X} \) theory.
Bibliography


Fioretta, Josef. 1993. NYU Seminar Presentation on Unification and Morphological Agreement.


Bibliography / 235


Joshi, Aravind K., and K. Vijay-Shanker, 1989, Treatment of Long Distance dependencies in LFG and TAG; Functional Uncertainty in LFG is a Corollary in TAG. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics.*


Keenan, Edward L. 1974. The Functional Principle: Generalising the Notion of 'Subject'. In *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society.*


from the Twenty First Regional Meeting of the Chicago Linguistic Society.
University Press.
Ravin, Yael. 1976, A Decompositional Approach to Predicates Denoting
MIT.
Riemsdijk, H. Van. 1978, A Case Study in Syntactic Markedness: The Binding
Ross, Háj. 1995, Defective Noun Phrases. In Papers from the 31st Regional
Meeting of the Chicago Linguistic Society.
Ross, J. R. 1967, Constraints on Variables in Syntax. Doctoral dissertation,
MIT.
Ross, J. R. 1974, There, There (There, (There, (There, ...))). In Papers from
the Tenth Regional Meeting of the Chicago Linguistic Society.
In Views on Phrase Structure, ed. K. Leffel and D. Bouchard. Dordrecht:
Kluwer.
Rouveret, Alain, and Jean-Roger Vergnaud. 1980. Specifying Reference to
manuscript.
Sag, Ivan A., and Janet D. Fodor. 1994, Extraction Without Traces. In Pro-
cedings of the Thirteenth West Coast Conference on Formal Linguistics,
ed. Raul Aronovich, William Byrne, Susanne Preuss, and Martha Senturia.
Stanford University. CSLI Publications/SLA.
Grammar of English and its Applications. Reading, Mass.: Addison-
Wesley Publishing Company, Inc.
Sanders, Gerald. 1967, Some General Grammatical Processes in English. Doc-
toral dissertation, Indiana University.
Sanders, Gerald. 1974, Precedence Relations in Language. Foundations of
Language 11:361–400.
Shieber, Stuart, and Yves Schabes. 1990, Synchronous Tree-Adjoining Gram-
on Computational Linguistics.
Shieber, Stuart M. 1986, An Introduction to Unification-based Approaches to


Stowell, Tim, 1978. What was There Before There was There. In *Papers from the Fourteenth Regional Meeting of the Chicago Linguistics Society*.


