A Unification-Based Approach to Government and Binding Theory

by

Adam Meyers

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Linguistics Department

New York University

April 21, 1994

______________________

Approved
Acknowledgments

I thank the NYU Linguistics Faculty and all members of my committee. In particular, Ray Dougherty has directed my research, helped me find a proper mix of theoretical and computational linguistics, edited various drafts of this dissertation beyond the call of duty, and alerted me to employment prospects. David Johnson was the other half of a thoroughly enjoyable collaboration at IBM. He also steered me towards an interdenominational approach to linguistics and gave me important criticism of this dissertation. Mark Baltin directed me in a fascinating research assistantship on the nature of selection. Conversations I’ve had with him over the past few years have had a significant influence on this dissertation. Paul Postal has provided guidance on various linguistic, philosophical and political topics by email, in person and through his published work. John Singler has always provided me with thorough and intelligent criticism of anything I’ve ever given him to read. John Costello was the one who talked me into enrolling in the department. Noriko Umeda was a major reason I got through the comprehensive exams.

I am grateful to: Profs. Rob French, Rob Jeffers, Robert Fowkes, Steve Seegmiller and Martin Braine; Lois Bloom, Richard Beckwith and Erin Tinker and others from the Teacher’s College Child Language Project; Warren Plath, Claudia Leacock, Fred Damerau, Ezra Black and Don Decain of IBM; Ralph Grishman, Catherine Macleod, Kanna Rajan, Evan Christenson, Leslie Barrett, Beth Craig, Isaac Glasman, Cassandra Glasco, John Sterling and Tomek Strzalkowski of NYU’s Proteus Group; NYU students including: Angel Arzan, Scott Browne, Peter Jeong, Lynn Ohira, Mark Brand, Josef Fiorella, Zupeczana Vizić, Svetlana Vasina and Roman Yangarber; Yolando and Anndu of the ACF; and close friends and family: Mom, Dad, Beth, Rosie, Happy, Nana, Mara, Marvin, Will, Muriel, Greg, Aviva, Casey, Molly, Neal, David, Fernando, Judy, Francesca, Joel, Jed, Eric, Ed. I thank the NYU Linguistics Department, Teachers College, the IBM T.J. Watson Research Center and the Courant Institute of the Mathematical Sciences for the intellectual and financial support necessary for me to complete this dissertation.

Most of all, I thank Jenny, the love of my life, who pampered me through the process of getting my degree and without whom I could never have made it this far.
# Table of Contents

Acknowledgments ........................................................................................................................ iv  
List of Abbreviations and Acronyms ........................................................................................... x

## 1. Introduction .............................................................................................................................. 1
   1.1 Problem Statement and Approach ........................................................................... 1
   1.2 A GBUG-based Model of GB Theory ..................................................................... 4
   1.3 Summary of Dissertation ......................................................................................... 7

Notes .............................................................................................................................. 11

## 2. A Unification-based Model for Government and Binding Theory .......................................... 13
   2.1 Introduction .............................................................................................................. 13
   2.2 Feature Structures .................................................................................................... 13
   2.3 Directed Acyclic Graphs ......................................................................................... 25
   2.4 Subsumption ............................................................................................................ 27
      2.4.1 Introduction ............................................................................................ 27
      2.4.2 Defining Subsumption ............................................................................ 28
      2.4.3 The Role of Subsumption in our Grammar and Lexicon ....................... 34
      2.4.4 Summary ................................................................................................ 39
   2.5 Unification ............................................................................................................... 40
      2.5.1 Introduction ............................................................................................ 40
      2.5.2 Defining Unification ............................................................................... 40
      2.5.3 The Role of Unification in Grammar, the Lexicon and Parsing .......... 46
      2.5.4 Summary ................................................................................................ 58
   2.6 Properties of Label Subsumption and Unification ................................................... 59
      2.6.1 Introduction ............................................................................................ 59
      2.6.2 Node Label Subsumption and Unification .................................................. 59
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6.3</td>
<td>Arc Label Subsumption and Unification</td>
<td>60</td>
</tr>
<tr>
<td>2.6.4</td>
<td>Conclusion</td>
<td>63</td>
</tr>
<tr>
<td>2.7</td>
<td>Summary</td>
<td>63</td>
</tr>
<tr>
<td>Appendix 2A</td>
<td>A GBUG Unification Algorithm</td>
<td>64</td>
</tr>
<tr>
<td>Notes</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>78</td>
</tr>
<tr>
<td>3.2</td>
<td>Constituent Licensing and X-bar Theory</td>
<td>80</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Introduction</td>
<td>80</td>
</tr>
<tr>
<td>3.2.2</td>
<td>A Licensing Relation Based Version of X-bar Theory</td>
<td>81</td>
</tr>
<tr>
<td>3.2.3</td>
<td>The Previous GB Version of X-bar Theory</td>
<td>102</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Summary</td>
<td>107</td>
</tr>
<tr>
<td>3.3</td>
<td>Predicate Licensing</td>
<td>108</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Introduction</td>
<td>108</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Selection Restrictions</td>
<td>110</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Theta Relations</td>
<td>113</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Modification</td>
<td>118</td>
</tr>
<tr>
<td>3.3.5</td>
<td>A Comparison of Theta and Modification Relations</td>
<td>119</td>
</tr>
<tr>
<td>3.3.6</td>
<td>Degree</td>
<td>130</td>
</tr>
<tr>
<td>3.3.7</td>
<td>Quantification</td>
<td>131</td>
</tr>
<tr>
<td>3.3.8</td>
<td>Possessives</td>
<td>135</td>
</tr>
<tr>
<td>3.3.9</td>
<td>Conclusion</td>
<td>141</td>
</tr>
<tr>
<td>3.4</td>
<td>Agreement Licensing Properties</td>
<td>142</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Introduction</td>
<td>142</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Case as an Agreement Relation</td>
<td>145</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>5.4 Conclusion</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td>261</td>
<td></td>
</tr>
<tr>
<td>6. What do Empty Categories Represent?</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td>6.1 Introduction</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td>6.2 The Role of Empty Categories in GB Theory</td>
<td>266</td>
<td></td>
</tr>
<tr>
<td>6.3 The Constituent Sharing Analysis</td>
<td>269</td>
<td></td>
</tr>
<tr>
<td>6.4 Selection Restrictions</td>
<td>281</td>
<td></td>
</tr>
<tr>
<td>6.5 ECs and Binding Theory</td>
<td>283</td>
<td></td>
</tr>
<tr>
<td>6.6 Null Phrase Structure Rules and Nonterminating Derivations</td>
<td>288</td>
<td></td>
</tr>
<tr>
<td>6.7 Quantifiers in Japanese</td>
<td>291</td>
<td></td>
</tr>
<tr>
<td>6.8 Wanna Contraction and the Status of Empty Categories</td>
<td>296</td>
<td></td>
</tr>
<tr>
<td>6.8.1 Introduction</td>
<td>296</td>
<td></td>
</tr>
<tr>
<td>6.8.2 The Contraction Debate</td>
<td>298</td>
<td></td>
</tr>
<tr>
<td>6.8.3 The To Subsumption Analysis</td>
<td>305</td>
<td></td>
</tr>
<tr>
<td>6.8.4 Wanna Contraction and Case Marking</td>
<td>311</td>
<td></td>
</tr>
<tr>
<td>6.8.5 Descriptive Adequacy</td>
<td>313</td>
<td></td>
</tr>
<tr>
<td>6.8.6 Summary</td>
<td>315</td>
<td></td>
</tr>
<tr>
<td>6.9 Conclusion</td>
<td>316</td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td>318</td>
<td></td>
</tr>
<tr>
<td>7. What is Constituency?</td>
<td>325</td>
<td></td>
</tr>
<tr>
<td>7.1 Introduction</td>
<td>325</td>
<td></td>
</tr>
<tr>
<td>7.2 Surface and Non-Surface Constituent Relations</td>
<td>331</td>
<td></td>
</tr>
<tr>
<td>7.3 Believe Type Verbs</td>
<td>336</td>
<td></td>
</tr>
<tr>
<td>7.3.1 Introduction</td>
<td>336</td>
<td></td>
</tr>
<tr>
<td>7.3.2 Expletives in Complement Positions</td>
<td>340</td>
<td></td>
</tr>
</tbody>
</table>
7.3.3 Sentences with No S-structure ............................................................... 354
7.3.4 W-Verbs and Raising to Object ............................................................. 364
7.3.5 Conclusion .............................................................................................. 370
7.4 The Minimalist Framework .............................................................................. 371
7.4.1 Introduction ............................................................................................ 371
7.4.2 A GBUG-based Approach to the Minimalist Framework ..................... 372
7.4.3 Why S-structure Positions are Required in GB Analyses ..................... 376
7.4.4 Believe-type Verbs in the Minimalist Framework ................................. 381
7.4.5 Summary ................................................................................................ 385
7.5 Conclusion ............................................................................................................... 385
Notes .............................................................................................................................. 387
8. Conclusions and Future Research ............................................................................................ 396
8.1 Conclusions .............................................................................................................. 396
8.2 A GBUG Representation of Quantifier Scope and WH Extraction .......... 399
8.3 Verbal Inflection: Syntactic Category or Lexical Feature ......................... 402
8.4 A Parsing Algorithm ......................................................................................... 404
Notes .............................................................................................................................. 406
References .............................................................................................................................. 407
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-Chain</td>
<td>Argument Chain</td>
</tr>
<tr>
<td>~-Chain</td>
<td>Non-argument Chain</td>
</tr>
<tr>
<td>Acc</td>
<td>Accusative</td>
</tr>
<tr>
<td>ACL</td>
<td>Association for Computational Linguistics</td>
</tr>
<tr>
<td>Adj</td>
<td>Adjective</td>
</tr>
<tr>
<td>Adv</td>
<td>Adverb</td>
</tr>
<tr>
<td>AG</td>
<td>Attribute Grammar</td>
</tr>
<tr>
<td>AGR</td>
<td>Agreement Category</td>
</tr>
<tr>
<td>APG</td>
<td>Arc Pair Grammar</td>
</tr>
<tr>
<td>CFR</td>
<td>Context free phrase structure rule</td>
</tr>
<tr>
<td>CG</td>
<td>Categorial Grammar</td>
</tr>
<tr>
<td>CMT</td>
<td>Case marked trace</td>
</tr>
<tr>
<td>CS</td>
<td>Constituent Sharing</td>
</tr>
<tr>
<td>CSLI</td>
<td>Center for the Study of Language and Information</td>
</tr>
<tr>
<td>DAG</td>
<td>Edge-labeled single-rooted directed acyclic graph</td>
</tr>
<tr>
<td>Dat</td>
<td>Dative</td>
</tr>
<tr>
<td>Def</td>
<td>Definite</td>
</tr>
<tr>
<td>EC</td>
<td>Empty Category</td>
</tr>
<tr>
<td>ECM</td>
<td>Exceptional Case Marking</td>
</tr>
<tr>
<td>ECP</td>
<td>Empty Category Principle</td>
</tr>
<tr>
<td>ESCOL</td>
<td>Eastern States Conference on Linguistics</td>
</tr>
<tr>
<td>EST</td>
<td>Extended Standard Theory</td>
</tr>
<tr>
<td>Fem</td>
<td>Feminine</td>
</tr>
<tr>
<td>FPF</td>
<td>Finite Partial Function</td>
</tr>
<tr>
<td>FS</td>
<td>Feature Structure</td>
</tr>
<tr>
<td>FSEP</td>
<td>Feature Structure Economy Principle</td>
</tr>
<tr>
<td>GB</td>
<td>Government and Binding</td>
</tr>
<tr>
<td>GBUG</td>
<td>Government and Binding Unification Grammar</td>
</tr>
<tr>
<td>Gen</td>
<td>Genitive</td>
</tr>
<tr>
<td>GL</td>
<td>GB Logic</td>
</tr>
<tr>
<td>GPSG</td>
<td>Generalized Phrase Structure Grammar</td>
</tr>
<tr>
<td>Head-Proj</td>
<td>Head Projection</td>
</tr>
<tr>
<td>HFC</td>
<td>Head Feature Convention</td>
</tr>
<tr>
<td>HPSG</td>
<td>Head-Driven Phrase Structure Grammar</td>
</tr>
<tr>
<td>HS</td>
<td>Have Subsumption</td>
</tr>
<tr>
<td>I</td>
<td>Inflection</td>
</tr>
<tr>
<td>INFL</td>
<td>Inflection</td>
</tr>
<tr>
<td>IP</td>
<td>Inflection Phrase</td>
</tr>
<tr>
<td>Indef</td>
<td>Indefinite</td>
</tr>
<tr>
<td>LF</td>
<td>Logical Form</td>
</tr>
<tr>
<td>LFG</td>
<td>Lexical Function Grammar</td>
</tr>
<tr>
<td>LHS</td>
<td>Left Hand Side</td>
</tr>
<tr>
<td>Loc</td>
<td>Locative</td>
</tr>
<tr>
<td>LP</td>
<td>Linear Precedence</td>
</tr>
<tr>
<td>LRP</td>
<td>Licensing Redundancy Principle</td>
</tr>
<tr>
<td>LSP</td>
<td>Linguistic String Project</td>
</tr>
<tr>
<td>MAC</td>
<td>Minimal A-chain</td>
</tr>
<tr>
<td>Masc</td>
<td>Masculine</td>
</tr>
<tr>
<td>MGG</td>
<td>Metagraph Grammar</td>
</tr>
<tr>
<td>Morph-Agr</td>
<td>Morphological Agreement</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Morph-Case</td>
<td>Morphological Case</td>
</tr>
<tr>
<td>MSRI</td>
<td>Mathematical Sciences Research Institute</td>
</tr>
<tr>
<td>N</td>
<td>Noun</td>
</tr>
<tr>
<td>Neg</td>
<td>Negative</td>
</tr>
<tr>
<td>NELS</td>
<td>North Eastern Linguistic Society</td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>NOC-PRO</td>
<td>Nonobligatorily Controlled Pro</td>
</tr>
<tr>
<td>Nom</td>
<td>Nominative</td>
</tr>
<tr>
<td>NQ</td>
<td>Nominal Quantifier</td>
</tr>
<tr>
<td>Obj</td>
<td>Object</td>
</tr>
<tr>
<td>OC-PRO</td>
<td>Obligatorily Controlled Pro</td>
</tr>
<tr>
<td>P</td>
<td>Preposition</td>
</tr>
<tr>
<td>Pass</td>
<td>Passive</td>
</tr>
<tr>
<td>PATR</td>
<td>Parse and Translate</td>
</tr>
<tr>
<td>PCH</td>
<td>Priority of Constituency Hypothesis</td>
</tr>
<tr>
<td>PF</td>
<td>Phonetic Form</td>
</tr>
<tr>
<td>Plur</td>
<td>Plural</td>
</tr>
<tr>
<td>PP</td>
<td>Postal and Pullum</td>
</tr>
<tr>
<td>Pred</td>
<td>Predicate</td>
</tr>
<tr>
<td>PS92</td>
<td>Pollard and Sag (1992)</td>
</tr>
<tr>
<td>Quant-Value</td>
<td>Quantification-Value</td>
</tr>
<tr>
<td>RG</td>
<td>Relational Grammar</td>
</tr>
<tr>
<td>RHS</td>
<td>Right Hand Side</td>
</tr>
<tr>
<td>RNR</td>
<td>Right Node Raising</td>
</tr>
<tr>
<td>RO</td>
<td>Raising to Object</td>
</tr>
<tr>
<td>SFG</td>
<td>Stratified Feature Grammar</td>
</tr>
<tr>
<td>SOV</td>
<td>Subject Verb Object</td>
</tr>
<tr>
<td>SS</td>
<td>Subject Sharing</td>
</tr>
<tr>
<td>Subj</td>
<td>Subject</td>
</tr>
<tr>
<td>Sgraph</td>
<td>Stratified Feature Graph</td>
</tr>
<tr>
<td>T</td>
<td>Tense</td>
</tr>
<tr>
<td>TAG</td>
<td>Tree Adjoining Grammar</td>
</tr>
<tr>
<td>TG</td>
<td>Transformational Grammar</td>
</tr>
<tr>
<td>TS</td>
<td>To Subsumption</td>
</tr>
<tr>
<td>V</td>
<td>Verb</td>
</tr>
<tr>
<td>VOS</td>
<td>Verb Object Subject</td>
</tr>
<tr>
<td>VSO</td>
<td>Verb Subject Object</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

1.1 Problem Statement and Approach

This dissertation formulates Government and Binding Unification Grammar (GBUG), a Feature Structure (FS) based model for Government and Binding (GB) Theory. In GBUG, fundamental GB relations including constituent relations, predicate relations and agreement relations are cast in terms of what we call LICENSING RELATIONS. Each licensing relation is modeled in an edge-labeled single-rooted directed acyclic graph D, as a pair of same source arcs: one licensor arc A and one licensee arc B, where D denotes a FS representation of some lexical entry, phrase or syntactic construction; the subgraph rooted at A represents the licensor of the relation, e.g., the case assigner, the theta assigner, the head, etc.; and the subgraph rooted at B represents the licensee of the relation, e.g., the theta recipient, the case recipient, the specifier, the complement, the adjunct, etc.

Licensing relations provide the basis for (1) the development of a theory of word order; (2) the elimination of most empty categories in GB analyses; (3) the elimination of constituent relations which have the same interpretation as other licensing relations e.g., we show that the meaning of the relation Internal-Theta(X,Y) always includes the meaning of some relation Complement(X,Y) and therefore any pair of relations Internal-Theta(X,Y) and Complement(X,Y) always reduces to the single relation Internal-Theta(X,Y) (cf. Chomsky’s 1991 principle of economy of representation); and (4) a comparison of GB analyses with analyses in unification-based versions of other linguistic frameworks such as Relational Grammar (including Arc-Pair Grammar and Stratified Feature Grammar), Lexical Function Grammar, Categorial Grammar, Generalized Phrase Structure Grammar, Head-Driven Phrase Structure Grammar, Functional Unification Grammar, Parse and Translate II (PATR-II), among others.

This dissertation shows that the descriptive and explanatory power of GB theory derives from the interaction between licensing relations and that the feature structure is an excellent data structure for representing these crucial relations. We show that licensing relations cannot be expressed in a
constituent structure tree, which was the data structure used in previous GB analyses. Not only are theta and case relations are not modeled in constituent structure trees at all, but GB’s constituent relations are represented inadequately.

GBUG is based on other FS based and graph-based formalisms, especially the ones used for Head-Driven Phrase Structure Grammar, Stratified Feature Grammar and Arc-Pair Grammar. We chose a unification-based formalization because: (1) the mathematical properties of unification and FS formalization are well understood; and (2) many other linguistic theories are modeled in unification-based frameworks—therefore, adopting a unification-based model facilitates comparison of GB analyses with analyses in these other frameworks.

We propose and develop the Feature Structure Economy Principle (FSEP), a principle of economy of representation in the sense of Chomsky (1991). Under the FSEP, given two Feature Structures (FSs) $FS_1$ and $FS_2$, representing the same licensing relations, $FS_1$ is preferred if $FS_1$ represents these licensing relations using fewer features (arcs) than $FS_2$.

The semantics\(^1\) of licensing relations and the FSEP motivates the replacement of most empty category analyses with an analysis in which constituent licensor/licensee arcs share the same target. We show that every empty category (ec) with a single, unambiguously assigned antecedent (NP-trace, WH-trace, Obligatorily Controlled PRO, but not nonobligatorily controlled PRO) superficially encodes that the same constituent is a licensor or licensee of more than one constituent relation. For example, if N and NP-trace are coreferential and the relations Specifier(X,N) and Complement(Y,NP-trace) hold, we can eliminate all arcs representing coreference if we replace the relation Complement(Y,NP-trace) with Complement(Y,N). In the edge-labeled single-rooted directed acyclic graph (DAG) representation of these relations, the two arcs representing the Complement Licensee features (Complement arcs) share the same target—the root node of the DAG representing N. Key facts about binding theory and selection restrictions follow from a structure sharing approach which must be stipulated for ec analyses. Therefore, FSEP and other principles clearly support a structure sharing analysis over an ec analysis for a wide range of constructions.
The FSEP also motivates the elimination of redundant portions of GB analyses based on our observation that the relations \( \text{theta} \) and \( \text{case} \) are subtypes of the relations \( \text{specifier} \) and \( \text{complement} \). For example, Internal-Theta(X,Y) alone conveys all the information that the two relations Internal-Theta(X,Y) and Complement(X,Y) convey in combination. Therefore Complement(X,Y) is extraneous. So-called exceptional case marking (ECM) analyses (cf. Chomsky 1980b, 1981, among others) are potential counter-examples to this claim. However, we provide evidence for choosing a raising to object analysis (cf. Rosenbaum 1967, Postal 1974, Postal and Pullum 1988, Lasnik and Saito 1991, among others) for constructions for which an ECM analysis was previously assumed.

These simplifications do not follow in the constituent structure tree model because the relevant relations are not represented. An explicit GBUG-based model of GB theory is crucial to each of these results. We first determine precisely what empty categories, case relations, theta relations, specifier relations and complement relations represent. Then it becomes clear how we can represent the same information in a more efficient manner, eliminating redundant portions of the representation.

A major strong point of GBUG is its similarity to formal models for Generalized Phrase Structure Grammar (GPSG), Head-driven Phrase Structure Grammar (HPSG) Lexical Function Grammar (LFG), and Relational Grammar (RG) (including Arc-Pair Grammar and Stratified Feature Grammar) analyses. GBUG provides a medium for comparing GB analyses with analyses in any linguistic framework formalized in a unification-based or Feature Structure model. Comparing competing analyses of some syntactic phenomenon in two different frameworks is difficult because different frameworks have different basic assumptions. For example, RG (among others) assumes that grammatical relations like subject, direct object, etc. are primitive and bases principles of the theory on these relations. GB’s basic relations include case, theta, specifier, complement, adjunct, and others. However, if two different frameworks are represented using similar formal systems, e.g., two unification-based formalisms, it becomes easier to see which aspects of each analysis translates into the other framework, which aspects do not and what are the crucial differences.

A long range goal of this dissertation is to provide a principled way to develop a descriptively and explanatorily adequate computational model based on GB Theory. The development of an explicit
data structure (GBUG) and (the possibility for) a descriptively adequate grammar are essential. In the
course of this dissertation we attempt to fill gaps of GB theory while maintaining a GB-style analysis.
Along these lines we provide the basis for more complete GB analyses of modification and degree
constructions, word order, among other topics.

1.2 A GBUG-based Model of GB Theory

We begin with the basic assumption that the optimal logical language for writing a GB grammar
consists of statements of different kinds of logic—we call this logic GB Logic (GL). The maximally
simple grammar written in GL expresses each principle of grammar in the lowest possible order of
logic. GBUG is a subset of GL which only expresses linguistic principles that may be described in
feature structure logic—in particular we assume a variety of feature structure logic based on Kasper
and Rounds (1986, 1990) and Johnson and Moss (1993, 1994). This dissertation formalizes a subset of
GB theory in GBUG and GL. Other principles are stated more informally using a combination of
GBUG, higher order logic (e.g., quantifiers) and plain English. We cite current work in formalizing
GB principles as statements of logic, e.g., Stabler (1993), as a possible source for a more formal
account of these more complex statements.

Previous approaches to formalizing GB theory typically use a formal language capable of
encoding all GB concepts relevant to their work. For example, Correa (1987, 1988, 1991, 1992a,b)
proposes an attribute grammar formalism for GB Theory. In principle, Correa’s approach could for-
alize all the same concepts formalized here, as well as many of the GB principles that are too com-
plex for FS representation. Additionally, attribute grammars are well understood mathematically.
Basically the same things may be said about the logic-based approaches of Fong (1990, 1991a, 1991b,
1992), Stabler (1993), among others. We prefer the unification-based approach for two reasons: (1)
We feel it is important to represent grammatical principles in the simplest logic available and the GL
approach outlined above ensures this; and (2) Modeling GB in GBUG invites comparison with these
other unification-based frameworks. This is important because a large number of linguistic frameworks
are modeled in unification-based grammars.
In GBUG, a lexical entry $L$ of a word $W$ is an ordered pair $[F_L, O_L]$. $F_L$ is a FS and $O_L$ is a set of word order (or linear precedence) constraints imposing a partial order on arcs in $F_L$. $W$ may have one or more lexical entries, each representing a phrase which $W$ ANCHORS in the sense of SFG.\(^2\) $W$ anchors a phrase if $W$ justifies the existence of that phrase. We show that it is an empirical issue whether $W$ is the head of the phrase it anchors or some other nonhead constituent. In this respect, we follow Johnson and Moss (1993, 1994) and Johnson, Meyers and Moss (1993a,b).

Chomsky (1970) and much subsequent work based on $\lambda$ theory assumes that a phrase is always projected from the head of that phrase. PROJECTION of a phrase from a lexical item is the process by which a lexical item is associated with a corresponding phrase structure tree licensed by that item. Both projection and anchoring associate a lexical item with a data structure containing information licensed by that item. Chomsky assumes that $\lambda$ theory, a module of grammar, projects or builds phrase structure from the lexicon as part of the process of building D-structure from the lexicon. In our version of GB theory, $\lambda$ theory guarantees that each $F_L$ has a certain form, for every $L = [F_L, O_L]$, a lexical entry for any $W$. However, $W$ may or may not be the head of $F_L$. For example, we show that an adjective modifier anchors a nonterminal of category noun, even though the adjective is an adjunct, not the head of the phrase. A principle of $\lambda$ theory, namely a version of GPSG’s head feature principle, insures that a phrase has the same category as its head.

We show that many GB principles thought to be clause-level or phrase-level principles are most economically viewed as parts of lexical entries or constraints on the construction of lexical entries.\(^3\) This kind of move is typical of unification-based frameworks (linguistic frameworks using unification-based formalisms) because this allows unification, an operation for combining FSs to represent many of the combining operations in that theory. For example, we use unification to implement the instantiation of a licensing relation, i.e., what GB literature calls the assignment of case, the assignment of a theta role or lexical insertion in the case of complement, specifier and adjunct relations.
F_L is a FS representation representing W’s morphological and phonological and properties, as well as all licensor and licensee features which W anchors, where a licensing relation R(x,y) is modeled by each pair of (1) one licensor feature X with value x; and (2) one licensee feature Y with value y. X and Y are modeled in a DAG as same source arcs A_X and A_Y. For example, a relation Internal-Theta(Verb, N) is represented by a feature Head-Proj with a value Verb and a feature Internal-Theta with a value N, where Head-Proj is an Internal-Theta Licensor feature and Internal-Theta is an Internal-Theta Licensee feature, and Verb and N are FSs representing a verb and an N. For each licensing relation R, such that W or a phrase anchored by W is the licensor, F_L includes a FS representing the licensee of R.

O_L is a partial order on the surface constituent licensor/licensee features represented in F_L. We follow SFG in representing word order in this way. In a graph-based representation of FSs, F_L is instantiated as a DAG and each licensor and licensee feature is interpreted as an arc. O_L is a partial order on the subset of the arcs in F_L that represent surface constituent positions, i.e., arcs representing S-structure relations labeled with the features Head-proj (head), Complement, Specifier or Adjunct. O_L is a partial order, because some surface constituent features may be unordered. For example English adverbs may occur to the left or right of the verb phrases they modify (John quickly [left the room]; John [left the room] quickly.) O_L is derived from word order parameters settings and adjacency constraints, e.g., case adjacency.

This dissertation provides a program for creating a more fully specified and more falsifiable version of GB theory. Under this program, it will be possible to compare GB analyses with analyses in other frameworks, to create more descriptively adequate GB grammars and to build computational models of GB theory.
1.3 Summary of Dissertation

This section outlines the remaining chapters in this dissertation.

Chapter 2 provides definitions of basic concepts of unification-based formalisms from the literature. The terms Feature Structure (FS), Subsumption, and Unification are defined in terms of a model in which FSs are described alternately as DAGs or finite partial functions. We provide FS representations of lexical items, phrases and grammatical constraints and definitions of constraint satisfaction and theta role/case assignment in terms of subsumption and unification; we discuss the role of unification and subsumption in the structure of the lexicon and parsing; and we show that the unification-based approach provides a concrete criterion for formulating a principle of economy of representation (cf. Chomsky 1991)—the Feature Structure Economy Principle (FSEP.)

Chapter 3 describes the syntactic properties of licensing relations and explicates their role in GB theory. Licensing relations include:

- Constituent relations, such as: complement, specifier, adjunct. These relations are represented in terms of the feature Head-Proj—a constituent licensor feature, and a constituent licensee feature named for each relation. We show that previous attempts to define these relations in terms of phrase structure trees were inadequate for two reasons: (1) In GB literature, the terms complement, specifier and adjunct are given definitions which fail to match many standard GB analyses; and (2) If these definitions are expanded, the resulting definitions overlap so it is impossible to unambiguously determine whether elements in constituent structure trees are complements, adjuncts or specifiers. Our licensing relation definitions solve this problem.

- Predicate relations, such as: theta, modification, degree and quantification. We define these relations in terms of selection restrictions which the licensor imposes on the licensee.

- Agreement licensing relations, which must be explicitly distinguished from morphological agreement relations. We define agreement licensing relations, as sets of conditions which must be satisfied for predicate relations to be well-formed—these sometimes include conditions based on morphological agreement. We limit our discussion to N-AGREEMENT, conditions on various N-internal predicate relations and ABSTRACT CASE, a set of conditions on theta relations which have
a long history in GB theory.

Chapter 3 also provides definitions of command relations and chains, GB theoretical objects which are defined in terms of licensing relations.

Chapter 4 develops a theory of word order. Word order constraints in GBUG are partial orders on surface constituent licensor/licensee features, where a surface constituent licensor/licensee feature is a constituent licensor/licensee feature which represents an S-structure position. We show that the predicate licensor (the value of the predicate licensor feature) of a nonterminal constituent determines word order for that constituent. For example, most modifiers occur to the right of their modifiees in English, e.g., the president [of the country]; but adjectives occur to the left of their modifiees, e.g., the blond man; and a small number of adjectives occur to the right of their modifier, e.g., the president elect. We provide a systematic pattern of exceptions showing that exceptional word order is determined by the predicate licensor of a nonterminal constituent. We represent both canonical and exceptional word order in a principled and systematic way using a default inheritance hierarchy (cf. Wilensky 1981, Watanabe and Johnson 1990, Johnson and Watanabe 1991). Word order parameters are set for types of selection, syntactic categories and individual lexical items. A parameter setting is assigned to an item in the hierarchy based on either (a) information associated with that element; or (b) information associated with its superclass. (b) is only checked if (a) is not provided. Thus both canonical and exceptional word order can be described. We show that our account is more descriptively and explanatorily adequate than previous GB accounts of word order based solely on the Head parameter and Case theory. This approach is based on Johnson, Meyers and Moss (1993a,b).

Chapter 5 discusses how GBUG represents coreference for the Binding Theory and Trace Theory modules of GB theory. We compare our mode of representation to that of Pollard and Sag (1992.) We show that our mode of representation is superior for two reasons: (1) We provide a means of representing intersecting reference, e.g., the reading of (1.1) in which they refers to a group of people which includes (inclusively or exclusively) John and Mary. In contrast, Pollard and Sag only allow coreference; and (2) Pollard and Sag’s representation of coreference requires that coreferents agree in number, gender and person. Since neither John nor Mary are plural, their number features would each
clash with the number of the pronoun *they* in (1.1). Thus Pollard and Sag’s approach fails to correctly represent examples like (1.1). In contrast, we assume that agreement is accomplished by a more complex mechanism.

(1.1) John told Mary that they would leave together

Chapter 6 shows that empty categories (ecs) in GB represent features of lexical items, not phonologically null phrases. We show that once this role of ecs becomes explicit, most ecs may be eliminated in favor of more compact representation. In our theory, a chain consisting of one antecedent phrase P and one or more coreferential empty categories may be replaced by a set of arcs with target P, each representing a distinct constituent licensor/licensee feature.

Chapter 7 shows that predicate and agreement relations in GB theory are actually types of constituent relations. On this basis, we purge our syntactic representation of superfluous information. For example, based on the assumption that the relation *Internal-Theta* is actually a type of Complement relation, an explicit representation of Complement(X,Y) is unnecessary in a FS representing the relation Internal-Theta(X,Y). Our assumption that agreement relations, which include, case, are a type of constituent relations has implications for our treatment of *believe*-type verbs. In this regard we offers evidence for preferring a raising to object type analysis over an exceptional case marking analysis of these verbs.

Chapter 8 provides our conclusions. We discuss the successes and limitations of the GBUG formalism, as well as some of our theoretical findings. The main contribution of this dissertation is an explicit syntax and semantics for licensing relations and the consequences of this formulation. GBUG is a superior model to the constituent-structure-tree-based model previously assumed because GBUG clarifies GB analyses. The constituent-tree-based model is misleading because it does not represent GB’s fundamental relations. The most important aspects of GB analyses are usually discussed informally because they cannot be represented in constituent structure trees. We provide a solution to this problem. Chapter 8 also suggests some avenues for future research including: (1) a proposal for representing properties of Logical Form including quantifier scope in terms of licensing relations; (2) a proposed comparison of two alternative analyses of finite verbal inflection (INFL): (a) INFL represents
features of verbs, and (b) INFL is the head of a phrase and $V$ is its complement; and (3) a parsing algorithm for GBUG representations adapted from the algorithm presented in Johnson, Meyers and Moss (1993a,b) for SFG.
Notes

1 We use the term *semantics* throughout this dissertation to identify the interpretation given to a theoretical object. Only explicit references to the semantics of a natural language should be interpreted as such, e.g., *natural language semantics, the semantics of English, the semantics of Russian*, etc.

2 This idea of FSs and properties thereof being anchored by lexical items is similar to the trees associated with lexical items in the TAG framework (Schabes, Abeille and Joshi 1988, Joshi and Schabes 1991).

3 Abney (1986) uses the term *Licensing* to generalize selection on the model of theta roles. In contrast, we use the term *Licensing* to refer to selectional, case as well as constituent or $\bar{X}$ relations. Furthermore, the details of our theory of selection differs from that of Abney.

4 Like Arc-Pair Grammar (APG), SFG and LFG, GB analyses formulated in GBUG represent all relations previously described in the GB literature in terms of syntactic movement and control relationships. However, like APG/SFG and unlike LFG, a GBUG lexical entry represents the correspondence between canonical and derived entries. For example, a GBUG and SFG passive verb entry includes a representation of the passive verb’s relationship with its active form. In contrast, LFG does not overtly represent this relationship in the lexical entry.

APG, SFG, LFG and GB use different descriptive tools for these constructions. For example, NP movement in GB theory is analogous to promotion in RG, e.g., in passive the direct object is promoted to subject. SFG represents this facet of RG by means of a "stratified label" like [2, 1], indicating that the value of the feature labeled [2, 1] is an initial 2 (direct object) which is promoted to a 1 (subject).

GB uses movement between syntactic positions as a means for comparing a construction with canonical word order to a derived passive with different word order. Movement from one syntactic position to another represents the change in syntactic properties resulting from the derivation of a passive form from a non-passive form. In contrast, RG uses numbers as a similar type of metaphor—the change of an item with grammatical relation 2 in one strata to grammatical relation 1 in the next strata is promotion on the analogy of our number system in which 1 is ordered before 2. In RG, the change is
not thought of in terms of syntactic position, but in terms of abstract grammatical relations.
CHAPTER 2

A Unification-based Model for Government and Binding Theory

2.1 Introduction

This chapter formulates the basic concepts of Government and Binding Unification Grammar (GBUG), our unification-based model for Government and Binding Theory and defines the basic unification grammar and graph theory terminology used in the remaining chapters of the dissertation. GBUG provides:

(1) Data structures for representing GB descriptions of lexical entries, phrases, and some constraints;

(2) Formal mechanisms for instantiating: licensing relations; steps in parsing algorithms; multiple inheritance in a hierarchical lexicon; and constraint satisfaction.


We depart from most previous work in GB theory in assuming a data structure as compatible as possible with the data structures of other (unification-based) linguistic frameworks. Thus when GB analyses are encoded in GBUG: differences between GB and other linguistic frameworks become less of a barrier for contrasting analyses of the same syntactic phenomenon; GB concepts may be more readily adopted by other linguistic frameworks; and concepts from other linguistic frameworks may be more readily adopted into GB. The formulation of GBUG is a major step towards determining which differences between GB theory and other frameworks are superficial and which differences are significant.

2.2 Feature Structures

This section defines our basic data structure, the feature structure (FS), and demonstrates how FSs are used to represent lexical entries, phrases and constraints. We also demonstrate how we use the
Feature Structure Economy Principle (FSEP) to evaluate competing FS-based analyses. The FSEP is a straight-forward and intuitive means for comparing FSs which represent the same licensing relations.

The term FEATURE STRUCTURE is defined as follows:

**Definition 2.1 Feature Structure:** A FS is either:

1. **A COMPLEX FS** — a set of features and their values.
2. **An ATOMIC FEATURE STRUCTURE** — a FS without any internal structure.

Figure 2.1 is a simplification of a GBUG lexical entry for *eat* represented as a FS consisting of three feature value pairs. The value of the feature Category is the atomic FS Verb. The value of the feature Head-Proj is a complex FS representing the head of the phrase. The value of the feature Specifier represents an NP-trace. The indices i and j indicate shared values of two or more features—the Category features share the value: Verb; the features Complement, Internal-Theta and Complement-Case (in the FS value of the feature Head-Proj) share the value Category: Noun. This indicates that the licensing relations Complement-Case(Verb, \( \overline{N} \)), Theta(Verb, \( \overline{N} \)) and Complement(Verb, \( \overline{N} \)) hold, i.e., the verb assigns (accusative) case and a theta role to its \( \overline{N} \) complement.\(^3\)

<table>
<thead>
<tr>
<th>An initial lexical entry for eat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category:</td>
</tr>
<tr>
<td>Specifier:</td>
</tr>
<tr>
<td>Head-Proj:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.1**
Kay (1979, p.143) requires well-formed FSs to be functional where FUNCTIONALITY is defined as follows:

**Definition 2.2 Functionality:** Given FS X, a set of feature value pairs \([F_1, V_1], [F_2, V_2], \ldots, [F_n, V_n]\), X is functional if no feature \(F_i\) occurs in more than one feature value pair in X. Thus the feature (or attribute) uniquely identifies a value in X.

For example, in Figure 2.1, the features *Category*, *Complement-Case* and *Head-Proj* each have one corresponding value. None of these features are repeated with different values. A perfectly well-formed feature structure may contain two differently valued instances of the same feature at different levels of embedding, e.g., there are five instances of the *Category* feature in Figure 2.1: three with the value *Verb*, one with the value *Noun*, and one with the value *NP-Trace*. However, only one *Category* feature occurs in each set of feature value pairs: the set of feature value pairs at the outer most level; the set of feature value pairs making up the value of each of the *Head-Proj*, *Specifier* and *Complement* features.⁴ As discussed in later sections of this chapter, this functionality requirement insures that subsumption and unification relations are well-defined for FSs.

Figure 2.1 approximately corresponds to Figure 2.2, a constituent structure tree diagram, also representing the lexical entry for *eat*. Observe that there is no explicit indication of case and theta relations. For this constituent structure tree, it may be possible to derive definitions for *specifier* and *complement* based on the positions of constituents in the diagram. However in Chapter 3 we show that definitions of the terms *specifier*, *complement* and *adjunct* in terms of constituent structure tree positions is not possible in general, i.e., once all standard GB analyses of phrase structure are taken into account.
A Complex FS is an abstract concept modeled alternatively as edge-labeled directed acyclic graphs (DAGs) or finite partial functions (FPFs). An Atomic FS is modeled alternatively as a labeled node (in a graph) or a logical constant. Modeling FSs as DAGs or FPFs has become a standard practice in the literature on unification-based grammars and unification-based parsing.

Both FPFs and DAGs have well-understood mathematical properties. For the most part, we use the DAG model. However, the FPF model is useful for dealing with negation and disjunction of FSs, as well as so-called "typed" FSs (cf. Carpenter 1992), among other topics. Figures 2.3 to 2.6 are different ways of representing the same FS, which we will call FS₁. These diagrams should not be viewed as representing actual FSs, but rather descriptions of FSs. As is the practice in the literature, the matrix diagram in Figure 2.3 and the graph diagram in Figure 2.4 both depict DAG representations of FSs. Of these two, it is easier to display FSs in the tabular format in text, especially for large examples, while the graph format is a clearer representation of the arcs and nodes making up
a DAG. We use the graph theory terminology defined below in reference to both. Figures 2.5 and 2.6 are both conjunctions of logical formulas written in GB Logic (GL), a logic for describing FSs (A symbolizes logical conjunction) based on Kasper and Rounds (1986, 1990) and Johnson and Moss (1993, 1994), among others. Generally, we find it most convenient to describe lexical entries and phrases with DAGs, but to describe constraints using GL formulas. The main difference between the
two representations is that it is easier to extend FS logic to more expressive logics, incorporating disjunction, negation, etc. in order to model constraints which cannot be represented as FSs alone.

Figure 2.5

Feature1 (Feature2 (Atom1) \ \& \ \text{Feature3} (\text{Atom2}))
\& \ \text{Feature1} \ \text{Feature3} = \ \text{Feature4}
\& \ \text{Feature5 (Atom3)}

Figure 2.6

Feature1 \ \text{Feature2 (Atom1)}
\& \ \text{Feature1 Feature3 (Atom2)}
\& \ \text{Feature1 Feature3 = Feature4}
\& \ \text{Feature5 (Atom3)}

In Figure 2.3, the value of the feature corresponding to the label Feature1 is a complex FS consisting of two feature value pairs. The values of the features represented by the labels Feature4 and Feature5 are atomic FSs represented by the atomic labels Atom2 and Atom3.\textsuperscript{6} The \[\square\] indexes indicate that both Feature3 and Feature4 have Atom2 for a value. Although the shared value in this example is an atomic FS, both complex and atomic FS values of features may be shared. The difference between Figures 2.3 and 2.4 is merely a question of notation since both figures represent precisely the same data structure. The tabular format (Figure 2.3) is easier to use for representing large examples in text and the graph format (Figure 2.4) is a clearer representation of the data structure being represented.

Figure 2.4 is a graph representation of FS\textsubscript{1}.\textsuperscript{7} (We provide a brief description of the DAG in Figure 2.4 here, leaving more complete definitions of DAG and other graph theory terms to Section 2.4). In Figure 2.4, features are represented by arcs, connecting pairs of nodes. All arcs and some nodes bear labels. A label $L_{f}$ on an Arc $A$ indicates that $A$ represents feature $F$. A label $L_{X}$ on a node $n$ typically indicates that that node represents an atomic FS $X$ (We also use node labels for other purposes...
The first of the two nodes connected by an arc is its SOURCE and the second node is its TARGET.\(^8\) The node labeled \(r\) is the root of \(\text{FS}_1\), where a ROOT is a node which is the source of one or more arcs, but is not the target of any arc. In Figure 2.4, the arcs labeled \(\text{Feature3}\) and \(\text{Feature4}\) meet at the node labeled \(\text{Atom2}\). This phenomenon is called MULTIATTACHMENT, NODE SHARING, STRUCTURE SHARING or GRAPH REENTRANCY. Node sharing, inter alia, graphically represents that features \(\text{Feature3}\) and \(\text{Feature4}\) both have \(\text{Atom2}\) as their value. This is equivalent to the indexing notation in Figure 2.3.\(^9\)

Figures 2.5 and 2.6 describe \(\text{FS}_1\) in terms of the logical formula which it satisfies. In FS Logic, a FS is described with a logical constant (an Atom) denoting itself or a conjunction of zero or more logical formulas of the following types:

1. A pairing of a PATH and its value, where a path is a sequence of one or more feature labels and its value is a FS.

2. Two paths separated by an equal sign, representing that the two paths have the same value. This is called a PATH EQUIVALENCE formula. It represents in FS logic what node sharing represents in DAGs;

3. A disjunction of FS formulas; or

4. The negation of a FS formula\(^{10}\)

Consider the feature value pairs in Figures 2.5 and 2.6. The formula \(\text{Feature5} (\text{Atom3})\) consists of the path \(\text{Feature5}\) and its value \(\text{Atom3}\); the formula \(\text{Feature1 Feature2(Atom1)}\) consists of the path \(\text{Feature1 Feature2}\) and its value \(\text{Atom1}\); and the formula \(\text{Feature1 (Feature2 (Atom1) \∧ Feature3(Atom2))}\) represents the path \(\text{Feature1}\) and its value a complex FS \(\text{Feature2 (Atom1) \∧ Feature2 (Atom2)}\). Figures 2.5 and 2.6 represent that the two paths \(\text{Feature1 Feature3}\) and \(\text{Feature4}\) have the same value using the path equivalence formula \(\text{Feature1 Feature3} = \text{Feature4}\). The path equivalence formula does not explicitly specify the value of \(\text{Feature4}\), but \(\text{Feature4}\)'s value is derivable from the path equivalence formula and the formula stating the value of the path \(\text{Feature1 Feature3}\). As shown in Kasper and Rounds (1986, 1990), it is pos-
sible to transform formulas consisting of paths with complex FS values into formulas consisting solely of paths with atomic FS values using a simple algorithm. In this way Figure 2.5 can be shown to be logically equivalent to Figure 2.6. Following Kasper and Rounds (1986, 1990), Figure 2.6 is in the normal form for FS logic formulas used in this paper: all paths are local, where LOCAL paths are paths which have atomic FS values.

Figures 2.1, 2.7 and 2.8 each model the lexical entry for *eat* as a FS which they each partially describe either as a DAG (another FS) or a logical formula satisfied by the lexical entry for *eat*. The actual FS representing *eat*, may have details left out in these descriptions—that is why they are called partial descriptions.

Above, we only discuss examples of FSs representing lexical entries. However, FSs may represent a variety of other linguistic objects, including (1) complete phrases such as *John saw Mary* and *eat chocolate* in Figures 2.9 and 2.10; and (2) some linguistic constraints, such as the case filter\textsuperscript{11}, as in Figure 2.11. (See Section 2.4 for a more complete listing of the sorts of linguistic objects we represent as FSs.) The relation between lexical entries and phrases is transparent. Figure 2.1, the lexical entry for *eat*, is a partial specification of all phrases anchored by *eat* including Figure 2.10, a representation of the \( \overline{V} \text{eat chocolate} \) (as in *John wants to eat chocolate*.) Figure 2.10 is just like Figure 2.1, except that the value of the path *Head-Proj Complement* is more fully specified. In a sense, Figure 2.1 may be viewed as a constraint that each instantiation of that lexical entry for *eat* must satisfy and Figure 2.10 may be viewed as a FS which satisfies this constraint.
Figure 2.7
Category (Verb)
Λ Specifier Category (NP-trace)
Λ Head-Proj Category = Category
Λ Head-Proj Head-Proj Category = Category
Λ Head-Proj Head-Proj Phonology (eat)
Λ Head-Proj Complement Category (Noun)
Λ External-Theta = Specifier
Λ Head-Proj Internal-Theta = Head-Proj Complement
Λ Head-Proj Complement-Case = Head-Proj Complement

Figure 2.8

A FS representation of
John saw Mary

Category: [I] INFL
Tense: [I] Past
Specifier: [K] Category: Noun
Phonology: John
Antecedent: [I]
Head-Proj: Category: [I]
Tense: [I]
Head-Proj: Category: [I]
Tense: [I]
Complement: [M] Category: [N] Verb
Specifier: [O] Category: NP-trace
Same-reference: [I]
Head-Proj: Category: [I]
Head-Proj: Category: [I]
Phonology: saw
Complement: [P] Category: Noun
Phonology: Mary
Internal-Theta: [P]
Complement-Case: [P]
External-Theta: [O]
Internal-Theta [M]
Specifier-Case [K]

Figure 2.9
For simple and coherent linguistic analyses, it is advantageous to model syntactic constraints using the same abstract model as the phrases and lexical entries they constrain. This enables us to use a finite number of formal mechanisms to represent constraint satisfaction, case/theta role assignment, relationships between lexical entries, lexical entry construction, parsing steps, etc. For example, Figure 2.11 represents the GB constraint known as the case filter. In GB theory, the Case Filter requires that Ns and Gerunds that are assigned theta roles are also assigned case (see note 11). FSs which satisfy Figure 2.11 represent analyses which satisfy the case filter, where a precise definition of constraint satisfaction is provided in Section 2.4. Intuitively, each FS X which satisfies Figure 2.11 is partially described by Figure 2.11.

The explicitness of the FS model provides us with means of comparing the simplicity of one syntactic representation to another:
Definition 2.3 The Feature Structure Economy Principle (FSEP): Given two Feature Structures $FS_1$ and $FS_2$ representing the same licensing relations, $FS_1$ is preferred iff $FS_1$ represents these licensing relations using fewer features (arcs) than $FS_2$.

The FSEP is a principle of economy of representation in the sense of Chomsky (1991)—it is a concrete specification of one criterion for comparing syntactic representations by Occam’s Razor. For example, Figure 2.12 (below) is preferred to Figure 2.9 by the FSEP. There are two steps for making this determination:

1. Based on evidence from Chapters 6 and 7, we show that Figures 2.9 and 2.12 represent the same licensing relations.

2. We observe that Figure 2.12 has fewer arcs than Figure 2.9

Chapter 6 shows that the mechanism for marking coreference between the NP-trace and its antecedent in Figure 2.9 has the same semantics as the structure sharing representation in Figure 2.12. It follows from our account that the mechanism for representing and determining coreference need not apply here, i.e., the NP-trace can be eliminated from Figure 2.9, along with the Same-Reference and Antecedent arcs which represent coreference resulting in the representation shown in Figure 2.12. Chapter 7 shows that all relations of type case and theta are actually subtypes of relations of type complement and type specifier. For example, the relation Internal-Theta(X,Y) implies the relation Complement(X,Y) because the former is a subtype of the latter. Therefore, Internal-Theta(X,Y) A Complement(X,Y) always reduces to Internal-Theta(X,Y). We create a new type of arc labeled as MINIMAL A-CHAIN (MAC) which is both a type of Internal-Theta arc and a type of Complement-Case arc. Thus the path Head-Proj Complement Head-Proj MAC in Figure 2.12 represents that its value is the licensee of a complement-case, internal-theta and complement relations and the value of the path Head-Proj Complement Head-Proj Head-Proj is the licensor of each of these relations. These facts show that Figures 2.9 and 2.12 represent the same licensing relations, but Figure 2.12 uses fewer features (arcs) for this purpose. Therefore Figure 2.12 is preferred by the FSEP.
This section introduced and motivated a FS-based model for GB analyses. We showed how FSs may be used to represent aspects of GB theory and briefly indicated some of the advantages over the constituent structure tree model. The remaining sections of this chapter: (1) refine the FS model, defining graph theory terms and operations on FSs like subsumption and unification; and (2) show how the various formal devices are used.

2.3 Directed Acyclic Graphs

This section introduces the concepts DAG, arc, node and other graph theory terms used in this dissertation. The purpose of this section is to review some of the key graph theoretic terms needed to talk about FSs and indicate (in graph theoretic terms) how GBUG relates to previous unification based formalisms. Shieber (1986) notes that:

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. (Shieber 1986, pp. 20-21)

The use of DAGs and other graph theoretic entities originates with Relational Grammar (RG) and Arc-Pair Grammar (APG, a framework which has since been renamed Metagraph Grammar). See Johnson (1974), G. Lakoff (1977), Postal (1977), Perlmutter and Postal (1977) and Johnson and Postal (1980). Many of the more recent frameworks (e.g., HPSG) have adopted RG-based analyses, including RG analyses of raising and control phenomena: in place of RG graphs, they use FSs; where RG uses the term multi-attachment, they employ the equivalent notion node sharing.

As is clear from the term edge-labeled single-rooted directed acyclic graph, DAGs have several formal properties. We define these properties below (in the citations, the emphasis is our own):

- Berge (1973, p.3) defines the term graph as follows:

  Formally, a graph G is defined to be a pair (X,U), where
  
  (1) X is a set \( \{x_1, x_2, \ldots, x_n\} \) of elements called vertices [nodes], and

  (2) U is a family \( (u_1, u_2, \ldots, u_m) \) of elements of the Cartesian product \( X \times X \), called arcs. This family will often be denoted by the set \( U = \{1,2, \ldots, m\} \) of its indices. An element \( (x,y) \) of \( X \times X \) can appear more than once in this family. A graph in which no element of \( X \times X \) appears more than \( p \) times is called a \( p \)-graph.

Berge’s definition is essentially the definition of graph assumed in APG as well as unification-based frameworks. It differs from some other definitions of graph in graph theory in that it assumes that:

(1) the graph is directed, i.e. that the indices of each arc are ordered; and (2) the arcs are primitive, not defined as pairs of nodes (as in other versions of graph theory).

- A (directed) graph is ACYCLIC if it contains no cycles, i.e., no path of one or more arcs connects a node to itself, where in graph theoretic terms a PATH is a set of arcs \( \{a_1, \ldots, a_n\} \), such that the target of \( a_i \) is the same as the source of \( a_{i+1} \), for \( 1 \leq i \leq n - 1 \). For example, a directed graph containing the set of arcs in Figure 2.13 would not be acyclic.

- A graph G is SINGLE-ROOTED if there is only one node r in the graph which is not the target of any arc. r is called the ROOT of G.\(^{13}\)
EDGE-LABELED graphs have the additional property that each arc bears a label. Edge (or arc) labels provide one way of distinguishing arcs which connect the same two nodes, each arc bearing a different label corresponding to a distinct feature. For example in Figure 2.7, the arc labeled Complement and the arc labeled Internal-Theta represent distinct features (FSEP considerations aside), but have the same source and target nodes. Following Moss (1992) and David Johnson (personal communication), we assume that label arc is defined as a partial function from a node and a label to a node.

GBUG follows SFG in that it permits node labeling. In GBUG, as in SFG, atomic FSs are represented as labeled nodes, as in Figure 2.4. We (informally) use a label on node n as an abbreviation for a FS rooted at a node n (or as an abbreviation for a constraint on all FSs which may be rooted at node n), as in Figure 2.10, where the node label \( \equiv \) represents that the value of the arc labeled Specifier must be an \( \equiv \).

2.4 Subsumption

2.4.1 Introduction

This section defines subsumption, a key concept in unification-based grammars. Subsumption, is a relation between two elements X and Y, such that if X SUBSUMES Y, X is more general than Y, or Y is an instance of X. We use Subsumption for the following purposes:
Subsumption refines the notion ‘Y is a subtype of X’ as ‘X subsumes Y.’ For example, we organize the lexicon into various classes of lexical items defined in terms of the subsumption relation. Thus given \( FS_v \) representing the set of lexical entries of type verb, \( FS_t \) representing the set of lexical entries of type transitive verb and \( FS_{eat} \) representing the lexical entry for the verb eat, \( FS_v \) subsumes \( FS_t \) and \( FS_t \) subsumes \( FS_{eat} \).

Subsumption provides a clear means of representing some instances of constraint satisfaction. Given \( FS_C \) representing a constraint \( C \) and a \( FS_P \) representing a phrase \( P \), if \( FS_C \) subsumes \( FS_P \), then \( P \) satisfies \( C \). For example, for each phrase \( P \) satisfying the case filter, we propose that \( FS_{CF} \) subsumes \( FS_P \), where \( FS_{CF} \) represents the Case Filter.

Subsumption is crucial to the definition of the unification operation discussed in Section 2.5: \( Z \) is the unification of \( X \) and \( Y \), iff \( X \) is the most general entity (the smallest description) such that \( X \) subsumes \( Z \) and \( Y \) subsumes \( Z \).

This section has two basic objectives: (1) to define the subsumption relation with regard to features (arc labels), atomic FSs (node labels), disjunctive labels, and complex FSs (DAGs) and (2) to describe how subsumption is utilized in our model of grammar and our lexicon.

### 2.4.2 Defining Subsumption

We define subsumption as follows:

**Definition 2.4 Subsumption:** \( A \subseteq B \) if the set of all \( B \) is a subset of the set of all \( A \).

This definition provides the basis for determining that the following subsumption relations hold between the FSs listed in Figure 2.14:

- \( FS_1 \subseteq FS_2 \) and \( FS_1 \subseteq FS_3 \), where \( FS_2 \) and \( FS_3 \) are the values of \( Head-Proj \) in \( FS_2 \) and \( FS_3 \). \( FS_2 \) and \( FS_3 \) differ from \( FS_1 \) in the following respects: (1) they specify a value for the \( Category \) feature; and (2) they contain some feature value pairs not in \( FS_1 \). These differences narrow the set of items which \( FS_2 \) and \( FS_3 \) can describe. Thus \( FS_2 \) and \( FS_3 \) each represent a subset of the set of items \( FS_1 \) represents. \( FS_2 \) and \( FS_3 \) are each a type of \( V^0 \) and \( FS_1 \) represents the set of all \( X^0 \)s.
\( FS_2 \subseteq FS_3 \). The difference between these FSs is that the value of \textit{Head-Proj} in \( FS_3 \) is more specific (informally, it contains more information) than the value of \textit{Head-Proj} in \( FS_2 \). \( FS_2 \) represents a member of the set of verbs which assign external theta roles (i.e., in which the relation \text{External-Theta}(V,N) holds); and \( FS_3 \) represents a member of the set of transitive verb. \( FS_2 \subseteq FS_3 \), since a transitive verb is a type of verb taking a subject.

\( FS_1 \) and \( FS_4 \) contain different feature value pairs with different features: thus while there are items which are characterized by both, neither subsumes the other. \( FS_4 \) represent an immediate projection (\( \overline{X} \) or \( \overline{\overline{X}} \)) of any case licensor (case assigner). \( FS_1 \) subsumes both case licensors and noncase licensors. Therefore it is not surprising that these FSs do not conflict with each other or that there are in fact objects partially described by both \( FS_1 \) and \( FS_4 \), i.e. immediate projections of \( X^0 \) case assigners.

\( FS_2 \) and \( FS_5 \) are incompatible with each other because they have different values for the same feature: \( FS_2 \) contains the feature value pair \textit{Category: Verb}, whereas \( FS_5 \) contains the feature value pair \textit{Category: Preposition}. Due to this incompatibility, neither subsumes the other and there are no items which can be characterized by both FSs simultaneously. \( FS_5 \) represents a nonterminal projection.
of $P$ containing a $P^0$ case-assigner and $FS_2$, is a nonterminal projection of $V$. This conflict fits our linguistic assumptions, i.e., we assume that a phrase cannot simultaneously be of category *Verb* and category *Preposition*.
In the following citation from Shieber (1986, pp. 14-17)\textsuperscript{15} the subsumption relation is defined for complex FSs. Shieber follows previous research in unification-based formalisms in assuming that: (1) only the values of identical arc labels are checked for the subsumption relation; and (2) an atomic FS F subsumes a FS F’ only if F and F’ are identical or if F is a variable, where a VARIABLE is a FS which subsumes all other FSs. A variable is typically represented by the symbol Nil since a variable is the least specified FS. Nil is represented in GBUG as a node with node label Nil.\textsuperscript{16}

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 $D \subseteq D'$) 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 $D(l) \subseteq D'(l)$ 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 $\vec{N}$. They are not incompatible, but neither subsumes the other]... or because they have conflicting information [e.g., a 3rd person $\vec{N}$ and a 1st person $\vec{N}$]... (Shieber 1986, pp. 14-17)

Following Kay (1979), we assume special Feature values including Nil and True, the latter corresponding to Kay’s ANY. We also assume the feature value False, which is equivalent to $\neg$True:

- **Nil** — The least specified FS. Nil subsumes all other FSs. In GBUG, in matrix notation, Nil is expressed by either listing Nil as the value of some feature or leaving the value of the feature unspecified. In graph notation, Nil is expressed as an atomic FS with the label Nil. Alternatively, we refer to Nil as a ‘Null FS’ since, conceptually, Nil is a FS consisting of zero feature value pairs.

- **True** — True subsumes all complex and atomic FSs, except Nil and False. (Nil $\subseteq$ True). We give True the interpretation ‘this value is required.’ We assume that a FS representing a well-formed complete phrase cannot have True as the value of any of its features. For example, when the value of a Complement feature is True in a lexical entry, this indicates that the complement is obligatory.

- **False** — Nil $\subseteq$ False and False does not subsume nor is subsumed by any other FS. A feature with False has the interpretation ‘this feature cannot have a value,’ i.e., False is equivalent to $\neg$True. We use the FS value False in constraints.
Based on innovations of SFG (cf. Johnson and Moss 1993, 1994 and Johnson, Meyers and Moss 1993a,b), we allow subsumption to hold between feature labels. Based on the recent literature on typed FSs (cf. Carpenter 1992, among others), we allow the subsumption relation to hold on atomic FS (node) labels. Accordingly, we revise the definition of subsumption on complex FSs to allow for arc label and node label (atomic FS) subsumption. GBUG allows the subsumption relation to be asserted between two labels (arc labels or node labels). The assertion ‘A ⊆ B’ means something like ‘B is a subtype of A’, assuming that every element is also a subtype of itself to allow for the possibility that Type A = Type B. This conforms to Definition 2.4 since if B is a subtype of A, then B characterizes a subset of the items characterized by A. Since labels are not divisible into components in the same way as FSs, it would be hard to formulate an understanding of this type of subsumption in terms of Definition 2.4. For example, the following statements assert that: the atomic FS label (atomic label) Nominal subsumes the label Noun and the label Gerund and that the arc label Theta subsumes the label Internal-Theta and the label External-Theta.

- Nominal ⊆ Noun,
- Nominal ⊆ Gerund,
- Theta ⊆ Internal-Theta,
- Theta ⊆ External-Theta.

In each of these cases a more general term is asserted to subsume a more specific term.

GBUG defines the subsumption relation on DISJUNCTIVE labels, labels consisting of sets of labels separated by the operator |, representing inclusive or, e.g., Gerund | Noun and Internal-Theta | External-Theta. The disjunction of n items may be viewed as a union of sets. Given Label\(_1\), a disjunction of labels \(L_1...L_n\) and Label\(_2\), a disjunction of labels \(L_{i1}...L_{jn}\) \(1 \leq i \leq j \leq n\), the set of disjuncts in Label\(_1\) is a superset of the set of disjuncts in Label\(_2\). Label\(_1\) represents the set of items \(I_1\), the union of the sets of items represented by \(L_1...L_n\). Label\(_2\) represents the set of items \(I_2\), the union of the sets of items represented by \(L_{i1}...L_{jn}\), a subset of \(I_1\). Thus, by Definition 2.4, Label\(_1\) ⊆ Label\(_2\). For example, the label Gerund | Noun subsumes the label Gerund as well as the label Noun because the set of all items...
which are of type Gerund-or-Noun is a superset of both the set of items of type Gerund and the set of items of type Noun. Similarly, Internal-Theta | External-Theta subsumes the label Internal-Theta as well as the label External-Theta because the set of features of type Internal-Theta-or-External-Theta is a superset of both the set of features of type Internal-Theta and the set of features of type External-Theta. In each case, a disjunctive label D represents a wider range of features, than a disjunctive label \( D' \) consisting of a subset of the disjuncts in D. The subsumption relation on disjuncts is defined as follows:

**Definition 2.5 Subsumption of Disjunctions:** A Subsumes B, if A and B are disjunctions and the disjuncts in B are a subset of the disjuncts in A. More formally:

For all \( X_i \) and \( X_j \) in \( X_1 \cdots X_n \), \( 1 \leq i \leq n, i \leq j \leq n \):

\[
X_1 | X_2 | \cdots X_n \subseteq X_i | X_{i+1} | \cdots X_j.
\]

Once we allow label subsumption, it is possible for two distinct labels to describe the same feature or atomic FS (each label providing information which varies in content or specificity). Thus Shieber’s definition of subsumption for complex FSs (see above) must be replaced with the following definition:

**Definition 2.6 Subsumption for Complex FSs:**

\( FS_A \subseteq FS_B \) iff: (1) For each [F, V] in FS\(_A\), there exists some [\( F' \), \( V' \)] in FS\(_B\) such that F \( \subseteq F' \) and V \( \subseteq V' \), and (2) For each pair of paths p and q in FS\(_A\) such that p = q, there exist paths \( p' \) and \( q' \) in FS\(_B\) such that: each arc in p subsumes the corresponding arc in \( p' \), each arc in q subsumes the corresponding arc in \( q' \) and \( p' = q' \).

In Figure 2.15, FS\(_1\) \( \subseteq \) FS\(_3\) and FS\(_2\) \( \subseteq \) FS\(_3\). For each feature value pairs [F, V] in FS\(_1\) (or FS\(_2\)) and [\( F' \), \( V' \)] in FS\(_1\), such that F \( \subseteq F' \), V \( \subseteq V' \). The features Category, Head-Proj, and Complement in FS\(_1\) and FS\(_2\) each subsume a feature in FS\(_3\) under identity. The feature Theta in FS\(_1\) subsumes both the feature Internal-Theta in FS\(_3\) and the feature Internal-Theta | External-Theta in FS\(_2\) by the mechanisms defined above. Furthermore, all structure sharing paths in FS\(_1\) and FS\(_2\) have corresponding paths in FS\(_3\) which satisfy the definition above.
Above we have provided two ways in which one label\textsuperscript{18} may subsume another: (1) by assertion as in $\text{Nominal} \subseteq \text{Noun}$; and (2) according to the logical properties of disjunction, as in $\text{Noun} \mid \text{Gerund} \subseteq \text{Noun}$. (2) is perhaps an instance of a more general case, also exemplified by stratified labels in SFG. (See Chapter 6 and Johnson, Meyers and Moss (1993a,b) and Johnson and Moss (1993, 1994) for details.) We could devise other operations for calculating subsumption on special types of labels in addition to disjunctive labels. Like disjunctive labels, these special labels would be formulated such that the internal syntax and semantics of the labels provide a means for determining subsumption. This is an interesting area for future research.

2.4.3 The Role of Subsumption in our Grammar and Lexicon

This subsection defines the role that the subsumption operation plays in our grammar and lexicon. We show how subsumption is used to organize our lexicon hierarchically into classes and subc-
lasses and how subsumption provides a basis for some instances of constraint satisfaction. Each class may be represented by a FS subsuming all the members of that class. For example, given FS_v representing the class verb, FS_T representing the class transitive verb and FS_{eat} the lexical entry for eat, 
FS_v ⊆ FS_T ⊆ FS_{eat}. Then we discuss two examples of constraint satisfaction using subsumption: (1) A GB version of GPSG’s Head Feature Constraint and (2) GB’s case filter. We represent each of these constraints as a FS FS_{C}, the minimal FS which satisfies constraint C. We claim that only FSs subsumed by FS_{C} represent analyses that satisfy C.

Any FS X may be viewed as a constraint which is satisfied by any FS Y such that X ⊆ Y. If X represents a class of linguistic objects, X is a constraint that is satisfied by each Y representing one or more members of X. For example, Figure 2.1 is satisfied by each use of that form of the verb eat in the sense that each lexical entry is a constraint satisfied by instances of a word using that lexical entry. In one use of eat, the value of the Complement feature may be a FS representing the \( N \) fish, and in another the value may be a FS representing the \( N \) a small piece of avocado. Figure 2.16 represents the linguistic object X⁰, viewed as a constraint which is satisfied by all objects of type X⁰, i.e., every X⁰ token can be represented by a FS which contains at least as much information as (is subsumed by) Figure 2.16.

<table>
<thead>
<tr>
<th>FS representing X⁰</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category: True</td>
</tr>
<tr>
<td>Head-Proj: False</td>
</tr>
</tbody>
</table>

Figure 2.16

As noted above, we assume that our FSs are well-typed in the sense of Carpenter 1992. Head-Proj (False) in combination with cooccurrence restrictions would prevent other features which can only cooccur with Head-Proj from occurring in X⁰’s (David Johnson, personal communication). For example, the feature Specifier, could not occur in an X⁰ assuming a cooccurrence restriction of the form: \( \neg \text{Head-Proj} \rightarrow \neg \text{Specifier} \).
In organizing a lexicon and a grammar, it is sometimes useful to categorize all lexical items and phrases into categories and subcategories in what is known as an inheritance hierarchy. A small portion of this hierarchy of types and subtypes of lexical items is provided graphically in Figure 2.17—each arc represents that the category at its source is a supertype of the category at its target. For example, the FS representing objects of type

![Figure 2.17](image)

*lexical-entry-for-verb* subsumes the FS representing objects of type *lexical-entry-for-verb-that-licenses-one-internal-theta-relation*, which in turn subsumes the type *lexical-entry-for-unaccusative-verb*. Viewed as constraint satisfaction, a FS X representing a lexical entry for any verb type listed in Figure 2.17 satisfies a constraint represented by a FS Y representing some verb type listed above it in the hierarchy such that there is a path of zero or more arcs connecting X and Y. It is quite common to organize unification-based lexicons this way. For example, see Carpenter (1992) and the sources cited.
there. In Carpenter’s approach, not only is the lexicon organized in an inheritance hierarchy, but the values of each feature are as well. Ultimately, we would like to adopt this approach for GBUG, but will leave the details to future research.

We model the portion of \( \bar{X} \) theory which requires the projections of the same head to be of the same category as a constraint written in GL. We view this constraint as a special case of GPSG’s Head Feature Convention (HFC) (cf. Gazdar, Klein, Pullum and Sag 1985, pp. 94-99). For each feature \( F \) assumed to be a Head Feature, the following constraint holds:

**Definition 2.7 The Head Feature Convention (GB version):**

For each feature \( F \), a Head Feature:

\[
F = \text{Head-Proj } F
\]

For GB theory’s version of \( \bar{X} \) theory, the only claim is that the feature \( \text{Category} \) is a head feature, i.e. \( F = \text{Category} \) in Definition 2.7. Each FS above representing a non-terminal satisfies this constraint, i.e., we describe this constraint as the formula:

(2.1) \( \text{Category} = \text{Head-Proj } \text{Category} \)

Formula (2.1) partially describes, i.e. subsumes, or is satisfied by, each of the FSs in Figures 2.1, 2.9, 2.10, 2.12, 2.14 and 2.15. Other possible head features include \( \text{Tense}, \text{Number}, \text{Person} \) and \( \text{Gender} \).

Figure 2.11, repeated below, represents our initial representation of the Case filter. The value of the path \( \text{Head-Proj Complement Head-Proj} \) in Figure 2.9, repeated below, satisfies this version of the Case filter. In GB theory, the case filter is satisfied if an \( N \) that receives a theta role (is the licensee of a theta relation), also receives case. Figure 2.11 is actually too strict, requiring that the \( \text{Theta} \) and \( \text{Case} \) arcs have the same source. In order to produce a more accurate representation of the case filter:

1. We must use **INVERSE PATHS**, a concept borrowed from SFG. Basically an inverse path describes the ascent from the target to the source of a path. Figure 2.18 describes a FS subsumed by The Feature Value pair \( \text{Category: } (\text{Noun} \mid \text{Gerund}) \) which is the value of one case feature and the value of one theta feature. The 2 unary inverse paths \( \text{Case}^{-1} \) and \( \text{Theta}^{-1} \) mean
that if you can, in effect, "travel up" one case arc or one theta arc from source of the category arc. Structure sharing between the case and theta arcs is implied since both inverse paths start at the same node.

(2) We must account for how chains of NP-traces relate to their antecedents with respect to the Case filter—we leave this latter task to Chapter 6 where we eliminate NP-traces (see Figure 2.12 above).
Formulas containing inverse paths of SFG are examples of areas of FS logics which can be used to describe FSs, but cannot be modeled as DAGs themselves. This example illustrates how we can expand our FS logic as needed, keeping the complexity of most FS logic formulas to a minimum, introducing complications only when necessary.

2.4.4 Summary

This section defined the subsumption relation and provided examples of its function in a GBUG-based grammar and lexicon. We extended the traditional notion of subsumption to range over arc and node labels as well as complex FSs (based on innovations of SFG and typed FS logics). We also introduce inverse paths from SFG. Subsumption provides a formal way for expressing linguistic generalizations of the sort item X is of type T or item X satisfies constraint C. However, subsumption on ordinary FSs is not always sufficient for modeling GB constraints.
2.5 Unification

2.5.1 Introduction

This section defines unification, a basic operation for combining (atomic and complex) FSs and features. GBUG follows SFG and work in typed FS logic, in that nonidentical arc and node labels can unify, where the definition of label unification follow from our definitions of label subsumption above.

The UNIFICATION of A and B, notated \( A \cup B \), is the minimal entity C, such that A subsumes C and B subsumes C, where MINIMAL means that there is no entity D such that D is subsumed by A and B, but D is not subsumed by C. Unification FAILS if A and B are incompatible, i.e., if an item C cannot exist such that C is subsumed by both A and B. For example, the unification of the atomic FSs Preposition and Verb (both possible values of the feature Category) fails because there does not exist any atomic FS which is subsumed by both.

This section defines the unification operation on arc labels, node labels, and FSs. The role of unification in a GBUG-based lexicon, grammar and parser is also discussed. Unification represents:

1. The instantiation of abstract agreement (including case), predicate (including theta) and constituent relations (including complement, specifier and adjunct);

2. The relationship among lexical items in an inheritance hierarchy — Lexical items are represented by the unification of FSs representing their superclasses. This fact is useful for organizing our lexicon.

When unification represents (1), it may be implemented as a step in parsing for combining constituents, e.g., combining an \( \overline{N} \) and a verb to produce a \( \overline{V} \).

2.5.2 Defining Unification

Shieber (1986) provides the following formal definition of unification defined on DAGs (FSs).

For our purposes, however D should be taken to be a FS, a node label or an arc label.

In formal terms, we define the unification of two feature structures \( D' \) and \( D'' \) as the most general feature structure D, such that \( D' \subseteq D \) and \( D'' \subseteq D \). We notate this \( D = D' \cup D'' \).

As we have seen, not all pairs of feature structures can be unified in this way; they may contain conflicting information. In this case, unification is said to fail. (Shieber, 1986, pp. 17-18)
Definitions 2.8 defines unification in terms of subsumption based on Definition 2.4:

**Definition 2.8 Unification:** \( C = A \cup B \) iff the set of all \( C \) is the intersection set of the set of all \( A \) and the set of all \( B \), i.e., \( C \) is a subset of \( A \) and \( C \) is a subset of \( B \) and there is no item \( X \), a member of \( C \), such that \( X \) is not a member of both \( A \) and \( B \). Unification fails when \( C \) is the empty set.

We now consider some possible applications of unification to the FSs in Figures 2.19 and 2.20.

1. \( \text{FS}_1 \cup \text{FS}_2 \) is \( \text{FS}_3 \). \( \text{FS}_1 \) and \( \text{FS}_2 \) each subsume \( \text{FS}_3 \) because \( \text{FS}_3 \) contains a superset of the feature value pairs contained in both \( \text{FS}_1 \) and \( \text{FS}_2 \) separately. Since \( \text{FS}_3 \) is made up of the set union of the sets of feature value pairs in \( \text{FS}_1 \) and \( \text{FS}_2 \), there can be no FS \( X \) subsumed by both \( \text{FS}_1 \) and \( \text{FS}_2 \), but not subsumed by \( \text{FS}_3 \). Therefore \( \text{FS}_3 \) is the minimal FS subsumed by both \( \text{FS}_1 \) and \( \text{FS}_2 \).

2. \( \text{FS}_1 \cup \text{FS}_4 \) fails because \( \text{FS}_4 \) and \( \text{FS}_1 \) conflict in that they have different values for the Category feature. Thus there can be no FS representation of a linguistic object representing a nonterminal that is simultaneously of type Category: Preposition and Category: Verb.

3. \( \text{FS}_6 \) is the result of unifying the value of the Complement feature of \( \text{FS}_5 \) with \( \text{FS}_4 \). The value of the Complement feature in \( \text{FS}_5 \) is the same as \( \text{FS}_4 \), except that \( \text{FS}_4 \) contains the additional feature value pair Complement-Case: \( \overline{N} \). Therefore \( \text{FS}_5 \) subsumes \( \text{FS}_4 \). \( \text{FS}_4 \) is subsumed by both itself (every FS subsumes itself) and the value of the Complement arc of \( \text{FS}_5 \). Therefore \( \text{FS}_4 \) is the unification of \( \text{FS}_4 \) and the value of the Complement arc of \( \text{FS}_5 \). \( \text{FS}_4 \) is identical to the value of the Complement feature in \( \text{FS}_6 \).
Figure 2.19
Recall that in Section 2.4, subsumption is defined on nonidentical arc labels and nonidentical node labels, an innovation we borrowed from the literatures on SFG and typed FSs. These definitions of subsumption interact with the general definitions of unification above. The unification of two labels A and B (node or arc labels) is the label C, subsumed by both A and B. Thus, based on the definitions of subsumption in Section 2.4, the unification of the pairs of labels in the first two columns in Figure 2.21, is found in the corresponding row in the third column. In each case, the unification of A and B is
always at least as specific as (defines at least as small a set of elements as) both A and B. In Section 2.4, we defined \textit{Nil} as the least specified feature value, the FS consisting of zero feature value pairs and representing the set of all linguistic objects. All FSs, complex or atomic are subsumed by \textit{Nil}. Thus the unification of \textit{Nil} and any FS A is A. We assert that \textit{Nominal} subsumes \textit{Noun} and that \textit{Nominal} subsumes \textit{Gerund}. Therefore the results of \textit{Gerund} \cup \textit{Nominal} and \textit{Noun} \cup \textit{Nominal} fall out of the definition of unification provided. Section 2.4 defines subsumption on two disjunctive labels X and Y, so that if X consists of the set of disjuncts D, Y consists of the set of disjuncts D' and D' is a subset of D, then X \subseteq Y. Given the above definition of unification, the unification of two sets of disjuncts A and B, is some set of disjuncts which is a subset of both A and B, namely the set intersection of A and B, as exemplified in Figure 2.21.

<table>
<thead>
<tr>
<th>LABEL A</th>
<th>LABEL B</th>
<th>A UNIFICATION B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>Noun</td>
<td>Noun</td>
</tr>
<tr>
<td>Noun</td>
<td>Nil</td>
<td>Noun</td>
</tr>
<tr>
<td>Noun</td>
<td>Nominal</td>
<td>Noun</td>
</tr>
<tr>
<td>Gerund</td>
<td>Nominal</td>
<td>Gerund</td>
</tr>
<tr>
<td>Internal-Theta</td>
<td>Theta</td>
<td>Internal-Theta</td>
</tr>
<tr>
<td>External-Theta</td>
<td>Theta</td>
<td>External-Theta</td>
</tr>
<tr>
<td>Noun</td>
<td>Gerund \mid Noun</td>
<td>Noun</td>
</tr>
<tr>
<td>Gerund</td>
<td>Gerund \mid Noun</td>
<td>Gerund</td>
</tr>
<tr>
<td>Internal-Theta</td>
<td>Internal-Theta \mid External-Theta</td>
<td>Internal-Theta</td>
</tr>
<tr>
<td>External-Theta</td>
<td>Internal-Theta \mid External-Theta</td>
<td>External-Theta</td>
</tr>
</tbody>
</table>

\textbf{Figure 2.21}

Figures 2.22 and 2.23 contain examples of FS unification, in which nonidentical labels unify. FS_4 in Figure 2.22 is the result of unifying the value of the \textit{Complement} feature of FS_1 with FS_3 and FS_4 is also the result of unifying the value of the \textit{Complement} feature of FS_2 with FS_3. FS_7 is the unification of FS_5 and FS_7. FS_7 is also the unification of FS_6 and FS_7.
Figure 2.22

**FS_1**

Category: [i] Preposition
Head-Proj: Category: [i]
   Phonology: about
Complement: [i] Category: Noun | Gerund
Complement-Case: [j]

**FS_2**

Category: [i] Preposition
Head-Proj: Category: [i]
   Phonology: about
Complement: [i] Category: Nominal
Complement-Case: [j]

**FS_3**

Category: [i] Noun
Head-Proj: Category: [i]
   Phonology: Gertrude

**FS_4**

Category: [i] Preposition
Head-Proj: Category: [i]
   Phonology: about
Complement: [i] Category: [k] Noun
Head-Proj: Category: [k]
   Phonology: Gertrude
Complement-Case: [j]
2.5.3 The Role of Unification in Grammar, the Lexicon and Parsing

Unification plays a variety of roles in our models of grammar and processing:

1. Unification represents the instantiation of a licensing relation between two constituents, e.g., if an $\bar{N}$ unifies with the target of one $\Theta$ arc, one $\text{Case}$ arc and/or one $\text{Complement}$ arc, this represents that the relations $\Theta(X, \bar{N})$, $\text{Case}(X, \bar{N})$ and/or $\text{Complement}(X, \bar{N})$ have been instantiated for some $X$.

2. Unification provides a means of representing the intersection set of classes of lexical items, e.g., $\mathcal{FS}_T = \mathcal{FS}_C \cup \mathcal{FS}_E \cup \mathcal{FS}_I$, where $\mathcal{FS}_E$ represents the class of verbs assigning one external theta role, $\mathcal{FS}_I$ represents the class of verbs assigning one Internal theta role and $\mathcal{FS}_C$ represents the class of verbs assigning one case. This assumes that the set of verbs of the class
represented by $\text{FS}_T$ is the intersection set of the sets of verbs in the classes of represented by $\text{FS}_I$, $\text{FS}_E$ and $\text{FS}_C$.

(1) is implemented in parsing, as a step which builds constituents from subconstituents. For example, a parsing step may consist of unifying an $\overline{N}$ with the value of a Complement feature in the lexical entry for a verb. In this way a FS representing a $\overline{V}$ (or $\overline{\overline{V}}$) is derived by unification from a FSs representing one $V^0$ and one $\overline{N}$. All licensing relations for which that $\overline{N}$ is a licensee will be instantiated simultaneously.

Consider the FSs in Figures 2.24 and 2.25 representing lexical entries of the bare infinitive form of *eat* and *chocolate*. The FS for *chocolate* can unify with the value of the path Head-Proj Complement in Figure 2.24 resulting in the FS in Figure 2.26 representing the $\overline{V}$ *eat chocolate* (the bare infinitive form of *eat* is the morphological form of *eat* appearing in *I want to eat chocolate* and *I can eat chocolate*.) The value of the Head-Proj Complement path is structure shared with the paths Head-Proj Internal-Theta and Head-Proj Complement-Case. By unifying Figure 2.25 with the value of the path of arcs Head-Proj Complement, we provide a value for each of the structure sharing paths. Thus in one step, we instantiate the relations Complement(eat,chocolate), Internal-Theta(eat,chocolate) and Complement-Case(eat,chocolate). This instance of unification could be implemented as a parsing step resulting in a complete $\overline{V}$ constituent, complete in the sense that all its surface constituents have been found. We define a SURFACE constituent as any constituent which is part of an S-structure representation (see Chapter 6 for details). Some constituents require more than one unification operation before being completed, e.g., the verb phrase *give Mary tea* requires that the values of two Complement features be filled.

Unification provides a means of representing a lexical item or class of lexical items alternatively as (1) a subclass of more than one other class of items; or (2) a combination of certain sets of characteristics. Each FS, a set of feature value pairs, alternatively represents some set of characteristics or the set of items represented by that set of characteristics. Therefore for some $\text{FS}_1 = \text{FS}_2 \cup \text{FS}_3$, where each FS represents the classes of lexical items $C_1$, $C_2$ and $C_3$ respectively: (1) the members of
C₁ form the intersection set of the members of C₂ and C₃; and (2) the characteristics which make an item a member of C₁ are a union of the characteristics that make items members of C₂ and C₃. Thus unification provides a way of both: (1) characterizing subclass/superclass relations among classes of lexical items; and (2) representing a particular item in a hierarchy as the combination of the characteristics of its superclasses. Below we use unification to characterize (1) the relationships between various forms of the verb *eat*; and (2) the relationships between the class of verbs *transitive* and its superclasses. Once again, see Carpenter (1992) among others, for a more complete inheritance hierarchy which includes both the type of lexicon specified here and a hierarchical account of the possible values of each feature.

An initial lexical entry for *eat*

- **Category:** Infin-Verb
- **Specifier:** Category: Noun
- **Head-Proj:** Category: Noun
  - **Head-Proj:** Category: Noun
    - **Phonology:** eat
  - **Complement:** Category: Noun
    - **Internal-Theta:**
    - **Complement-Case:**
    - **External-Theta:**

**Figure 2.24**

An initial lexical entry for *chocolate*

- **Category:** Noun
- **Head-Proj:** Category: Noun
- **Phonology:** chocolate
- **Morph-Agr:** Third-Person-Singular

**Figure 2.25**
Figure 2.26

The $\bar{V}$
\[ eat \text{ chocolate} \]

Category: \[ I \] Infin-Verb
Specifier: \[ J \] Category: Noun
Head-Proj: Category: \[ I \]
  Head-Proj: Category: \[ I \]
  Phonology: eat
  Complement: \[ K \] Category: \[ I \] Noun
  Head-Proj: Category: \[ I \]
  Phonology: chocolate
  Morph-Agr: Third-Person-Singular

Internal-Theta: \[ K \]
Complement-Case: \[ K \]
External-Theta: \[ I \]

Figure 2.26

Compare Figures 2.24, 2.27, 2.28 and 2.29:

(1) Figure 2.24 represents the infinitive form of \textit{eat}, as noted above.

(2) Figure 2.27 represents the finite third-person-singular present form of \textit{eat}, the form found in \textit{They eat chocolate}.

(3) Figure 2.28 represents the entry for \textit{eaten}, the past participle form of \textit{eat}, the head of the complement of the auxiliary verb \textit{have}.

(4) Figure 2.29 represents the entry for \textit{ate}, the finite past form of \textit{eat}.

Figure 2.24 and 2.28, as well as the value of the path in Figures 2.27 and 2.29, which we will refer to as FS$_{27}$ and FS$_{29}$, differ in their values of exactly two paths: \textit{Category} and \textit{Head-Proj Head-Proj Phonology}. Figure 2.30 subsumes Figures 2.24 and 2.28, as well as FS$_{27}$ and FS$_{29}$, assuming that the following subsumption relations holds between node labels: \textit{Verb} \subseteq \textit{Infin-Verb}, \textit{Verb} \subseteq \textit{Finite-Verb}, \textit{Verb} \subseteq \textit{Past-Participle-Verb} and \textit{Noun} \subseteq \textit{NP-trace}. Figure 2.30 represents the class of uninflected/generic transitive verbs, which partially describes one lexical entry for the verb \textit{eat}. Figure 2.31 illustrates that Figure 2.30 can be unified with FS representations of distinguishing characteristics,
(Finite) Lexical entry for

\textit{eat}

- Category: INFL
- Tense: Present
- Specifier: Category: Noun
  - Morph-Agr: Third-Person-Singular
  - Antecedent: [l]
- Head-Proj: Category: [i]
  - Tense: [j]
  - Head-Proj: Category: [i]
  - Tense: [j]
  - Complement: Category: [m]
  - Specifier: Category: NP-trace
  - Same-Reference: [l]
  - Head-Proj: Category: [i]
    - Head-Proj: Category: [i]
    - Phonology: eat
    - Complement: Category: Noun
    - Internal-Theta: [p]
    - Complement-Case: [p]
    - External-Theta: [o]
  - Internal-Theta: [m]
  - Specifier-Case: [k]

\textbf{Figure 2.27}

An initial lexical entry for

\textit{eaten}

- Category: Past-Participle-Verb
- Specifier: Category: Noun
- Head-Proj: Category: [i]
  - Head-Proj: Category: [i]
    - Phonology: eaten
    - Complement: Category: Noun
    - Internal-Theta: [k]
    - Complement-Case: [k]
  - External-Theta: [i]

\textbf{Figure 2.28}
resulting in Figures 2.24, 2.27, 2.28 and 2.29.

Figure 2.17 (repeated below) represents part of the subsumption hierarchy of verb types. Each type is represented by a FS which subsumes all subtypes. Figure 2.30 above represents the class of
transitive verbs, a subclass of the following verb classes: Licenses-External-Theta-Role (Figure 2.32), Licenses-One-Internal-Theta-Role (Figure 2.33), and Licenses-One-Case (Figure 2.34). Since these three superclasses exhaustively cover verbs of type Transitive Verb, then Figure 2.30 = Figure 2.32 ∪ Figure 2.33 ∪ Figure 2.34. This unification represents that members of the class Transitive Verb form the intersection sets of members of all its superclasses.

We propose an inheritance hierarchy consisting of each lexical entry, each class of lexical entries, each class of phrases, etc. A FS representing any member in the hierarchy may be derived by unifying the FSs representing all its superclasses. In addition, it may be necessary to add some phonological and morphological information in generating lexical entries (we do not attempt to represent the phonological relationships between morphologically related lexical items here.) Each class of lexical entries, e.g., Transitive, has a precise relationship with its superclasses, e.g., Verbal Case Licensor, Verbal Internal Theta Licensor, Verbal External Theta Licensor. It should also be clear that more general types are possible such as Case Licensor, Internal Theta Licensor, Verb, Generic Word, Generic Licensor, etc.
Morphological Forms of 
\textit{eat}

Figure 2.24 = Figure 2.30 ∪ (Category (Infin-Verb) \\
\quad A \text{ Head-Proj Head-Proj Phonology (Eat)})

Figure 2.27 = Category (INFL) \\
\quad A \text{ Tense (Present)} \\
\quad A \text{ Specifier Category (Noun)} \\
\quad A \text{ Specifier Morph-Agr (}$\neg$\text{Third-Person-Singular)} \\
\quad A \text{ Head-Proj Category = Category} \\
\quad A \text{ Head-Proj Tense = Tense} \\
\quad A \text{ Head-Proj Head-Proj Category = Category} \\
\quad A \text{ Head-Proj Head-Proj Tense = Tense} \\
\quad A \text{ Specifier Antecedent =} \\
\quad \quad \text{Head-Proj Complement Specifier Same-Reference} \\
\quad A \text{ Specifier-Case = Specifier} \\
\quad A \text{ Head-Proj Complement (Figure 2.30 ∪} \\
\quad \quad \text{Category (Finite-Verb) A Head-Proj Head-Proj Phonology (eat)}) \\
\quad A \text{ Head-Proj Internal-Theta = Head-Proj Complement}

Figure 2.28 = Figure 2.30 ∪ (Category (Past-Participle-Verb) \\
\quad A \text{ Head-Proj Phonology (Eaten)})

Figure 2.29 = Category (INFL) \\
\quad A \text{ Tense (Past)} \\
\quad A \text{ Specifier Category (Noun)} \\
\quad A \text{ Head-Proj Category = Category} \\
\quad A \text{ Head-Proj Tense = Tense} \\
\quad A \text{ Head-Proj Head-Proj Category = Category} \\
\quad A \text{ Head-Proj Head-Proj Tense = Tense} \\
\quad A \text{ Specifier Antecedent =} \\
\quad \quad \text{Head-Proj Complement Specifier Same-Reference} \\
\quad A \text{ Specifier-Case = Specifier} \\
\quad A \text{ Head-Proj Complement (Figure 2.30 ∪} \\
\quad \quad \text{Category (Finite-Verb) A Head-Proj Head-Proj Phonology (eat)}) \\
\quad A \text{ Head-Proj Internal-Theta = Head-Proj Complement}

\textbf{Figure 2.31}
Figure 2.17

A FS representing a Generic One Case Verb

Category:  Verb
Head-Proj:  Category:
  Head-Proj:  Category:
Complement:  Category: Noun
Complement-Case: 

Figure 2.32
Our representation of the relationship between passive and active verbs in a GBUG-based GB analysis requires a slight variation on ordinary FS logic. We add a REPLACEMENT formula of the form

\[ X \rightarrow Y \]

to the inventory of possible FS logic formulas, with the interpretation that the unification of \( X \rightarrow Y \) with some feature value pair \( X' \) that unifies with \( X \) is \( Y \).\(^{25}\) Figure 2.35 is a set of logical formulas which may be used to change an active transitive verb to a passive verb. When Figure 2.35 is combined with some other set of logical formulas (a FS): (1) the External-Theta path and its value is replaced with the path Head-Proj Suppressed-Theta with the structure shared value of the optional
ComplementX Complement;\textsuperscript{26} and (2) the value of Head-Proj Complement-Case is changed to False.

For example, Figure 2.36 is the result of combining Figure 2.35 with Figure 2.30 to yield a FS representing a generic passive transitive verb.

### The Passive Rule

\begin{align*}
\text{External-Theta} & \rightarrow \text{Head-Proj Suppressed-Theta} \\
\Lambda \text{ Head-Proj Complement-Case} & \rightarrow \text{Head-Proj Complement-Case (False)} \\
\Lambda \text{ Category (Passive-Verb)} & \\
\Lambda \text{ Complement Category (NP-trace)} & \\
\Lambda \text{ Specifier Antecedent} &= \text{Complement Same-Reference} \\
\end{align*}

(Optional)

\begin{align*}
\text{ComplementX Category (Preposition)} & \\
\Lambda \text{ ComplementX Head-Proj Category} &= \text{ComplementX Category} \\
\Lambda \text{ ComplementX Complement} &= \text{Head-Proj Suppressed-Theta} \\
\Lambda \text{ Complement Complement-Case} &= \text{Head-Proj Suppressed-Theta} \\
\end{align*}

**Figure 2.35**

### A FS representing a Passive Transitive Verb

\begin{align*}
\text{Category:} & \underline{\text{Passive-Verb}} \\
\text{Specifier: Category: Noun} \\
\text{Antecedent:} & \underline{k} \\
\text{Head-Proj: Category:} & \underline{l} \\
\text{Head-Proj: Category:} & \underline{l} \\
\text{Complement1: Category:} & \underline{l} \\
\text{Same-Reference:} & \underline{k} \\
\text{Complement2: Category:} & \underline{n} \\
\text{Preposition} & \\
\text{Head-Proj: Category:} & \underline{n} \\
\text{Complement:} & \underline{m} \\
\text{Category: Noun} & \\
\text{Complement-Case:} & \underline{m} \\
\text{Internal-Theta:} & \underline{l} \\
\text{Suppressed-Theta:} & \underline{m} \\
\text{Complement-Case: FALSE} & \\
\end{align*}

**Figure 2.36**
Figure 2.35 is our representation of a Jaeggli (1986) analysis of passive. According to Jaeggli, passive is a morphological process with syntactic effects—it suppresses a theta role and absorbs a case. Each of these are destructive operations in that they prevent licensing relations instantiated in the active form from being instantiated in the passive (assuming that the active form is basic and the passive form is derived.) We represent each of these changes with replacement formulas. Suppression differs from absorption in that the former process is more absolute. When a case is absorbed, it can no longer be assigned—therefore the new value for the Complement-Case feature is False. When a theta role is suppressed, it can be optionally reassigned. Therefore we use the replacement formula to simply change the external-theta role to a Suppressed-Theta role, a term we derive from Grimshaw (1990). Following GB assumptions, the licensee of this suppressed-theta relation is the complement of the optional by-phrase. The rest of the rule is modeled as ordinary unification—we assume that the node labels Verb ⊆ Passive-Verb and NP-Trace ⊆ Noun. ComplementX in the rule is instantiated by Complement2 in Figure 2.36. We gloss over the problems caused by attempting to number multiple complements in this situation. We are proposing a concrete mechanism for representing case absorption and theta role suppression. Prior to this dissertation, concrete mechanisms for these processes were not formulated because data structures for representing theta roles and case were not employed. Linguists writing about case absorption and theta role suppression would describe these processes in terms of their input and output, but failed to show how the mechanisms actually worked.

Jaeggli’s approach to passive suggests that the following GL formulas are the basis of the passive rule and that the remaining formulas listed in Figure 2.35 follow from these:

(1) External-Theta → Head-Proj Suppressed-Theta (NIL)

(2) Head-Proj Complement-Case → Head-Proj Complement-Case (False)

In a Jaeggli-style approach: (1) the optional part of the passive rule follows from the definition of theta role suppression (formula 1); (2) the presence of the NP-trace and the identity of its antecedent follow from case absorption (formula 2) as well as case theory, trace theory and binding theory principles; and (3) the formula Category (Passive-Verb) is unnecessary because there is no passive construction per se, but rather that the properties of the passive follow from formulas 1 and 2. We are hesitant to
alter our passive rule so that it consists solely of these two formulas because the necessary principles and mechanisms required to determine all the syntactic properties of passive based on these formulas have not been sufficiently formalized. Figure 2.35 may therefore be considered a starting point which may be simplified after further investigation.

Unification and subsumption alone can capture most of the relationships between the syntactic structures associated with lexical items. However the relationship between passive and active verbs is one that cannot be captured in this representation of GB Theory without relying on at least one other mechanism: replacement. We expect that other mechanisms may also be necessary.\(^{30}\)

2.5.4 Summary

This section defined the unification operation and provided examples of its function in a GBUG-based grammar and lexicon. The fact that unification is defined in terms of subsumption has a number of consequences. All of our extensions to the subsumption relation discussed in Section 2.4 led to extensions in the functioning of the unification operation. As a result, GBUG allows unification of non-identical labels, as in SFG and unification-based frameworks which assume typed FSs, e.g. HPSG. The role of unification in the lexicon is based in part on the role of subsumption in the lexicon because any item in a subsumption hierarchy is the unification of all items subsuming it in that hierarchy. This provided us with a way of building lexical items from the sum of their characteristics (this has some obvious implications for building a computational lexicon.) Unification is also used to combine constituents into larger constituents and to instantiate licensing relations. This is useful for deriving a representation of a phrase from the some of its parts, e.g., in parsing.
2.6 Properties of Label Subsumption and Unification

2.6.1 Introduction

This section explores the complications to our FS logic GL which result from label subsumption and unification. GBUG defines two types of labels: arc labels and node labels and three types of label subsumption:

1. Defined subsumption relations like \( \text{Nominal} \subseteq \text{Noun} \);
2. Subsumption based on properties of disjunctive labels like \((\text{Noun} \mid \text{Gerund}) \subseteq \text{Noun}\); and
3. Subsumption relations involving negation like \(\neg \text{Third-Person-Singular} \subseteq \text{First-Person-Singular}\)

We show that each of these types of label subsumption complicate GL in the same way. We also show that label subsumption presents the same complications as disjunctive and negative feature values discussed in Karttunen (1984, 1986), Kasper and Rounds (1986, 1990), Nakazawa (1991), among others. These findings are partially based on the research conducted in SFG in connection with Johnson, Meyers and Moss (1993a,b).

2.6.2 Node Label Subsumption and Unification

Karttunen (1984, 1986) shows how disjunctive and negated Feature values complicate unification and recommends ways of dealing with these complications. Karttunen (1986, p.27) argues for viewing negative feature values as constraints on future unification, rather than as regular feature values. Intuitively, the unification of a negative value \(\neg X\) and a value \(Y\) is \(Y\), since \(Y\) is more specific. However, it is possible for later unifications of \(Y\) with other items to yield \(X\). This violates the spirit of combining the information of \(\neg X\) and \(Y\) together—the fact that \(Y\) may become \(X\) by unification indicates that the information instantiated as \(\neg X\) has been lost. Karttunen solves this problem by carrying along the negative value as a constraint on all future unifications.\(^{31}\)

Karttunen (1986, pp. 27-30) shows that unification of a disjunctive feature value is complicated by the fact that more than one disjunct may be unifiable with a given FS. This means that unification can actually complicate disjunction instead of simplifying (or eliminating) it. For example the
unification of \((A \mid B)\) and \((C \mid D)\) is \(((A \cup C) \mid (A \cup D) \mid (B \cup C) \mid (B \cup D))\). Furthermore, the problem of simplifying disjunctions is a well-known computationally complex problem.

Nakazawa (1991, p.46) points out that all disjunctive feature values are logically equivalent to a conjunction of negative values, assuming that each feature has a finite set of possible values. For example, the morphological feature person typically has only three possible values: 1st, 2nd and 3rd. Therefore, the feature values 1st \(\mid\) 2nd and \(\neg3rd\) are equivalent. Similarly, if we assume that the complete set of syntactic categories is \{Noun, Verb, Preposition, INFL, Adverb, Gerund, Complementizer, Degree\}, then the node label \(\neg\text{Verb}\) may be reinterpreted as the following disjunctive value of the Category feature:

\[
\text{Noun} \mid \text{Gerund} \mid \text{Preposition} \mid \text{INFL} \mid \text{Adverb} \mid \text{Complementizer} \mid \text{Degree}
\]

assuming that Verb and the categories listed in this disjunction exhaustively cover the possible syntactic categories for our domain, e.g., a grammar of English.

Assuming that Nominal \(\subseteq\) Gerund and Nominal \(\subseteq\) Noun are the only defined subsumption relations in which Nominal subsumes something else, then Nominal is equivalent to Noun \(\mid\) Gerund. In general, a label \(L\) which subsumes labels \(L_1 \cdots L_n\) is logically equivalent to \(L_1 \mid L_2 \mid \cdots L_n\).

Therefore, it appears our use of disjunctive node labels, negated node labels and node labels with defined subsumption relations may all be viewed as disjunctive node labels assuming that each feature has a finite set of possible values, i.e., each feature is WELL-TYPED in the sense of Carpenter (1992). See Carpenter (1992) and other work on typed FS logic for a detailed analysis of the properties of FSs constructed from well-typed features.

2.6.3 Arc Label Subsumption and Unification

This section shows that FSs containing features which enter into subsumption relations complicate the unification of complex FSs in two respects:

(1) Problem 1: Unification may result in two identical labels on arcs with the same source.

(2) Problem 2: Unification may have more than one result.
We show that Problem 1 is resolved by recognizing the equivalence of the unification relation and conjunction and that Problem 2 is a consequence of the equivalence of these special features and their values with disjunctions of FSs.\(^{32}\)

In a unification formalism in which arc labels only unify when equal, the unification of two functional FSs is always functional. However, by allowing two nonequal labels to unify, it is possible that given two distinct arc labels \(L_1\) and \(L_2\) in FS\(_i\) and two other distinct arc labels \(L_1'\) and \(L_2'\) in FS\(_j\), the unification of FS\(_i\) and FS\(_j\) will have two identical labels with the same source (we refer to this above as Problem 1). Problem 1 occurs if (1) \(L_1 \cup L_1'\) is the same as \(L_2 \cup L_2'\); or (2) There is no \(L_2'\) and \(L_1 \cup L_1'\) is the same as \(L_2\). For example the label \(X \mid Y \mid Z\) unifies with \(W \mid X\) to yield \(X\); and the label \(A \mid X \mid B\) unifies with \(D \mid X\) to yield \(X\). Notice that this can only happen if arc labels on two arcs with the same source are unifiable with each other. Thus, in order to prevent the production of ill-formed FSs, we must (1) impose a stricter version of functionality on FSs, requiring that arc labels on arcs with the same source cannot be unifiable (can never instantiate the same feature); (2) stipulate that unification fails if the output is not functional; or (3) require that all unifiable arcs with the same source are unified together—if unification fails, the FS is ill-formed. We assume choice (3), pending future research, since a complex FS is a conjunction of feature value pairs and unification is equivalent to conjunction.

Given that nonidentical labels can unify, it is possible for unification of two complex FSs to result in more than one distinct result depending on which arc is assumed to unify with which arc. Assume that some label \(L_1\) unifies with two or more different distinct labels \(L_2\) or \(L_3\). For example, if \(L_1\) equals \(X \mid Y\), \(L_2\) equals \(X\), and \(L_3\) equals \(Y\), then \(L_1\) unify \(L_2\) is \(X\) and \(L_1\) unify \(L_3\) is \(Y\). Let \(L_1\) label some arc \(A_1\) in FS\(_j\) and let \(L_2\) and \(L_3\) label two possibly corresponding arcs \(A_2\) and \(A_3\) in FS\(_j\). Depending on whether \(A_1\) unifies with \(A_2\) or \(A_1\) unifies with \(A_3\), two different FSs result from the unification of FS\(_j\) and FS\(_j\), one FS with an arc labeled \(X\) and another with an arc labeled \(Y\). There are two possible solutions to this problem:
(1) We can stipulate that unification fails if more than one result is possible, e.g., in the above example if $A_1$ can unify with either $A_2$ or $A_3$, then unification fails.

(2) We can assume a weaker definition of unification in which the output is the set of possible results, e.g., the above example would have two outputs: (a) a FS containing the arcs $A_1 \cup A_2$ and $A_3$; and (b) a FS containing the arcs $A_1 \cup A_3$ and $A_2$.

Pending future research, we choose solution (1) and reject solution (2) because: solution (2) has the potential of causing a combinatorial explosion; and we have not found any negative empirical consequences to this decision.

A feature value pair $F \cdot V$ is equivalent to a disjunction of feature value pairs with the same value if $F$ is disjunctive or if $F$ subsumes other features. For example, the logical formula

$$X \mid Y \cdot (V)$$

is logically equivalent to the formula

$$(X \cdot (V) \mid Y \cdot (V)).$$

This is the intended interpretation of a disjunctive label (no proof is necessary). Defined subsumption relations on labels are logically equivalent to disjunctive labels, as noted above. For example, the labels $\text{Theta}$ and $\text{Internal-Theta} \mid \text{External-Theta}$ are equivalent provided that $\text{Internal-Theta}$ and $\text{External-Theta}$ are the only labels subsumed by $\text{Theta}$.\textsuperscript{33}

Therefore, it appears that for all features $F$ which subsume other features, an instance of $F$ and its value is equivalent to the disjunction of feature value pairs, i.e., the disjunction of a particular type of FS. It is not surprising that the unification of FSs containing these special features have the two properties noted in this section. A disjunction of FSs unified with a single FS should yield a disjunction of FSs. Our reason for limiting the output of unification to a single result is based on our intended use of these features, not their logical properties.
2.6.4 Conclusion

This section has clarified how the presence of arc and node labels which can subsume other labels complicate unification. Following the literature on the logical properties of FSs, we show that each node label of this type is actually equivalent to a disjunction of node labels, assuming that the possible values of each feature is limited to a finite set, i.e., is well-typed. We have shown that each arc label of this type and its value is equivalent to a disjunction of feature value pairs, i.e., a disjunction of a type of FSs. Mindful of these complexities, we believe that a linguistically constrained grammar should not run into too many problems.

2.7 Summary

The formulation of GB analyses in terms of FSs, unification and subsumption, provides a language for making GB concepts more explicit than previously and for making it easier to compare GB based analyses with syntactic analyses using other unification based frameworks. This Chapter provides a brief overview of our unification-based model of GB theory. We introduce various formal terms based on previous work in unification-based grammar formalisms. We also demonstrate several ways these mechanisms are used to formalize GB concepts in the remainder of this thesis.

Above, we point out the limits of certain formal devices for expressing GB concepts and proposed ways of extending the formalism. We introduced SFG’s inverse paths, replacement formulas, and arc/node labels which subsume other labels. It is important that these extensions be carefully cataloged, so that their properties are open to further analysis. This makes our work falsifiable and makes it possible to analyze it in terms of mathematical complexity. We do not hesitate to use formal devices that more complex than standard unification. However, it is essential that the details of these devices be given—this is the only way that other researchers may improve them or show that other devices are superior. We find much previous work in GB theory to be deficient in this regard.
Appendix 2A - A GBUG Unification Algorithm

The definitions of FSs above define unification as a logical relation between three elements: one result and two input objects. While this is useful for analyzing the logical and formal properties of unification, we provide another type of description below: an algorithm for deriving the output of unification of two complex FSs. This algorithm represents unification as a combinatorial operation, rather than a relation. This kind of specification is also essential for implementing unification computationally. The following algorithmic specification from Karttunen (1986) describes how to derive the unification of two complex FSs, assuming that only identical labels are compatible with each other.

We provide a new specification below which corresponds to our definition of GBUG unification.

In trying to unify two structures, A and B, one proceeds as follows. If A and B contain the same attribute but have incompatible values for it, they cannot be unified. When A and B are merged, every attribute that appears only in one of the structures is included in (Unify A B) with its original value. If some attribute appears both in A and B, then the value of that attribute in (Unify A B) is the unification of the two values ... (Karttunen 1986, pp. 20-21)

Deriving FS_3, the GBUG unification of two complex FSs FS_1 and FS_2 requires the procedures 2.1, 2.2 and 2.3 (based on the unification algorithm for SFG described Johnson, Meyers and Moss 1993a,b). Unify(FS_1, FS_2) is implemented as Unify-nodes(r_1, r_2). These procedures unify both the input nodes/arcs and and all nodes/arcs below them in the DAG containing these nodes/arcs. We provide a simple algorithm for deriving the unification of two FSs. This algorithm is nondestructive in the sense that it does not alter its input. We abstract away from the differences between implementing unification destructively, nondestructively and various other ways. While genuine efficiency issues arise in choosing an implementation strategy, these issues do not alter the basic properties that make GBUG unification different from previous unification based approaches. We assume the following notation in Procedures 2.1 to 2.3: Label(A_i) is the arc label of A_i; and Unify(Label_1, Label_2) represents the unification of two labels.
**Procedure 2.1:** UNIFY-NODES($n, n'$)
1. Unify(Label($n$), Label($n'$))
2. Unify-Sets-of-Arcs(Out-Arcs($n$), Out-Arcs($n'$))

**Procedure 2.2:** UNIFY-ARCS($A, A'$)
1. Unify(Label($A$), Label($A'$))
2. Unify-Nodes(Target($A$), Target($A'$))

**Procedure 2.3:** UNIFY-SETS-OF-ARCS($Set_1, Set_2$)
$Set_1$ and $Set_2$ are sets of arcs. This procedure returns a set of arcs $Set_3$ if it succeeds. For each $A$ in $Set_1$, attempt to find some $A'$ in $Set_2$ such that Step 1 of Unify-Arcs($A, A'$) succeeds. There are three possibilities:
1. If no $A'$ is found, $A$ $\in$ $Set_3$
2. Unify-Arcs($A, A'$) $\in$ $Set_3$ if Unify-Arcs succeeds.
3. If Step 1 of Unify-Arcs($A, A'$) succeeds, but Step 2 fails, then the procedure fails.

Consider the abstract example in Figure 2.37, simple in that FSs $FS_1$ and $FS_2$ contain no unifiable nonidentical labels. Unify-nodes recursively calls Unify-Sets-of-Arcs. Features 5 and 6, along with their values are included in the unification of $FS_1$ and $FS_2$ by Procedure 2.3, Step 1. The instance of Feature 1 and its value $V$ in $FS_1$ unifies with a corresponding instance of Feature 1 and value $V'$ in $FS_2$ by Procedure 2.3, Step 2. Step 2 calls Procedure 2.2 which recursively calls Procedure 2.1 with the roots of $V$ and $V'$ as input.
FS₄ in Figure 2.38 fails to unify with FS₁ in Figure 2.37 because without some special stipulation that Atom 1 and Atom 7 both subsume some label, these labels cannot unify. Therefore Atom 1 ∪ Atom 7 fails and FS₄ ∪ FS₁ fails as per Procedure 2.3 Step 3. However, let's take the same examples and assert that Atom 1 ⊆ Atom 7 and Feature 8 ⊆ Feature 4. Then FS₁ ∪ FS₄ = FS₅ in Figure 2.38. The only difference between this unification and the previous example is the unification of nonidentical arc and node labels.
In Section 2.6.3, we noted that arc label subsumption complicates the unification of complex FSs in two respects:

1. Unification may result in two identical labels on arcs with the same source.
2. Unification may have more than one result.

In that section we proposed (at least) two solutions to these problems:

1. A requirement that all unifiable arcs with the same source are unified together. If unification fails, the FS is ill-formed.
2. Imposing a constraint causing unification to fail if there is more than one result.

Implementing these solutions requires that we modify Procedure 2.3, as follows:
Procedure 2.3 (revised): UNIFY-SETS-OF-ARCS(Set₁, Set₂)

Set₁ and Set₂ are sets of arcs. This procedure returns a set of arcs Set₃ if it succeeds. For each A in Set₁, attempt to find some A' in Set₂, such that Step 1 of Unify-Arcs(A, A') succeeds. There are three possibilities:

(1) If no A' is found, A ∈ Set₃
(2) A'' = Unify-Arcs(A, A') if Unify-Arcs succeeds. If the following three steps can be completed without failure, then A₃ ∈ Set₃.
   (a) If Step 1 of Unify-Arcs(A, B') succeeds for any B' in (Set₂ - A'), then, the procedure fails.
   (b) If Step 1 of Unify-Arcs(A', B) succeeds for any B in (Set₁ - A), then, the procedure fails.
   (c) For every arc A₄ ∈ Set₃, Unify-Arcs(A₃, A₄). There are three possibilities:
      (i) If Steps 1 fails, the procedure succeeds.
      (ii) If Step 1 succeeds, but Step 2 fails, this step fails.
      (iii) If Steps 1 and 2 succeed, remove A₄ and replace A₃ with A₃ ∪ A₄ and the procedure succeeds.
(3) If Step 1 of Unify-Arcs(A, A') succeeds, but Step 2 fails, then the procedure fails.

This appendix provides an overview of GBUG unification from a procedural or implementational point of view. It is mainly of interest for those interested in computational implementations of the theory.
Notes

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 descriptions* (Kay 1979, 1985) *functional structure* or f-structure in LFG (Bresnan 1982a,b), *categories* in GPSG (Gazdar, Klein, Pullum and Sag 1985), *terms* in Definite Clause Grammar (Pereira and Warren 1980), *attribute value structure* (Johnson 1988), and *S-graph* in SFG (Johnson and Moss 1993, 1994, Johnson, Meyers and Moss 1993a,b).

2Note that this Figure 2.1, like other FSs in this chapter, lacks a representation of bar levels of categories—the $\overline{X}$ theory representation of whether an object of category X is an independent phrase ($\overline{X}$) a terminal category ($X^0$ or X) or an intermediate nonterminal ($\overline{X}$). Chapter 3 provides means for determining the bar level of the syntactic category without explicitly representing it as a feature. However, the bar notation is a useful abbreviation for complex FSs. We therefore use $\overline{X}$ notation in this chapter to simplify diagrams, e.g., $\overline{N}$ in Figure 2.14 is an abbreviation for FS representing an $\overline{N}$. The feature value pair *Category: Noun* in Figure 2.1 is an underspecified representation of an $\overline{N}$. The value of the Complement feature in the actual lexical entry for *eat* is more fully specified. For example, an N$^0$ (such as the common noun *sandwich* with no determiner) cannot be the value of that Complement feature. Chapter 3 provides further details.

3We assume the VP internal subject analysis of Fukui and Speas (1986), Sportiche (1988), among others here. In this analysis the subject of the sentence originates as the specifier of the $\overline{V}$ and raises to the specifier of $\overline{I}$ to receive case. We discuss details of this analysis in later chapters. The crucial features to note are that the specifier of the $\overline{I}$ is assigned case and is coreferential with an NP-trace in the specifier position of the $\overline{V}$. Our notation for coreference is described in detail in Chapter 5. For now, simply note that the node sharing relation between the arc labeled *Antecedent* and the arc labeled *Same-Reference* indicates that the NP-trace and the $\overline{N}$ *John* are coreferential.
4Note that the theta role features are distinguished by type, e.g., External-Theta, Internal-Theta. When there are more than one internal theta role, these are distinguished by number, e.g., Internal-Theta1, Internal-Theta2. See Chapters 2 and 7, as well as Williams (1980) for an explication of the internal/external theta role distinction. In some GB analyses, theta roles have names like theme, agent, source and goal, originating in Gruber (1965). Others like Ravin (1976), Hoekstra (1984) and Levin and Rappaport (1986) maintain that the theta role names have no status in syntax. The use of numbered theta roles is a representation of the latter approach: the theta roles are distinct and thus they do not violate Kay’s functionality condition. However, we make no statements about the natural language semantic differences between theta roles in this dissertation.


6Hereafter, for ease of exposition, we use labels to represent both labels and the entities they represent, e.g., the term Feature1 is used to discuss both the label Feature1 and the feature represented by that label; and the term Atom1 is used to discuss both the label Atom1 and the atomic FS it represents. We take care so that the context makes it plain which use is intended.

7A TREE is a subtype of DAG which lacks node sharing, a constituent structure tree being one instance. Current GB phrase markers (PMs) representing the annotated surface structure of phrases are not, strictly speaking, trees. While they are often drawn as trees, the nodes of the trees also represent a variety of relations which could not possibly be represented as trees. Assuming that the tree-ness only represents the s-structure relations of type dominates, a series of other relations are not included in the tree (e.g., theta and case relations) or conveyed by means of annotation or coindexing. When PMs are represented as edge-labeled graphs, the edge-labels indicating the relation represented by the arcs, it becomes clear that the graph-theoretic object Tree is too constrained to be explicit representations of GB analyses of phrases.
The source is also called the tail of the arc. The target is also called the head of the arc. However, the term head is used in at least two modules of GB theory already. X theory refers to the head of a phrase. Trace theory refers to the head of a chain. We have adopted the terms source and target to avoid confusion which could result from multiple uses of the same term.

Some readers may misunderstand the tabular notation’s use of coindexing because previous GB representations use coindexing to indicate type identity. For example in

\[ [John, INFL, [sees himself,]]. \]

John and himself are of the same type in the sense that both refer to the same human being; and John and INFL are of the same type in that they are both Third-Person-Singular (they agree). In contrast the indices in Figures 2.1 and 2.2 have the same interpretation as the intersecting arcs in Figure 2.4—the same value (not just the same type of value) is shared by more than one feature.

Chapters 5 and 6 show that it is usually possible to reduce coindexing in GB theory to node sharing, once the coreference and agreement relations are made more explicit than current GB accounts of these phenomena. We draw on previous research in RG, HPSG and other frameworks for creation of this new means of representation.

Figures 2.5 and 2.6 do not contain disjunctions or negations of formulas. Figure i contains simple examples of each.

| Feature1 Feature2 (Atom1 | Atom2) |
| Feature1 Feature3 (¬Atom2) |

Figure i

Disjunction and negation of values is written using | for disjunction and ¬ for negation. Although Figure i only includes the disjunction and negation of atomic FSs, we allow more complex disjunctions and negations. FS logic provides a means for representing disjunctions and negations of all types of FS logic formulas as discussed in Moshier and Rounds (1987), Nakazawa (1991), among
others. In contrast, negations and disjunctions cannot be incorporated into the DAG model as easily.

11In this dissertation, we represent a constraint $X$ as a representation of the minimal FS which satisfies $X$. As pointed out by David Johnson (personal communication), our constraints are actually abbreviations for conditional statements. For example, Figure 2.11 is equivalent to the following logical statement:

$$\Theta\text{ Category (Noun | Gerund)} \rightarrow \Theta = \text{Case}$$

The fact that theta-marked phrases of these types MUST satisfy Figure 2.11 to meet the case filter, means that Figure 2.11 and other constraints discussed in this thesis are actually logical implications. However for purposes of exposition, we will, in effect represent implications as their right hand sides.

12We represent the Case Filter using FS Logic instead of a DAG data structure. The reason is that it is possible to extend FS Logic to a more complex logic which can describe constraints more complex than ordinary FSs. Therefore constraints of both complexities can be based on similar models, the more complex model being a superset of the FS model.

13In Johnson, Meyers and Moss (1993a,b) and some unpublished work, we propose multi-rooted graphs for lexical entries which license long distance dependencies. Each multi-rooted graph has one main root and one DANGLING ROOT. The subgraph rooted at the dangling root represents the extracted phrase. In this approach, only single-rooted graphs can represent complete phrases, i.e. well-formed parses. We propose special parsing devices to find where the dangling root is extracted from and create a single-rooted graph.

14Chapter 3 provides detailed definitions of $\bar{X}$ Theory terminology. For the purposes of this chapter, it is sufficient to assume the following definitions where $X$ is a variable over the syntactic categories N(oun), V(erb), P(reposition), etc. When a FS represents an $X^0$, $\bar{X}$ or $\bar{\bar{X}}$, it is representing a phrase of the generic category $X$. However, $X$ may also be substituted with the initial representing any syntactic category.

$X^0$ or $X$ is a terminal category.
\( \tilde{X} \) is a nonterminal constituent, a possible maximal projection of type \( X \). It can occur as an independent phrase, e.g., a complement, an answer to a question, etc.

\( \bar{X} \) is a nonterminal constituent of type \( X \), but is not an independent phrase, e.g., green book can be an \( \bar{N} \), but not an \( \bar{N} \) because it cannot as the complement of a verb, as the answer to a question, etc.

15 In Shieber’s notation: \( D(X) \) means the value of some feature \( X \) and \( \text{Dom}(X) \) means the domain of some \( \text{FS} X \), i.e. the set of feature value pairs in \( X \).

16 This note defines lattice structure, a term assumed in the following citation. A lattice structure is a way of formally characterizing a special sort of partial order and its domain. \( L \) is a lattice, where \( L = \{ D, \leq \} \), \( D \) is the domain of the partial order \( \leq \), if \( L \) has the following properties: (i) There exists another partial order \( \geq \) on \( D \) such that if \( X \leq Y \), then \( Y \leq X \); (ii) There is one element, the SUPREMUM (the sup\( (D) \)) such that sup\( (D) \) \( \leq \) \( X \) for each \( X \), a member of \( D \); (iii) There is one element, the INFIMUM (the inf\( (D) \)) such that inf\( (D) \) \( \geq \) \( X \) for each \( X \), a member of \( D \); For example \( L' \) is a lattice if, \( D' \) is the power set of \{a, b, c\} and \( \leq \) is the subset of relation, since \{a, b, c\} is a superset of every member of the power set of \{a, b, c\} and the empty set is a subset of every member of \( D \).

Two other operations: meet \( (\Lambda) \) and join \( (|) \) are defined on lattice structures such that \( (a \Lambda b) = \inf\{(a, b)\} \) and \( (a | b) = \sup(D) \). Combinations of elements connected by these operations follow certain rules, namely the rules on conjunction and disjunction in propositional logic. For details, see Partee, ter Meulen and Wall (1990, pp. 277-296) or a text on lattice theory.

The fact that both set theory and subsumption may be described in terms of lattices makes it easy to define subsumption (and unification) in terms of set theory. (Here and elsewhere \( \Lambda \) is the sign used for logical and, and \(|\) is the sign used for logical or.)

17 Properties of inclusive or include: (i) \( X \ | \ X = X \); (ii) \( X \ | \ \neg X \) is always true; and (iii) \( X \ | \ Y = Y \) iff \( Y \subseteq X \). Properties (i) and (ii) are standard in predicate calculus: (i) is the idempotent law for disjunction; and (ii) is a tautology. (iii) is derivable as a theorem from (ii). Informally: if \( Y \) subsumes \( X \), then \( Y = X \ | \ Z \), since \( Y \) includes all instances of \( X \) plus some set of other items which we are
representing as \( Z \); therefore \( X \mid Y = X \mid X \mid Z = X \mid Z = Y \), by substitution.

18 In general, allowing label subsumption and atomic FS subsumption complicate our FS logic in several ways (see Section 2.6).

19 We have not found the features \( \pm N \) and \( \pm V \) features of Chomsky (1970) to be useful in this regard. Stowell (1981) attempts to justify these features as natural classes, but fails in our view. For example, he formulates the CASE RESISTANCE PRINCIPLE which states that:

   (1) All \(-N\) categories (verbs, prepositions) may assign case and cannot be assigned case;

   (2) All \(+N\) categories (nouns adjectives) may be assigned case and cannot be assigned case.

Notice the use of the modal \( may \) in each of the above principles. In other words, there are some \(-N\) items like passives or unaccusative verbs like \( arrive \) which do not assign case and noun phrases in certain positions do not need case, e.g., the phrase \( a \ genius \) in

   (i) John graduated from nursery school a genius.

Thus, at best Stowell is showing that \( \pm N \) provides some idea as to what a canonical \(+N\) item and a canonical \(-N\) item acts like. However, this information is incorrect for the data noted above. Furthermore, the ability to assign case must still be marked on lexical items since it cannot be taken for granted that \(-N\) items assign case and the case filter determines when nouns and gerunds (not adjectives which are also \(+N\)) require case.

This is typical of Stowell’s arguments that \( \pm N \) and \( \pm V \) form natural classes. In some unpublished work (Meyers 1991), we provide a more complete refutation of Stowell’s arguments. However, a complete account is not required here. We are demonstrating how to organize a lexicon—we are not arguing for a particular organization.

20 The GPSG version of the HFC is actually more complex because it accounts for head conjunctions, cases in which there are more than one head of a constituent. In such cases, GPSG claims that the value of the feature \( F \) of a conjunction may be the generalization of the value of those \( F \) features for each of the conjoined heads (head daughters). The GPSG version also allows for certain features to be inherited from sources other than the head (FOOT FEATURES.) For our purposes, the version of
the HFC in Definition 2.7 is sufficient. However, we believe that future research probably requires a more complex version which is closer to the one found in GPSG.

21The requirement that there be no entity \( D \) such that \( D \) subsumes \( C \) but \( D \) does not subsume \( A \) or \( D \) does not subsume \( B \), is problematic for FSs containing negations or disjunctions. See Section 2.6 for details.

22The term *past participle* is often used to refer to the morphological form which heads the complement of perfective *have*, as well as the passive verb. We use this term to refer only to the former syntactic function. We use the term *passive form* to refer to the passive lexical entry which has the same phonology.

23Following the GB tradition, we could develop lexical entries for the various types of verbal inflection morphemes so that Figures 2.27 and 2.29 could be derived by combining these lexical entries with FS\(_{27}\) and FS\(_{29}\). However for expository purposes, we only represent lexical entries for independent words.

24The arguments presented in Chapter 6 eliminate NP-traces altogether. Assuming those arguments, the assertion \( \text{Noun} \subseteq \text{NP-trace} \) is unnecessary.

25Our replacement formulas are similar to PATR-II’s overwrite mechanism (cf. Shieber 1986, p.60) and nonmonotonic sort unification, especially the formulation presented in Young and Rounds (1993) (cf. Young and Rounds 1993 for references to other formulations of nonmonotonic sort unification). However, unlike these other mechanisms, our replacement formulas allow specific features to replace specific other features.

OVERWRITE is an operation on FS like unification, except rather than failing, all conflicts are resolved in favor of a designated argument to the overwrite relation.

Young and Round’s version of monotonic sort unification allows each value of a feature to be either STRICT or DEFAULT values, where a conflict between a strict value and a default value are resolved favor of the strict value.
We assume that there is some unspecified mechanism for distinguishing optional constituents (the $P$) from obligatory ones (see Chapter 8.)

In the GB tradition, parts of Figure 2.35 represent a lexical process and other parts are derivable from general principles. We call Figure 2.35 the passive rule, where the term rule may be understood as one possible compilation of lexical properties and general linguistic principles. This particular compilation is based on Jaeggli (1986).

GB principles are types of rules which apply very generally, e.g., principles of type X are used to derive rules of type X to account for the specific cases. The rule/principle distinction made in the GB literature, e.g. Chomsky (1986a) is really a matter of how general a rule applies or whether a rule is used to derive other rules.

There are a number of possible solutions—we only mention one here. Some linguists consider the passive by-phrase an adjunct. Others like Grimshaw (1990, pp. 109-110) note that the by-phrase has some complement-like traits (argument traits in Grimshaw’s terminology) and some adjunct-like traits. If we use the label Complement for the NP-trace object and Adjunct for the by-phrase, then Complement could unify with either Complement1 or Complement2 of verbs with two complements and the by-phrase would be unproblematic.

The passive rule represented in Figure 2.35 may apply to verbs like give in Mary gave John tea, which license two case-marked complements. In such instances either case may be absorbed. The case-less complement must be the one which becomes an NP-trace due to the properties of NP-traces. See Chapters 3 and 6 for further discussion about NP-traces.

One possible way of eliminating the need for the replacement formula may be to reject the assumption that passive lexical entries are derived from transitive lexical entries. The relationship between transitive and passive verbs could in principle be characterized in terms of some FS X such that X subsumes both transitive and passive verbs. For example, X could be Figure 2.33. A transitive entry would be the unification of Figures 2.32, 2.33 and 2.34. A passive entry would be the unification of Figures 2.33 and 2.35' below. Furthermore, maybe there would be some restriction much like a
feature co-occurrence restriction of the type used in GPSG.

(i) ¬(External-Theta (True) ∧ Head-Proj Suppressed-Theta (True))

<table>
<thead>
<tr>
<th>The Passive Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head-Proj Suppressed-Theta (NIL)</td>
</tr>
<tr>
<td>∧ Category (Passive-Verb)</td>
</tr>
<tr>
<td>∧ Specifier Antecedent = Complement Same-Reference</td>
</tr>
<tr>
<td>(Optional)</td>
</tr>
<tr>
<td>ComplementX Category (Preposition)</td>
</tr>
<tr>
<td>∧ ComplementX Head-Proj Category = ComplementX Category</td>
</tr>
<tr>
<td>∧ ComplementX Complement = Head-Proj Suppressed-Theta</td>
</tr>
</tbody>
</table>

**Figure 2.35’**

31 The difference in complexity between negative constraints and positive constraints is a key issue in Chapter 6 in our comparison of the possible syntactic constraints for determining the distribution of wanna contraction, the morphological process which contracts certain verbs (including want) and infinitival to. We argue that Occam’s Razor favors positive constraints—constraints represented by ordinary FSs, to negative constraints—constraints represented by logical formulas containing negation.

32 While we have disjunctive arc labels and labels with defined subsumption relations, we see no purpose in having negative arc labels. The Feature value pair ¬L(V), would be logically equivalent to L(¬V).

33 If we assert that Theta ⊆ Suppressed-Theta, then Theta is equivalent to Internal-Theta | External-Theta | Suppressed-Theta.
3.1 Introduction

This chapter provides a single formal mechanism for key relations from at least three modules of GB theory: $X$ theory, Case theory and Theta theory. We model the concept expressed by each of the following types of statements as a binary relation $X(A, B)$:

- $B$ is a specifier/adjunct/complement of $A$
- $A$ assigns a theta role to $B$
- $A$ assigns a case to $B$

In GBUG, $X(A, B)$ is represented as in Figure 3.1, where $A$ is the value of an $X$-licensor feature and $B$ is the value of an $X$-licensee feature. Pairs of $X$-licensor and $X$-licensee features represented by arcs with the same source have a special semantics: the values of these features are elements of the domain and range of the binary relation $X$ respectively. $X$ is called a LICENSING RELATION, $A$ an $X$-LICENSOR and $B$, an $X$-LICENSEE.

Figure 3.1

We sometimes refer to licensing relations $X(A, B)$ with statements like ‘$A$ assigns $X$ to $B$’ and ‘$A$ is the $X$ of $B$’, as this reflects the way these relations are talked about in the GB literature. When
discussing licensing relations in terms of assignment, we refer to \( X \) as a LICENSING PROPERTY,\(^2\) i.e., \( X(A,B) \) may be paraphrased as \( A \) assigns licensing property \( X \) to \( B \).

In Figure 2.10 (repeated below): (1) The feature Head-Proj is simultaneously a constituent-licensor, case-licensor and theta-licensor feature; (2) The feature Complement-Case is a Case-Licensee feature; (3) The features Internal-Theta and External-Theta are Theta-Licensee features; (4) The features Specifier and Complement are constituent-licensee features. Therefore, Figure 2.10 represents the following licensing relations:

1. The \( X^0 \) eat assigns a complement-case property, a complement property and an internal theta role property to the \( \bar{N} \) chocolate, i.e. Complement-Case(eat, chocolate), Internal-Theta(eat, chocolate) and Complement(eat, chocolate).

2. The \( \bar{X} \) eat chocolate assigns an external theta role property to an \( \bar{N} \), i.e. External-Theta(eat chocolate, \( \bar{N} \)).

3. The \( \bar{X} \) eat chocolate assigns a specifier property to that same \( \bar{N} \), i.e. Specifier(eat chocolate, \( \bar{N} \)).

![A FS representation of the \( \bar{V} \) eat chocolate](image)

**Figure 2.10**
Throughout this chapter, the reader may observe pairs of features which appear to be redundant. For example in Figure 2.10, the arcs labeled Complement-Case and Complement have the same source and target. Since each instance of Complement-Case will have the same source and target as some Complement arc, it is logical to assume that Complement-Case arcs should represent both a complement feature and a case feature, making the cooccurring complement arc redundant. In chapter 7, we do eliminate all redundant arcs of this type. However, the purpose of this chapter is to introduce the different types of licensing relations. For purposes of exposition, we leave many redundancies where they stand.

This chapter provides a clear semantics for different types of licensing relations, basic relations for which we use the syntax provided above. We also describe other GB relations which may be defined in terms of licensing relations. These derived relations include A-chains, V-chains and command relations. Clear formulations of these basic relations are crucial for the development of a lexicalized unification-based GB grammar that is both descriptively and explanatorily adequate.

3.2 Constituent Licensing and $X$ Theory

3.2.1 Introduction

In our version of $X$ theory: the terms $X^0$, $X$ and $\overline{X}$ denote sets of linguistic entities which satisfy constraints defined in this section; and the terms Specifier, Adjunct and Complement are constituent licensing relations. We show that our specification of $\overline{X}$ theory provides a clear and precise model for the GB analyses of phrase structure found in the literature.

In contrast previous GB accounts follow Chomsky (1970) in defining the terms Specifier, Adjunct, Complement, $X^0$, $X$ and $\overline{X}$ in terms of constituent structure tree configurations. We show that the constituent structure tree based accounts are inadequate in at least two respects: (1) previous specifications do not handle GB analyses of small clauses, relative clauses, head modifier constructions and multiple complement constructions, among others; and (2) If these definitions are expanded for descriptive adequacy, the resulting definitions overlap, making it impossible to unambiguously identify complements, adjuncts and specifiers from a constituent structure tree based representation of a
phrase.³

3.2.2 A Licensing Relation Based Version of $\bar{X}$ Theory

This section motivates a version of $\bar{X}$ theory in which the features Specifier, Adjunct and Complement are constituent licensee features and Head-Proj is a constituent licensor feature. Constituent licensing relations are represented as pairs of constituent licensor features and constituent licensee features, as shown in Figures 3.2 to 3.4. For example, Figure 3.2 represents the relation Specifier(H,S), where H is the value of the feature Head-Proj and S is the value of the feature Specifier. Alternatively, we say that H assigns the specifier property to S. Figures 3.2 to 3.4 each represent a nonterminal phrase $Z$ consisting of H and S. H is called the HEAD of $Z$ and $Z$ is called the IMMEDIATE PROJECTION of H. For any FS $Z$ containing a path P such that:

- the root of $Z$ is the source of P
- H is the target of P; and
- P consists solely of Head-Proj arcs,

$Z$ is a PROJECTION of H. $Z$ is an immediate projection of H if P consists of exactly one arc.

In Figure 3.4, we annotate the value of Head-Proj with $\langle \bar{X}, | X^0 \rangle$ to indicate that the head of that FS must satisfy the constraint $X^0$ represented by Figure 2.16 (repeated below) or the constraint $\bar{X}$ represented by Figure 3.5. The heads of the FSs in Figures 3.2 and 3.3 must be the same category as the FS as a whole, but are not restricted in any other way.

<table>
<thead>
<tr>
<th>A Nonterminal Containing a Specifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category: [ ]</td>
</tr>
<tr>
<td>Head-Proj Category: [ ]</td>
</tr>
<tr>
<td>Specifier:</td>
</tr>
</tbody>
</table>

Figure 3.2
In our version of $\chi$ theory, $X^0$, $\overline{X}$ and $\overline{\overline{X}}$ represent constraints on FSs. We say that each FS which satisfies a constraint $C$ is of type $C$. Thus each FS which satisfies $X^0$ is an $X^0$; each FS which satisfies $\overline{X}$ is an $\overline{X}$; and each FS which satisfies $\overline{\overline{X}}$ is an $\overline{\overline{X}}$. It is possible for a FS to satisfy more than one con-
restraint from the set \( \{X^0, X, \overline{X} \} \) and thus simultaneously be an \( X^0 \) and an \( \overline{X} \) or simultaneously an \( X \) and \( \overline{X} \).

For example, a FS representing the lexical entry for the proper noun \( John \) will be both an \( N^0 \) and an \( \overline{N} \) and the constituent \( big \ dogs \) is simultaneously an \( \overline{N} \) as in \( \text{the [big dogs]} \) and an \( N \) as in \( \text{I like [big dogs]} \). As shown below, it is impossible for a constituent to be simultaneously an \( X^0 \) and an \( \overline{X} \).

An \( X^0 \) is some FS that satisfies the constraint \( X^0 \) in Figure 2.16. For example, the FS in Figure 3.6 satisfies the constraint \( X^0 \) because Figure 3.6 has a TRUE value, for the feature Category and a FALSE value for the feature Head-Proj, where true and false values are FSs satisfying the atomic FSs True and False respectively. As discussed in Chapter 2, all FSs except Nil and False satisfy True and only Nil (or lack of any specification) satisfies False.

A FS representing the word chocolate which satisfies \( X^0 \)

- Category: Noun
- Phonology: chocolate

**Figure 3.6**

An \( \overline{X} \) is a FS that satisfies the constraint \( \overline{X} \) in Figure 3.5. Figures 3.7 and 3.8 both satisfy \( \overline{X} \) because: (a) Neither have a true value for the feature Specifier; (b) the value of Head-Proj in Figure 3.7 satisfies \( X^0 \); and (c) the value of Head-Proj in Figure 3.8 satisfies \( \overline{X} \).

A FS representing the phrase yummy chocolate which satisfies \( \overline{X} \)

- Adjunct: [i] Category: Adjective
  Phonology: yummy

- Modifier: [i] Category: Noun
  Phonology: chocolate

**Figure 3.7**
A FS representing the phrase *yummy dark chocolate* which satisfies $\bar{X}$

<table>
<thead>
<tr>
<th>Adjunct:</th>
<th>Category: Adjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonology:</td>
<td>yummy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modifier:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head-Proj: Adjunct:</td>
</tr>
<tr>
<td>Category: Adjective</td>
</tr>
<tr>
<td>Phonology: dark</td>
</tr>
<tr>
<td>Modifier:</td>
</tr>
<tr>
<td>Head-Proj: Category: Noun</td>
</tr>
<tr>
<td>Phonology: chocolate</td>
</tr>
</tbody>
</table>

**Figure 3.8**

The constraints $X^0$ and $\bar{X}$ conflict. $X^0$ contains the feature value pair $Head-Proj: False$ and $\bar{X}$ contains the feature value pair $Head-Proj: X^0 \mid \bar{X}$. Therefore a FS cannot simultaneously be an $X^0$ and an $\bar{X}$.

We define an $\bar{X}$ as a member of the set of FSs that satisfy a constraint called a COMPLETION CONSTRAINT. Completion constraints vary from category to category and language to language. An $\bar{X}$ is a nonterminal which can play the role of an independent phrase, i.e., it can be a nonhead daughter of another constituent (i.e., the value of a Specifier, Adjunct or Complement feature), $\bar{X}$ the answer to a question, etc. Below we give examples of completion constraints for different types of English $\bar{X}$s.

$\bar{V}$s and $\bar{I}$s must satisfy the constraint in Figure 3.9, i.e., they must contain specifiers. When the specifier is not overt, GB theory posits that an empty category (ec) occurs in specifier position, i.e., the value of the Specifier property is an ec. For example, the $\bar{I}$ in (3.1a) has an overt specifier *John*, but the $\bar{I}$ in (3.1b) has the empty category PRO as a specifier. The $\bar{V}$ small clause in (3.1c) has an overt specifier *John*, but the $\bar{V}$ in (3.1a) has the empty category NP-TRACE as a specifier.

(3.1) (a) $[\bar{I}_7\text{John} \rightarrow \text{INFL} \rightarrow \bar{V}_7\text{NP-trace} \rightarrow \bar{V}_7\text{saw Mary}]$

(b) Mary seems $[\bar{I}_7\text{PRO} \rightarrow \text{to see Mary}]$

(c) Mary made $[\bar{V}_7\text{John} \rightarrow \bar{V}_7\text{leave}]$
\( \overline{X} \)s of type *preposition, adjective, adverb* don’t need to satisfy any completion constraints at all. Words of this type in combination with all their complements (if any) are complete, as evident in (3.2a,d,f). Modifier and (non-predicative) complement phrases headed by prepositions, adjectives and adverbs can take exactly one optional specifier or adjunct modifier,\(^7\) as evidenced by (3.2b,e,g).\(^8\) The phrase with this specifier or adjunct can occur in all the same environments as the phrase without, except one: the \( \overline{X} \) containing the specifier/adjunct cannot occur inside another \( \overline{X} \) with an additional specifier/adjunct of this same type. In small clause constructions, as in (3.2c,h), phrases of type *Preposition* and *Adjective* have an \( \overline{N} \) specifier external to the \( \overline{X} \) (which can optionally include a specifier/adjunct) forming another larger \( \overline{X} \).\(^9\)

(3.2) (a) John slept \( \overline{X}_p \) [\( \overline{P}_p \) in \( \overline{N}_p \) his room]]

(b) John slept \( \overline{X}_p \) right \( \overline{P}_p \) in the middle of his road]]

(c) I want \( \overline{X}_p \) everyone \( \overline{P}_p \) instantly \( \overline{P}_p \) off my ship]]

(d) John ran \( \overline{Adv}_p \) quickly]

(e) John ran \( \overline{Adv}_p \) very \( \overline{Adv}_p \) quickly]]

(f) John bought the \( \overline{Adj}_p \) expensive] toothbrush.

(g) John bought the \( \overline{Adj}_p \) very \( \overline{Adj}_p \) expensive]] toothbrush.

(h) I consider \( \overline{Adj}_p \) [\( \overline{N}_p \) the toothbrush] \( \overline{Adj}_p \) very \( \overline{Adj}_p \) expensive]]

\( \overline{X} \)s must contain all mandatory complements to be complete. This fact need not be stated as part of a completion constraint because it follows from the following other factors:
(1) Requirements of lexical entries; and

(2) If a lexical item licenses complements, it must assign the Complement properties in a configuration satisfying the constraint in Figure 3.4. The value of Head-Proj in Figure 3.4 cannot be $\overline{X}$. Since $\overline{X}$s and $X^0$s cannot dominate $\overline{X}$ heads (see figures 2.16 and 3.8 above), the licensing property Complement can only be assigned inside of FSs representing $\overline{X}$s.

For example, the lexical entries for prepositions, transitive verbs, etc., specify that they require complements. The only way that both these lexical requirements and the Figure 3.4 constraint can be satisfied is if the complements occur inside the $\overline{X}$. Therefore, a completion constraint specifying that $\overline{X}$s contain all complements is redundant.

Based on some unpublished joint research by David Johnson and myself, conducted in an SFG-based version of Relational Grammar, we assume the following: (A) $\overline{N}$s in English must satisfy the completion constraint in Figure 3.10; (B) Quantification-Value, abbreviated Quant-Value, is a Head feature (see Chapter 2.); and (C) Indef $\subseteq$ Def (Indef and Def are abbreviations for Indefinite and Definite.), where Indef and Def are possible values for the feature Quant-Value.

\[\begin{array}{c}
\text{Completion Constraint} \\
\text{Quant-Value: True} \\
\text{Figure 3.10}
\end{array} \]

It follows from our account that terminals and nonterminals headed by different types of nouns can be $N^0$s, $\overline{N}$s and $\overline{N}$s according to where they appear in Figure 3.11. Examples (3.3) to (3.5) illustrate that this is indeed the correct distribution for English.
<table>
<thead>
<tr>
<th>Noun Type</th>
<th>N⁰</th>
<th>̅N</th>
<th>̅̅N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Proper Noun¹²,¹³</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Mass Noun</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Plur. Count Noun</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sing. Count Noun</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Specifier-less Nonterminal Headed by Mass Noun or Plur. Count Noun</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Specifier-less Nonterminal Headed by Sing. Count Noun</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Nonterminal with Specifier</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

(3.3) (a) \(\text{Mary bought the big book}\)

(b) *\(\text{Mary bought the big book}\)

(c) *\(\text{the big book fell off the shelf}\)

(3.4) (a) She completed the \(\text{examinations of her patients}\))

(b) She completed the \(\text{final examinations}\))

(c) Examinations should be conducted in private.

(d) \(\text{the (dirty) water spilled on the ground}\)

(e) \(\text{The (dirty) water spilled on the ground}\)

(3.5) (a) \(\text{Mary bought the (big) book}\)

(b) *\(\text{Mary bought the (big) book}\)

(c) \(\text{the (big) book fell off the shelf}\)

(d) *\(\text{the (big) book fell off the shelf}\)
(e) \[ \text{Mary conducted the (important) examination of John} \]

(f) *\[ \text{Mary conducted the (important) examination of John}. \]

(g) \[ \text{The (important) examination of John} \]

(h) *\[ \text{The (important) examination of John} \]

In Figures 3.12 to 3.19, only the FSs which satisfy the \( N \) completion constraint (Figure 3.10) are \( N \)-s; only those satisfying Figure 2.16 are \( N^0 \)-s; and only those satisfying Figure 3.5 are \( N \)-s. Figure 3.12 is an \( N^0 \), Figure 3.19 is an \( N \), Figures 3.17 and 3.18 are \( N \)-s, and Figures 3.13 to 3.16 are both \( N^0 \)-s and \( N \)-s.

<table>
<thead>
<tr>
<th>Lexical entry for</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>dog</strong></td>
</tr>
<tr>
<td>Category: Noun</td>
</tr>
<tr>
<td>Phonology: dog</td>
</tr>
<tr>
<td>Morph-Agr: Third-Person-Singular</td>
</tr>
</tbody>
</table>

**Figure 3.12**

<table>
<thead>
<tr>
<th>Lexical entry for</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>dogs</strong></td>
</tr>
<tr>
<td>Category: Noun</td>
</tr>
<tr>
<td>Phonology: dogs</td>
</tr>
<tr>
<td>Quant-Value: Indef</td>
</tr>
<tr>
<td>Morph-Agr: Plural</td>
</tr>
</tbody>
</table>

**Figure 3.13**
Lexical entry for

money

Category: Noun
Phonology: money
Quant-Value: Indef
Morph-Agr: Third-Person-Singular

Figure 3.14

Lexical entry for

John

Category: Noun
Phonology: John
Quant-Value: Def
Morph-Agr: Third-Person-Singular

Figure 3.15

Lexical entry for

he

Category: Noun
Phonology: he
Quant-Value: Def
Morph-Agr: Third-Person-Singular

Figure 3.16
Lexical entry for

*the*

Category: Noun
Quant-Value: Def
Morph-Agr: k
Specifier: Category: Determiner
Phonology: the
Quantifier: l
Head-Proj: Category: l
Quant-Value: l
Morph-Agr: k

Figure 3.17

---

Lexical entry for

*a*

Category: Noun
Quant-Value: Indef
Morph-Agr: k
Specifier: Category: Determiner
Phonology: a
Quantifier: l
Head-Proj: Category: l
Quant-Value: l
Morph-Agr: k

Figure 3.18
The $\bar{N}$ completion constraint requires all $\bar{N}$s to have a value satisfying True for the feature Quant-Value. According to this criterion: (1) lexical entries for plural count nouns, mass nouns, proper nouns, pronouns and determiners (Figures 3.13 to 3.18) represent $\bar{N}$; and (2) lexical entries for singular count nouns and adjectives (Figures 3.12 and 3.19) do not.

Quant-Value is a head feature in the sense of Chapter 2. Therefore the paths Quant-Value and Head-Proj Quant-Value structure share in all nonterminals of type noun, e.g., Figures 3.17 to 3.19. This fact has the following consequence:

If FS$_1$, FS$_2$, and FS$_3$ are three FSs of type noun, such that FS$_3$ is the result of unifying FS$_1$ with the value of the Head-Proj feature of FS$_2$, then: (1) the value of the feature Quant-Value in FS$_3$ is the unification of the values of Quant-Value in FS$_1$ and FS$_2$; and (2) FS$_3$ satisfies the $\bar{N}$ completion constraint only if FS$_1$ and/or FS$_2$ satisfies the $\bar{N}$ completion constraint.

For example, Figure 3.20 is derived by unifying Figure 3.12 with the value of Head-Proj in Figure 3.19; and Figure 3.21 is derived by unifying Figure 3.13 with the value of Head-Proj in Figure 3.19. Figure 3.13 has a value for Quant-Value satisfying True, but Figure 3.12 and Figure 3.19 do not have true values. Therefore Figure 3.21 is an $\bar{N}$, but Figure 3.20 is not. Figure 3.22 is the result of unifying Figure 3.12 with the value of the Head-Proj feature in Figure 3.18. Figure 3.23 is the result of unifying...
Figure 3.20 with the value of the *Head-Proj* feature of Figure 3.17. Figures 3.22 and 3.23 are Ns because Figures 3.17 and 3.18 are Ns.

**FS representing**

*big dog*

Category: Noun
Quant-Value: 1
Morph-Agr: Third-Person-Singular
Adjunct: Category: big

**Figure 3.20**

**FS representing**

*big dogs*

Category: Noun
Quant-Value: Indef
Morph-Agr: Plural
Adjunct: Category: big

**Figure 3.21**
Ns headed by plural count nouns (and mass nouns) optionally contain determiners, e.g., *the big dogs* and *the dogs* represented as Figures 3.23 and 3.24. The values of the Quant-Value features for plural nouns and determiners are *Indef* and *Def* respectively. We assert that *Indef* ⊆ *Def*. Therefore *Def* ∪ *Indef* = *Def* and the value of Quant-Value for any FS representing a plural noun with the
determiner *the is Def* as shown in Figures 3.23 and 3.24.

David Johnson (personal communication) observes that the account so far fails to rule out examples like *the big John’s dog* and *the big the dog*. Johnson proposes that the requirement that the lexical entry for *big* requires that its modified head be an $\overline{N}$. We take a similar approach in Section 3.3.7 in our discussion of the selectional properties of quantifiers.

We propose the FS in Figure 3.25 to represent the prototypical $\overline{N}$ anchored by its possessive $\overline{N}$ specifier. Figure 3.25 is an example of a PHRASAL ENTRY, a FS representing a phrase or type of phrase that is anchored by another phrase or a particular phrasal type. Figure 3.28 is the phrasal entry anchored by the specifier *the dog’s* derived by unifying Figure 3.27 with the value of the Specifier feature of Figure 3.25. Figure 3.27 is a FS representing the possessive $\overline{N}$ *the dog’s*, which is derived by unification of Figure 3.26 with the value of the Head-Proj feature of Figure 3.17.
Phrasal entry for Possessive $Ns$

Category: Noun
Quant-Value: j
Morph-Agr: k
Specifier: Category: Possessive Noun
Quant-Value: j
Head-Proj: Category: Noun
Quant-Value: j
Morph-Agr: k
Quantifier: l

Figure 3.25

Lexical entry for $dog's$

Category: Possessive Noun
Phonology: dog’s
Morph-Agr: Third-Person-Singular

Figure 3.26

FS representing $the
dog's$

Category: Possessive Noun
Quant-Value: Def
Morph-Agr: Third-Person-Singular
Specifiers: Category: Noun
Phonology: the
Head-Proj: Category: Noun
Phonology: dog’s
Quant-Value: j
Morph-Agr: k
Quantifier: l

Figure 3.27
Internally, possessive Ns are just like other Ns (Figure 3.27), but externally, possessive Ns are like determiners (Figure 3.28). Possessive Ns are like other Ns in that they must satisfy the completion constraint (Figure 3.10) (we assert that Noun $\subseteq$ Possessive-Noun.) The Quant-Value feature of the possessive form of some noun N is the same as N. For example, dog’s is unmarked for Quant-Value; dogs’ is Indef, John’s is Def and these nouns combine with determiners to yield the same results as the non-possessive nouns discussed above.

Figure 3.25, like a lexical entry for a particular type of phrase, specifies how one type of phrase anchors another. Figure 3.25 captures how possessive Ns are like determiners. A possessive N X, assigned a Specifier property by an N Y shares the value of its Quant-Value feature with Y. Like determiners, possessive Ns always have a value for the Quant-Value feature. Therefore nominal non-terminals with either possessive or determiners as specifiers are always Ns. The grammaticality of (3.6a,c,d) and ungrammaticality of (3.6b,e) are accounted for as follows: (1) Only Ns can be specifiers; and (2) Nonterminals headed by possessive nouns are only Ns if they satisfy the N completion con-
(3.6) (a) [[the/a dog’s] collar]
(b) *[[dog’s] collar]
(c) [[dogs’] collars]
(d) [[Fido’s] collar]
(e) *[[dog’s] collars]

If we assume that it is undesirable to posit phrasal entries and that we want to put all the information about possessives in the lexical entry for each possessive noun, a possible lexical entry for *John’s* is provided below in Figure 3.29. Possessive count nouns like *dog’s* or *dogs’* head arbitrarily long possessive $\bar{N}$s. A lexical entry like Figure 3.29 for these words would be problematic. We would have to allow for representations of an arbitrarily large number of modifiers and determiners to be inserted inside the value of the Specifier feature. In contrast, under the proposal that the phrase type *Possessive* $\bar{N}$ has a phrasal entry, Figure 3.29 is just a case of looking up the phrasal entry Figure 3.25 and inserting the possessive $\bar{N}$ *John’s* via unification. The same entry is looked up for the possessive $\bar{N}s$ *the dog’s*, *the giant green dog’s* and “[the man who is always walking his dog]’s”. Therefore, our analysis using Figure 3.25 is preferred. See section 3.3.8 for further constraints which will rule out, for example, *the giant green dog’s’s*’s. As pointed out by David Johnson, our present account unfortunately does not rule out *the man’s who is always walking his dog*. 
Below, we sketch an account of \(\bar{N}\)'s containing more than one determiner, e.g. (3.7), (3.8) and (3.9), which is compatible with the our basic approach. We refine this account in Section 3.3.7, where we introduce selection restrictions imposed by the quantifier on its head. The grammaticality and ungrammaticality of the (3.7) examples, follow from the analysis above. However, the (3.8) examples can be accounted for by refining the lexical entries of determiners in the manner proposed in (3.10) and the (3.9) examples are problematic for our definition of \(\bar{N}\) (as well as previous GB definitions of \(\bar{N}\)) due to the unusual properties of the determiner *only*. We show that *only* anchors a phrase that is neither an \(\bar{N}\), an \(\bar{N}\), nor an \(N^0\), e.g. *only (green) dog*. This analysis is based on the following facts: (1) the sequence *the only green dog* forms an \(\bar{N}\), but the sequence *only green dog* does not; and (2) *only* can follow or precede other determiners, e.g., *the only green dog*, *only the green dog*, but cannot follow prenominal modifiers, e.g., *the *green only dog*.

(3.7) (a) one book, three books, one examination of John,

three examinations of John

(b) the one book, the three books, the one examination of John,

the three examinations of John

(c) *a books, *one books, *three book, *an/one examinations of John,
*three examination of John

(3.8) (a) *the the dog, *a the dog, *a a dog, *a one dog, *these the dogs,

*three the dogs, *one the dog, *a one dog,

(b) *the a dog

(3.9) (a) the/an only dog, (the) only dogs, *only (green) dog

(b) *the/a green only dog, the only green dog, only the (green) dog

(c) only one dog, *one only dog, *the only one dog, the only three dogs,

only the three dogs

(3.10) (a) Determiners restrict the value of the Quant-Value feature of the nonterminal which they can combine with. For example, the lexical entry for the requires that only FSs satisfying the constraint Quant-Value: ¬Def can unify with the value of the Head-Proj feature of the lexical entry for the; numerals and the determiner a use the constraint Quant-Value: False.

(b) Some determiners, including the, require that the following word bear phonetic stress.

Some determiners, including a, can only occur in unstressed position.

(cf. Perlmutter 1970)

The analysis above accounts for why indefinite Ns with numbers as specifiers, e.g. (3.7a), may occur as part of definite Ns with the determiner the, as in (3.7b). The Def value of the Quant-Value feature for the unifies with the Indef value of the Quant-Value feature for the numbers one and three used as determiners, correctly predicting the grammaticality of (3.7a,b). The examples in (3.7c) are ruled out because the determiner and noun do not agree in number.

(3.10) refines our lexical entries in order to account for examples like (3.8). (3.10a) permits the relation Specifier(N, the) if N is an indefinite N⁰, N or N and permits the relation Specifier(N, a) if N does not have any value for the feature Quant-Value. However, the constraints described in (3.10a)
cannot be incorporated into the FSs representing lexical entries for *a and the (Figures 3.17 and 3.18), because the values of the structure sharing paths Quant-Value and Head-Proj Quant-Value would conflict in the resulting FSs, e.g., in the lexical entry for *the, the shared value would have to be simultaneously Def and ¬Def. Therefore (3.10a) must be viewed as constraints which are part of the lexical entries of determiners, but are external to the FS representations of the nonterminal anchored by the determiners.

The phrase *the one dog is grammatical, but (3.8b) is not because: (a) Only (phonetically) stressed determiners may occur following the in English; and (b) the determinant *a cannot be stressed. This is based on Perlmutter’s (1970) proposal that *a is actually the unstressed form of one, occurring in most unstressed syntactic positions occupied by numerals. (3.10b) could be incorporated into the FS portion of the lexical entries for the simply by providing a more articulated Phonology feature value. The value of the Phonology feature would be a complex feature structure representing various phonological properties of the word including its stress properties.

The quantifier only is like other determiners in that it must precede all adjectives and it may be preceded by some determiners, as shown in (3.9). However unlike other determiners, a phrase headed by a singular common noun and containing only, but no other determiners is not an $\bar{N}$, e.g., only (green) dog. If only is a specifier, then the nonterminal only dog cannot be either an $\bar{N}$ or an $\bar{N}$ by our definitions above. Only dog is not an $\bar{N}$ because it cannot occur as an independent phrase. Only dog cannot be an $\bar{N}$ if only is a specifier because that would violate the $\bar{X}$ constraint in Figure 3.5. One possibility is that only is an adjunct and that only dog is an $\bar{N}$. However, that proposal is difficult to maintain in light of only’s similarities with other determiners. In particular, we could not account for the ungrammaticality of the phrase *the green only dog if only dog was an $\bar{N}$. These facts are compatible with an analysis in which only is a specifier and only dog is a projection of N higher than $\bar{N}$, but lower than $\bar{N}$. The value of the path Quant-Value in the lexical entry for only must be some V such that Indef is subsumed by V. Furthermore, we must revise the $\bar{N}$ completion constraint as Figure 3.10’, i.e., an $\bar{N}$s must have a Quant-Value value which is subsumed by Indef.
Our account of the phrase structure of English nominals shows that there is a need for future research in this area. In particular, an examination of the interaction of a wider range of English determiners is necessary before an adequate $X$ based phrase structure can be devised. The intuitive distinction between an $N^0$ and an $\bar{N}$ is fairly clear. However, the discussion above shows that different classes of intermediate projections need to be distinguished. Determiners place idiosyncratic restrictions on their heads and it is not clear that the intermediate projections which satisfies one set of restrictions is necessarily of a higher level of projection than members of some other class of intermediate projections.

Adjectives and other prenominal modifiers require a phrase satisfying the constraint in Figure 3.5. Classifying a collocation like *only dog* as an $\bar{N}$, an $\bar{\bar{N}}$, or something else is less important than figuring out how these phrases are constrained, i.e., they are not $\bar{N}s$, they can be heads of phrases containing other determiners, they cannot be preceded by prenominal modifiers, etc. In our account $\bar{N}$ and $\bar{\bar{N}}$ are convenient abbreviations for FSs which satisfy certain constraints. The fact that *only dog* is a constituent of type *noun* which is not a member of class of $N^0$s, $\bar{N}s$ or $\bar{\bar{N}}$s is unproblematic. However, previous GB accounts would have to assume *only dog* is an $\bar{N}$ or an $\bar{\bar{N}}$, or alternatively, a new level of projection between $\bar{N}$ and $\bar{\bar{N}}$.

To sum up, in our version of $\bar{X}$ theory, *Specifier, Adjunct and Complement* are defined as constituent licensing relations between two daughters of a phrases, the head (the value of the feature *Head-Prof*) being the constituent licensor and the other daughter being the constituent licensee. $X^0$, $\bar{X}$ and $\bar{\bar{X}}$, each represent some set of categories which satisfy particular constraints. The constraints $X^0$ and $\bar{X}$ are fairly straightforward, but the constraint $\bar{\bar{X}}$ actually represents a set of completion constraints. A dif-

Our account of the phrase structure of English nominals shows that there is a need for future research in this area. In particular, an examination of the interaction of a wider range of English determiners is necessary before an adequate $X$ based phrase structure can be devised. The intuitive distinction between an $N^0$ and an $\bar{N}$ is fairly clear. However, the discussion above shows that different classes of intermediate projections need to be distinguished. Determiners place idiosyncratic restrictions on their heads and it is not clear that the intermediate projections which satisfies one set of restrictions is necessarily of a higher level of projection than members of some other class of intermediate projections. Adjectives and other prenominal modifiers require a phrase satisfying the constraint in Figure 3.5. Classifying a collocation like *only dog* as an $\bar{N}$, an $\bar{\bar{N}}$, or something else is less important than figuring out how these phrases are constrained, i.e., they are not $\bar{N}s$, they can be heads of phrases containing other determiners, they cannot be preceded by prenominal modifiers, etc. In our account $\bar{N}$ and $\bar{\bar{N}}$ are convenient abbreviations for FSs which satisfy certain constraints. The fact that *only dog* is a constituent of type *noun* which is not a member of class of $N^0$s, $\bar{N}s$ or $\bar{\bar{N}}$s is unproblematic. However, previous GB accounts would have to assume *only dog* is an $\bar{N}$ or an $\bar{\bar{N}}$, or alternatively, a new level of projection between $\bar{N}$ and $\bar{\bar{N}}$.
ferent completion constraint may be appropriate for each syntactic category within each language. The completion constraint for English $N$ is the most complicated case we investigated. We found a completion constraint which accounted for a large subset of English nonterminals of type noun. However, we found that some nonterminal nominals in English cannot be classified as either $Ns$ or $Ns$.

3.2.3 The Previous GB Version of $\bar{X}$ Theory

In previous GB accounts of phrase structure, definitions of of $\bar{X}$ terminology including $X^0$, $\bar{X}$, $\bar{X}$, specifier, adjunct and complement are syntactically defined as the terminal and nonterminal constituents found in a small set of $\bar{X}$ rules, schemata for phrase structure rules. We show that similar definitions of these terms would be inconsistent if based on the larger set of $\bar{X}$ schemata required for GB analyses of small clauses, relative clauses, head modifier constructions, multiple complement constructions, among others. In this larger set of $\bar{X}$ rules, either an $\bar{X}$ or an $\bar{X}$ can be broken down into an adjunct and an $\bar{X}$; and specifiers, complements and adjuncts can each be in the syntactic position: sister to $\bar{X}$/daughter to $\bar{X}$. Therefore, given an arbitrary constituent in a particular phrase structure position or phrase structure rule, it will not always be possible to unambiguously identify that constituent as an $\bar{X}$ or $\bar{X}$; and as a complement, specifier or adjunct. The version of $\bar{X}$ theory theory proposed in the previous section is preferred because it provides a coherent basis for GB analyses.

$\bar{X}$ rules in the sense of Chomsky (1970) are schemata for context free phrase structure rules. Under current assumptions, these rules are unordered. Context free phrase structure rules (CFRs) are rules of the general form:

\[(3.11) \ X \rightarrow Y \ Z\]

Y and Z are either nonterminal or terminal symbols and X is a nonterminal. In a CFR, the constituent represented by the symbol on the left hand side (LHS) may be broken down into the set of constituents represented by the symbols on the right hand side (RHS). Unordered CFRs do not constrain word order. Word order is determined by other principles (cf. Chapter 4.)

In GB, $\bar{X}$ schemata define a class of CFRs with the following properties:
The symbol on the right hand side (RHS) of a rule obeying X theory is of the same category as one of the symbols on the left hand side (LHS) called the head of the phrase. Bar levels provide a way of encoding that terminal and nonterminal symbols are of the same category or that two distinct nonterminal categories are of the same category.

The bar level of the head is less than or equal to the bar level of the RHS symbol. The bar level of \(\bar{X}^0\) is 2; the bar level of \(\bar{X}\) is 1; and the bar level of \(X\) or \(X^0\) is 0.

The symbols \(X^0\), \(\bar{X}\) and \(\bar{X}\) have the following interpretation: \(X^0\) is a terminal category representing a single word or morpheme; \(\bar{X}\) is a nonterminal constituent which does not constitute a complete phrase; and \(\bar{X}\) is a complete phrase.

Other properties are often attributed to \(X\) rules which do not apply to GB’s version of X theory. For details, see Jackendoff (1977), Kornai (1983) and Pullum (1985).

Figure 3.30 is a list of three X schemata taken from the GB literature. The terms *specifier*, *complement* and *adjunct* refer to nonhead constituents. (Ordered versions of) Rules 1 and 3 originate with Chomsky (1970) and Rule 2 is found in Baker (1978), Hornstein and Lightfoot (1981) and Radford (1988, pp. 196-218), among others. Examples of phrases generated from these rules are provided in (3.12) to (3.14). (3.12) is derived from the rule \(\bar{I} \rightarrow \bar{I}\) Specifier, (3.13) is derived from the rule \(\bar{N} \rightarrow \bar{N}\) Adjunct, and (3.14) is derived from the rule \(\bar{V} \rightarrow V^0\) Complement.

(3.12) \(\[\bar{I} \rightarrow [N\bar{I}\{\text{the man}\} [T\bar{I}\{\text{walked home}\}]\]

(3.13) \(\text{the } [N_{\text{adj}} [\text{very friendly}\} [N_{\text{N}} \{\text{professor of linguistics}\}]\]

(3.14) \(\text{John } [V_saw [N_{\text{N}} \{\text{the man}\}]\]

103
If, counterfactually, we assume that Figure 3.30 lists the complete set of schemata for all phrase structure rules assumed in GB theory, we would have clear unambiguous criteria for: (1) distinguishing between X^0s, Xs and X̅s; and (2) distinguishing between specifiers, complements and adjuncts. We could uniquely define the terms X^0, X and X̅ in terms of Figure 3.30, as follows: (1) An X^0 is a terminal (it cannot be subdivided); (2) An X consists of an X^0 and a Complement, or an X̅ and an Adjunct; and (3) An X̅ consists of an X and a Specifier. As shown in Figure 3.31, we could provide definitions of specifier, complement and adjunct in terms of the following relations:

**DAUGHTER-OF** — Daughter-of(X,Y) 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 arc A from the root of Y to X (or the root of X).

**SISTER-OF** — Sister-of(X,Y) if Daughter-of(X,Z) and Daughter-of(Y,Z).

The rules in Figure 3.30 and 3.32 would be equivalent because each Y̅ in Figure 3.32 would be uniquely identifiable from Figure 3.31 as a specifier, adjunct or complement.

<table>
<thead>
<tr>
<th>Property</th>
<th>Sister-of</th>
<th>Daughter-of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifier</td>
<td>X̅</td>
<td>X</td>
</tr>
<tr>
<td>Adjunct</td>
<td>X̅</td>
<td>X̅</td>
</tr>
<tr>
<td>Complement</td>
<td>X^0</td>
<td>X</td>
</tr>
</tbody>
</table>

**Figure 3.31**
Figure 3.33 lists the $\bar{X}$ schemata required for the GB analyses in (3.12) to (3.21). We show that a crucial consequence of assuming this larger range of GB analyses is that the definitions for *Specifier*, *Adjunct* and *Complement* cannot be uniquely defined in terms of phrase structure rules, the daughter-of relation and the sister-of relation. In Figure 3.33, *Specifier*, *Adjunct* and *Complement* are licensee features as in our licensing relation based version of $\bar{X}$ theory, e.g., Rules (1), (2) and (3) are roughly equivalent to the rules in Figure 3.30, the schemata required to account for (3.12) to (3.14). The small clause analyses in (3.15) require rule (4). The relative clause and $\bar{P}$ modifiers in (3.16) to (3.17) require rule (5). Binary branching analyses of phrases headed by verbs with multiple complements require rule (6) as in (3.18). Nouns with multiple specifiers like the examples in (3.19) require either rule (4) or rule (7) depending on whether N-x is an $\bar{N}$ or an $\bar{N}_x$. \[ \begin{align*}
1. \bar{X} & \rightarrow \bar{X} \bar{Y} \\
2. \bar{X} & \rightarrow \bar{X} \bar{Y} \\
3. \bar{X} & \rightarrow \bar{X}^0 \bar{Y}
\end{align*} \]

**Figure 3.32**

(3.15) (a) I saw $\overleftarrow{\bar{A}}_{\bar{A}_d} \bar{X} \bar{Y}$

(b) I consider $\overleftarrow{\bar{A}} \bar{X} \bar{Y}$

(c) He made $\overleftarrow{\bar{A}} \bar{X} \bar{Y}$

(3.16) $\overleftarrow{\bar{N}} \overleftarrow{\bar{N}} \overleftarrow{\bar{P}}$ the book $I \overleftarrow{\bar{C}}$ that John bought $t$.

(3.17) $\overleftarrow{\bar{N}} \overleftarrow{\bar{N}} \overleftarrow{\bar{P}}$ the book $\overleftarrow{\bar{P}}$ about John

(3.18) $\overleftarrow{\bar{P}}$ John $\overleftarrow{\bar{V}}$ past $\overleftarrow{\bar{V}} \overleftarrow{\bar{V}}$ give $\overleftarrow{\bar{V}}$ Mary $\overleftarrow{\bar{V}}$ a book]

(3.19) (a) $\overleftarrow{\bar{N}}$ the $\overleftarrow{\bar{N}}_x$ only $\overleftarrow{\bar{N}}_x$ three $\overleftarrow{\bar{N}}$ green books]
(b) \( \exists_{N} \) only \( \exists_{N-x} \) the \( \exists_{N-x} \) three \( \exists_{N} \) green books]]]]

(c) \( \exists_{N} \) the/one \( \exists_{N} \) green book]]

(3.20) (a) She completed the \( \exists_{N} \) the \( \exists_{N} \) examinations \( \exists_{p} \) of her patients]]]

(b) She completed \( \exists_{N} \) the \( \exists_{N} \) final \( \exists_{N} \) examinations]]]

(c) \( \exists_{N} \) Examinations] should be conducted in private.

(3.21) \( \exists_{N} \) [\( \exists_{N} \) Complicated \( \exists_{N} \) medical examinations]] upset most people.]

Specifiers, adjuncts and complements can be optional. As a result \( X_{0} \)s and \( X^{0} \)s can function as \( \overline{X} \)s; and \( X^{0} \)s can function as \( \overline{X} \)s. For example the lexical item examinations is an \( N^{0} \), but in (3.20a,b,c) it functions like an \( N^{0} \), \( N \) and \( \overline{N} \) respectively according to the \( X \) rules in Figure 3.33. Complicated medical examinations is an \( \overline{N} \) derived from two applications of Figure 3.33 rule (2). In (3.21), complicated medical examinations acts like an \( \overline{N} \) with respect to Figure 3.33 rule (1). We follow the GB accounts, e.g. Baltin (1989), which assume that all \( \overline{X} \)s in the LHS of the rules in 3.33 can be optionally replaced
by $\overline{X}$s or $X^0$s and all $\overline{X}$s can be optionally replaced by $X^0$s. An alternative way of dealing with optional phrases is to allow rules of the form $\overline{X} \rightarrow X; \overline{X} \rightarrow X^0$ and $\overline{X} \rightarrow X^0$. However, generating these non-branching nodes would serve no descriptive function, since they are ignored for purposes of c-command, m-command, government, etc. The nonbranching structure would be generated without theoretical motivation other than the ad hoc maintenance of multiple bar levels of phrase structure.

Given this range of overlapping definitions, the terms specifier, adjunct and complement cannot be defined unambiguously based on the syntax of phrase structure trees. A phrase $P$, sister to $\overline{X}$ and daughter to $\overline{X}$ can be a specifier like the man in (3.12), an adjunct like complicated in (3.21), or a complement like a book in (3.18). A phrase which is a sister to an $\overline{X}$ and a daughter to an $\overline{X}$ can be either an adjunct like the relative clauses and $\overline{P}$ adjuncts in (3.16) and (3.17) or a specifier like the subject of the small clauses in (3.15) or the outermost determiners in (3.19). Many $\overline{X}$s, $\overline{X}$s and $X^0$s are made up of the same set of constituents given the considerations above, e.g., water can be an $X^0$, $\overline{X}$ or $\overline{X}^0$.23,24

This section showed that the defining properties of $X^0$, $\overline{X}$, $\overline{X}$, complement, specifier and adjunct cannot be viewed as the phrase structure environment in which they occur. As pointed out in Pullum (1985), previous GB accounts have not specified the details of the version of $\overline{X}$ theory they assume apart from the $\overline{X}$ rules listed above. Therefore, no GB principles would have to be sacrificed by adopting constraint-based definitions. However, maintaining the phrase-structure-based definitions of $\overline{X}$ theory would result in the sacrifice of crucial GB analyses of small clauses, relative clauses, head modifier constructions, multiple complement constructions among others.

### 3.2.4 Summary

We have examined two versions of $\overline{X}$ theory: our FS based version and a phrase structure tree based version. We showed that our version of $\overline{X}$ theory may provide a basis for a descriptively and explanatory adequate theory of phrase structure. In contrast, the phase structure tree based version has never been sufficiently worked out for many crucial GB analyses. We showed that the phrase structure based version of $\overline{X}$ theory becomes inconsistent when extended to cover these analyses.
Our \( \bar{X} \) theory is based on subtypes of one general schema for licensing relations (Figure 3.1) in contrast with the three or more schemata needed for phrase structure tree based versions of \( \bar{X} \) theory. Our licensing schema covers \( \bar{X} \) relations as well as other types of relations discussed in later sections. Therefore our account is not only more descriptively adequate than previous GB accounts of \( \bar{X} \) theory, but it is also based on a more general mechanism.

### 3.3 Predicate Licensing

#### 3.3.1 Introduction

This section provides a GBUG-based account of predicate licensing relations, the syntacto-semantic relations between predicates and their arguments. We propose that if an element \( P \) imposes a natural language semantic restriction \( S \) on an element \( A \), this is sufficient evidence that there exists a relation \( \text{Predicate}(P,A) \). \( S \) is called a SELECTION RESTRICTION. Informally, \( \text{Predicate}(P,A) \) may be stated as "\( P \) assigns a predicate role to \( A \)" or "\( P \) is predicated of \( A \)". Predicate relations include: theta (the assignment of a theta role), modification (the assignment of a modification role), degree (the assignment of a degree role), and quantification (the assignment of a quantification role). These types of predicate roles are assigned by heads, modifiers, degree words and determiners/quantifiers.

We base our account of all types of predication on Ravin (1976), Hoekstra (1984), Levin and Rappaport (1986), among other GB accounts of theta role assignment which assume that the content of the theta relation has no status in syntax. In these accounts, it suffices to show that a theta relation exists and it is unnecessary for a syntactic description to specify the type of theta relation, e.g. agent, patient theme, etc. One problem with accounts that identify theta roles by name, is that it is not clear whether or not there is a finite number of different theta roles. Another problem is that distinctions between types of theta roles is not always clear in particular cases, e.g., is Mary in a patient or theme relation in the following sentence?

\[ \text{Bernice pushed Mary onto the plane} \]

The selection restrictions a head \( H \) imposes on some phrase \( X \) are key pieces of evidence that a relation \( \text{Theta}(H,X) \) exists. The meaning of this relation (e.g. the kind of information that the labels
agent, patient, theme, etc.) are left to the field of natural language semantics. For example, the requirement that the referent of the object of eat be physically possible to eat (i.e., not an idea) and the requirement that the subject of eat be animate are evidence that the relations Theta(eat, apple) and Theta(eat, Mary) are part of the syntactic description of the sentence Mary ate an apple. However, we do not provide specific labels which distinguish the meaning of the eater relation from the eaten relation.

We generalize this approach to other types of predication. For example in our account, the requirement that the modifier of pregnant be a female animal is evidence that the relation Modification(pregnant, mare) is part of the syntactic description of the N the pregnant mare. Our account is preferred on grounds of descriptive adequacy to previous GB literature which only characterize one type of predication: the theta relation.

In our lexicon, a FS representing a word W or its projection is always the predicate licensor of all FSs anchored by W. Therefore if W anchors a FS F, W or projections of W select all other FSs in F. For example, the adjective big anchors the FS in Figure 3.19 above. Big anchors a nominal in which big is the modifier and adjunct and some element of type noun is the head of the phrase and the modifier. Big requires the head of this phrase to refer to a type of thing that can vary in its dimensions, e.g., the big inch violates this restriction. Therefore big, imposes selection restrictions on its modifier and also determines the syntactic environment in which modification occurs. This approach is motivated by previous research in SFG and Categorial Grammar. In each predicate licensing relation R(licensor, licensee), our choice of licensor corresponds to to the lexical anchor of an Sgraph in SFG (see Johnson, Meyers and Moss 1993a,b) and the FUNCTOR of the phrase in a Categorial Grammar analysis, e.g., in categorial grammar, an adjective is a category of type N/N, which combines with a category of type N (a noun) to form a category of type N.25,26

3.3.2 Selection Restrictions

Our analysis of predication depends on the assumption that if A imposes a selection restriction on B (A selects B), then the relation Predicate(A, B) holds. This section provides the basis for making
the determination that A selects B.

Following the approaches of McCawley (1968a, 1968b, 1970) (attributed to Fillmore) and G. 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 presuppositions 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 *N* the pregnant mare and the pregnant ophthalmologist are compatible with our world knowledge. The *N* the pregnant bull is incompatible with our world knowledge. Therefore bull, violates the selection restrictions associated with the adjective pregnant, but mare and ophthalmologist satisfy these restrictions.

This approach explains the following facts which we use to distinguish a situation in which A selects B from a situation in which B selects A:

SELECTION RULE 1: Given a predicate 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.

SELECTION RULE 2: In the context of a belief, a joke, a riddle, a fantasy, a dream, a folktale, a science fiction story, or a contrived situation, phrases violating selection restrictions under normal assumptions are well-formed. For example, in the context of a science fiction book about a planet where male animals can be pregnant, the phrase the pregnant bull is well-formed. We call readings of phrases which rely on one of these fantastic or obscure settings to make sense FORCED READINGS of these phrases.

Selection Rule 1 is useful for determining whether A selects B, if A restricts the possible referents of B by means of selection restrictions. Selection Rule 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 of (3.22) and as animals when given as a complement of verbs in Verb-class-B of (3.22), as evidenced in (3.23a,b). Pronouns have but unspecified referents. In (3.23a) and (3.23b), the selection restrictions imposed by the verbs on their pronoun complements limit the possible referents of these pronouns. The pronouns in (3.23a) must refer to animals and the pronouns in (3.23b) must refer to texts. (3.23a) and (3.23b) provide evidence that the verbs select their complements because the refinement of reference in those example can only be attributed to the verb. In (3.23c), 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.27,28

(3.22) (a) Verb-class-A = \{read, write, peruse, edited, \ldots \}

(b) Verb-class-B = \{please, tease, tickle, annoy\}

(3.23) (a) The child teased it/something.

(b) Gertrude read it/something.

(c) They did something to the book/the child.

In a \(\overrightarrow{V}\) consisting of one verb \(V\) and one \(\overrightarrow{N}\) complement \(N\), forced readings in the sense of Selection Rule 2 allow our real world assumptions about \(N\) to change, but not our real world assumptions about \(V\). The sentences in (3.25) contain well-formed \(\overrightarrow{Vs}\) and the sentences in (3.26) contain \(\overrightarrow{Vs}\) which are incompatible with selection restrictions. This type of ill-formedness, which we indicate with a percent sign, is called SEMANTIC ANOMALY.29 In a fantasy world consistent with (3.26) being well-formed, our assumptions about the \(\overrightarrow{Ns}\) change, not our assumptions about the verbs. For example, in a fantasy world where all written material is personified, it is possible for a book to be teased or annoyed. In a fantasy world where people and animals are made up of text, a child may be edited. Based on Selection Rule 2, this shows that the verbs select their complements. 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 Selection Rule 2 supports the hypothesis that verbs select their complements, but not the hypothesis that complements select head verbs.\(^{30}\)

\[(3.24) \text{(a)} \quad \bar{N}_{-\text{class-A}} = \{\text{the book, the magazine, the screenplay, \ldots} \}
\]

\[(3.24) \text{(b)} \quad \bar{N}_{-\text{class-B}} = \{\text{Gertrude, Cecil, the child, the young acrobat, \ldots} \}
\]

\[(3.25) \text{(a)} \quad \text{The play}_{i} \left[ v_{1} t_{1} \right. \text{pleased/annoyed Gertrude/Cecil/the child/the young acrobat} \]

\[(3.25) \text{(b)} \quad \text{The babysitter}_{i} \left[ v_{1} t_{1} \right. \text{tickled/teased Gertrude/Cecil/the child/the young acrobat} \]

\[(3.25) \text{(c)} \quad \text{The expert}_{i} \left[ v_{1} t_{1} \right. \text{read/wrote/perused/edited the book/the magazine/the screenplay} \]

\[(3.26) \text{(a)} \quad \%\text{The play}_{i} \left[ v_{1} t_{1} \right. \text{pleased/annoyed the book/the magazine/the screenplay} \]

\[(3.26) \text{(b)} \quad \%\text{The babysitter}_{i} \left[ v_{1} t_{1} \right. \text{tickled/teased the book/the magazine/the screenplay} \]

\[(3.26) \text{(c)} \quad \%\text{The expert}_{i} \left[ v_{1} t_{1} \right. \text{read/wrote/perused/edited Gertrude/Cecil/the child/the young acrobat} \]

In determining selection, we are only interested in changing views about the world so that a literal interpretation is possible. Therefore the only interpretations of (3.27) relevant to our 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. For example, (3.27a) 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 metaphorical interpretations of sentences have entered common usage. For example (3.27b) 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.
3.3.3 Theta Relations

In our theory, THETA RELATIONS are those predicate relations Predicate(X,Y), such that X is the head of a phrase P and Y is one of the nonhead constituents of P. In our account, a relation Theta(X,Y) holds iff X selects Y, where X is the head of a phrase containing Y. We distinguish two subtypes of Theta relations: Internal-Theta(X,Y) if the relation Complement(X,Y) holds; and External-Theta(X,Y) if the relation Specifier(Z,Y) holds and Z equals X or a projection of X. (cf. Williams 1980).

The verb eat selects a complement that is possible to eat and a specifier that is animate. These selection restrictions are our basis for maintaining that, e.g., the relations Internal-Theta(eat, the cheese sandwich) and External-Theta(eat the cheese sandwich, Mary) hold in (3.28a).

(3.27) (a) a pregnant car

(b) The vending machine ate my quarter

Below we use the selection tests proposed here to determine the nature of predication in a given collocation. If A selects B, the relation Predicate(A,B) holds. If B selects A, the relation Predicate(B,A) holds.

(3.28) (a) Mary ate the cheese sandwich.

(b) The dog ate the cheese sandwich.

Following Marantz (1981: 48-51, 1984:23-31), we assume that the theta licensor of an external-theta relation includes all internal-theta licensees. Marantz supports his claim by showing that verbs with more than one interpretation are disambiguated by the internal arguments rather than the external (the subject.) For example, the Vs in (3.29) and (3.30) (Marantz 1984, nos. 2.19a,d.e,f.g.h) assign different semantic roles to their subjects depending on the internal arguments.

(3.29) (a) throw a baseball

(b) throw a party

(c) throw a fit

(3.30) (a) take a book from the shelf
(b) take a bus to New York

(c) take a nap

Marantz also claims that in general, there are idioms in which the indirect arguments are fixed which assign a semantic role to their subjects, e.g., *kick the bucket, put the cart before the horse, look a gift horse in the mouth* but very few idioms in which the subject is fixed, but the object varies, e.g., *What’s eating x, the spirit moved x, the jury is out on.*

If "subject idioms" really are extremely rare, this may support Marantz’s hypothesis. However, Bresnan (1982b) has argued that there are in fact many counterexamples. Marantz (1984, 30) suggests that determining whether the alleged counterexamples are unaccusatives is crucial. If they are then they are not true counterexamples because the "missing argument" is an initial object or internal argument. We will not attempt to resolve this issue here.

The selection restrictions associated with *eat* restrict the possible referents of *something* in (3.31a) to "something that is possible to eat", e.g. a cheese sandwich but not an idea, and the possible referents of *it* in (3.31b) to "something animate", e.g. a human being but not a table. However the *cheese sandwich* fails to provide any clear limitations on the possible interpretation of the verbal unit *do something with* in (3.31c). Therefore, according to Selection Rule 1, a transitive verb selects its complement and its specifier, but the complement and specifier do not select the verb.

(3.31) (a) Mary ate something.

(b) It ate the cheese sandwich.

(c) Mary did something with the cheese sandwich.

Complements of *eat*, whose referents are impossible to eat, are acceptable in the context of a fantasy world which assumes that such things can be eaten, and subjects of *eat* which are not animate in the real world are acceptable in a fantasy world in which those things are animate. The (3.31) examples are felicitous, but the (3.32) and (3.33) examples are not. However, (3.32a,b,c) are well-formed in a fantasy world in which Mary is an extraterrestrial or goddess with special abilities enabling her to eat mental energy and abstract concepts. (3.33a) is well-formed if we assume a science fiction world in which machines can eat and (3.33c,d) make sense in a fantasy world where songs and
colors can eat. However, fantasy worlds do not change our assumptions about eating in examples like (3.32) or (3.33), e.g., in (3.32c), eat cannot be interpreted as think up or forget except perhaps metaphorically. Therefore according to Selection Rule 2, the verb selects its complements and specifiers, but the complements and specifiers do not select the verb.

(3.32) (a) %Mary ate the color blue.
(b) %Mary ate the song.
(c) %Mary ate the idea.

(3.33) (a) %The robot ate the iron pellets.
(b) %The song ate the entire country.
(c) %The color blue will eat the entire world.

The locative sense of the preposition on selects a complement capable of being a location and a small clause specifier representing a concrete object. These selection restrictions are our basis for maintaining that, e.g., the relations Internal-Theta(on, the table) and External-Theta(on the table, t) hold in (3.34b), where t is bound to the phrase the book. We assume that in imposing selection restrictions on t, an NP-trace, the preposition is in effect imposing selection restrictions on (and assigning a theta role to) the book, the antecedent of that trace. (See Chapter 6 for an account of the proper interpretation of NP-traces.)

(3.34) (a) Mary examined the eggplant [on the table].
(b) The book, was [t, on the desk].

In (3.35), the selection restrictions associated with the preposition on limit the possible referents of both instances of it: it in (3.35a) must be a location, e.g., a table, not an abstract idea like sincerity and it in (3.35c) must be a concrete object, e.g., an eggplant, not a day of the week. In (3.35b), the phrase the song must refer to a physical manifestation of a song, e.g., a record or a song sheet, not a collection of sounds (though this interpretation would be possible, if the locative phrase was on the radio). Based on Selection Rule 1, this shows that the preposition imposes selection restrictions on its complement and specifier. We are unaware of any unspecified or pronoun-like preposition. Therefore
we cannot construct an example based on Selection Rule 1, which shows that the specifiers and complements of prepositions do not select the prepositions.

(3.35) (a) Mary examined the eggplant [on it].

(b) The song\textsubscript{i} was [\textsubscript{i} on the table].

(c) It\textsubscript{i} was [\textsubscript{i} on the table].

(3.36a) violates the selection restriction on the complement of \textit{on} and (3.36b) violates the selection restrictions on the specifier of \textit{on}. Tastes don’t, generally speaking, represent locations and concepts like \textit{his fondness for plant life} don’t, generally speaking, occupy locations. However, forced interpretations are possible if they adhere to the selection restrictions of \textit{on}. For example, \textit{the sour taste} could be a place in a fantasy land where places are characterized by their tastes, or \textit{the sour taste} could be the name of a boat.\textsuperscript{31} \textit{His fondness for plant life} could occupy a location in some fantasy world where personal opinions are physical objects. However, there are no presuppositions associated with the complements and specifiers of \textit{on} which provide the basis of a fantasy world in which \textit{on} has a different meaning. For example, the meanings of \textit{eggplant} and \textit{fondness} do not cause an interpretation of these sentences to readily come to mind in which \textit{on} must mean \textit{for} in (3.36a). Therefore, by means of Selection Rule 2, these data show that the preposition selects both its complement and specifier, and the complement and specifier do not select the head preposition.

(3.36) (a) %Mary examined the eggplant [on the sour taste].

(b) %[His fondness for plant life\textsubscript{i}] was [\textsubscript{i} on the desk].

(3.37) and (3.38) provide additional examples in which heads select their complements and specifiers. In these cases, the heads are nouns and adjectives. These examples provide our basis for claiming that for each head specifier and head complement pair in these examples, the relations Internal-Theta(Head, Complement) and External-Theta(Head, Specifier) hold. (3.37) illustrates arguments based on Selection Rule 1 and (3.38) illustrates arguments based on Selection Rule 2.

(3.37) (a) The digestion of it/something.

(b) Its digestion of the new type of food.
(c) It was pregnant.
(d) Mary was angry at it.

(3.38) (a) %The digestion of the three hours (by the lizard).
(b) %The time of day’s digestion of the food.
(c) %The bull was pregnant.
(d) %She was angry at the color blue.

In (3.37), the interpretation of it depends upon whether it is being digested, digesting something, being pregnant or subject to Mary’s anger. By Selection Rule 1, this shows that these nouns and adjectives select their complements and specifiers. In some fantasy world, in which our preconceptions were different about time, gender and color, the examples in (3.38) would be well-formed. Our preconceptions about what it means to be digested, be pregnant, or be angry would not change in any fantasy world as a result of their complements and specifiers. On the basis of Selection Rule 2, the relations Internal-Theta(Head, Complement) and External-Theta(Head, Specifier) hold for these examples.

This section gave examples of external and internal theta relations between various types of heads and their complements and specifiers in order to show that X selects Y in the environments in which it is has been claimed that the relation Theta(X,Y) holds. We showed that our selection criteria are compatible with previous views of theta roles, predication and selection. This is a crucial step for motivating the use of selection as a defining property of predicate relations in general.
3.3.4 Modification

MODIFICATION(X,Y) is a predicate relation such that the modifier X provides nonobligatory information about the modifiee Y. Y is the head of the phrase containing both X and Y. This section shows that X selects Y, as determined by the tests proposed above. Previous accounts which assume that the modifier selects for its modifiee include Katz and Fodor (1964), Katz (1964b, 1966), McCawley (1968a), among others. Following GB assumptions, we assume that another defining characteristic of modification is that if Modification(X,Y), then Adjunct(Y,X).

(3.39) and (3.40) illustrate arguments based on Selection Rule 1 that the modifiers select their modifiees, but modifiees do not select their modifiers.

(3.39) (a) The green thing
(b) The angry one
(c) The large one
(d) Mary did that with a hammer.
(e) John did that slowly.

(3.40) (a) Cecil found his keys somehow.
(b) They timed the race somehow.
(c) Cecil found his keys in some manner.
(d) They timed the race in some manner.

In (3.39), the modifiers impose selection restrictions which limit the possible referents of one, thing, do that, and do it. For example, thing in (3.39a) must be something capable of bearing a pigment, e.g., not sincerity. One in (3.39b) must refer to a sentient being. One in (3.39c) must refer to something with variable dimensions, e.g., not an inch. Did that in (3.39d) must refer to an action achievable by means of using a hammer, e.g., not learning math. Did that in (3.39e) must refer to an action which can vary in duration depending on who does it, e.g., not timing an event. Therefore these modifiers select their modifiees.
The adverbial modifiers *somehow* and *in some manner* bear meanings conducive to further specification. However, choice of modifiee fails to provide any further specification as illustrated by the fact that the referents of these adverbials are just as vague in each example in (3.40). Therefore, these modifiees do not select their modifiers.

(3.41) illustrates arguments based on Selection Rule 2 that modifiers select their modifiees, but modifiees do not select their modifiers. *Pregnant* requires that the referent of the nominal it modifies be of the sort that can be pregnant. In a forced interpretation of *pregnant car* our assumptions about cars change, e.g., in a fantasy land where cars are living things. There is no possible world which forces the meaning of *pregnant* to change on the basis of our assumptions about cars. Similarly *slowly* and *with a magnifying glass* modify verbs depicting action. Verbs like *know* describe states and cannot normally be modified by these modifiers. However, in a forced interpretation of (3.41d), *knowing* becomes a process. For example, imagine that knowing is the process of connecting the pathways in one’s brain and it takes different amounts of time to connect different pathways in one’s brain. In this context, there can be certain things one would know quickly and other things that one would know more slowly. *Gertrude knew the fact with a magnifying glass* is well-formed in a world in which magnifying glasses are used to focus the brain on particular pathways, making it easier to know facts associated with those pathways. In these examples, *slowly* and *with a magnifying glass* remain constant in meaning, but the modifiee changes its meaning in a forced interpretation.

(3.41) (a) That is a pregnant cow.
   (b) %You can’t drive the pregnant car.
   (c) Gertrude read the text slowly/with a magnifying glass.
   (d) %Gertrude knew the fact slowly/with a magnifying glass.

3.3.5 A Comparison of Theta and Modification Relations

Internal-Theta(X,Y), External-Theta(X,Y) and Modification(X,Y) are each instances of Predicate(X,Y). This section compares the two theta relations on the one hand with the modification relation on the other. We observe parallels between theta and modification relations which follow from our assumption that they are both predicate relations. Differences between these two types of
relations provide insight into the well-known problem of distinguishing optional complements from adjunct modifiers.

For some pairs of relations $\Theta(X,Y)$ and $\text{Modification}(X',Y')$, where $X$ equals $X'$, or $X$ is a constituent of $X'$, the same selection restrictions are imposed on $Y$ and $Y'$. This correspondence between pairs of relations follows from our assumption that the theta relation and the modification relation are both types of predicate relations. For example, a given adjective $A$ imposes the same selection restrictions on some $N$ regardless of whether the relation $\text{External-theta}(A,N)$ or the relation $\text{Modification}(A,N)$ holds.\textsuperscript{34,35,36} Thus the adjectives select the same type of head nouns in both the (a) and (b) sentences of (3.42) to (3.44). In early transformational grammar, the correspondence between postcopular and prenominal adjectives was captured by means of a transformation which converted a relative clause containing a copula, e.g., *the book that is blue* into a noun phrase in which the adjective modifies the noun, e.g., *the blue book*. Since this sort of transformation is not available in GB, it is important that we have found another way of capturing this correspondence.

\begin{itemize}
\item[(3.42)] (a) The book is blue; %The minute is blue
\item[(b)] The blue book; %The blue minute
\end{itemize}

\begin{itemize}
\item[(3.43)] (a) The day seems long; %Sincerity seems long
\item[(b)] The long day; %The long sincerity
\end{itemize}

\begin{itemize}
\item[(3.44)] (a) The gorilla is angry; %The rock is angry
\item[(b)] The angry gorilla; %The angry rock
\end{itemize}

The selection restrictions imposed by a relative clause on its modifiee correspond to the selection restrictions imposed on the gap contained in the relative clause. In GB theory, the gap (or wh-trace) is bound by the modifiee of the relative clause and either (1) is assigned a theta role (extraction of a complement); or (2) modifies some item in the relative clause (extraction of a modifier). Case (1) is illustrated by a comparison of the (a) and (b) examples in (3.45) and (3.46). The (a) sentences are matrix counterparts to the relative clauses in (b). The same selection restrictions are imposed on the subject or object of some verb or preposition in (a) as on the the head nouns in (b). In this case, we assume that
the relative clause construction converts a theta relation into a modification relation by means of some gap filling mechanism, e.g., binding of a wh trace in GB theory. The selection restrictions imposed by the relative clause on its modifiee derive from the selection restrictions imposed by the theta licensor on its theta licensee. Since theta and modification relations are both types of predicate relations, this fact is hardly surprising. (3.47) exemplifies modifier extraction in relative clauses, a case in which a modifier (a predicate licensor) is a gap filled by the modifiee (a predicate licensee) of the relative clause. This more interesting case requires further analysis, as noted by David Johnson (personal communication).

(3.45) (a) I read the book; %I read the frog
    (b) The book that I read; %The frog that I read

(3.46) (a) I drank the juice; %I drank the steak
    (b) The juice that I drank; %The steak that I drank

(3.47) (a) The thursday that I ate cheese; %The Thursday that two plus two equaled four
    (b) I ate cheese on Thursday; %Two plus two equaled four on Thursday.

The above data captures 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. Given that both modification and theta relations are types of predicate relations, these correspondences are not surprising. We assume that the selection restriction correspondences actually represent closer meaning correspondences between these pairs of modification and theta relations. For example in an analysis using theta roles with names like agent, patient, theme, source, goal, etc., the gap bears the same theta role to the subordinate verb, as the modifiee bears to the relative clause.

Generative grammar has always maintained a distinction between complements and adjunct modifiers. Intuitively, complements are items selected by the head and modifiers are items which are always optional and do not satisfy any requirement of the head. In practice, it is often difficult to
distinguish optional complements from adjunct modifiers. Fortunately, this difference is encoded in our definitions of the Modification and Internal-theta relations and is testable by means of our selection restriction tests. We show that given two constituents, a head X and a nonhead Y, such that Y is either a complement or an adjunct modifier, if X selects Y, Y is a complement and if Y selects X, Y is an adjunct modifier. We also allow for the possibility that a nonhead constituent Y is both a complement and an adjunct if Y both selects for the head and is selected by the head.

Above, we distinguished Internal-Theta(X,Y), External-Theta(X,Y) and Modification(X,Y) in terms of constituent relations, as follows:

- Internal-Theta(X,Y) iff Complement(X,Y)
- External-Theta(X,Y) iff Specifier(X,Y)
- Modification(X,Y) iff Adjunct(Y,X)

The Modification relation differs from the two theta relations in that Modification(X,Y) implies that Y is the head of the phrase containing X and Y, but Theta(X,Y) implies that X is the head of the phrase containing X and Y. Predicate(X,Y) is distinguished from Predicate(Y,X) by means of our selection restriction tests. Therefore, given a phrase X’ made up of a head X and a nonhead Y, such that we know that either a modification or theta relation holds between these two constituents, if Y selects X then Modification(Y,X) and Adjunct(X,Y) hold, but if X selects Y then Theta(X,Y) and Complement(X,Y) hold.

Verbs of traveling or transfer co-occur with many different types of Adverb and P phrases, all of which are optional, e.g., the phrases in parentheses in (3.48). We show that: (1) the phrases specifying the endpoints or the path of the item/people in motion (e.g. from Boston, to New York and across the country) are complements; (2) the phrases by + Noun (e.g. by courier and by truck) which specify the means of travel/transfer are both modifiers and complements at the same time; and the by means of + NP phrases (e.g. by means of a courier/a truck) are modifiers.

(3.48) (a) Mary sent a letter (from Boston) (to New York) (by courier)
(b) They drove/traveled (across the country) (by truck)
(c) Mary sent a letter to New York by means of a courier
(d) They drove/traveled across the country by means of a truck

_Somewhere_ in (3.49a) and (3.49b) refers to the path of travel as determined by the main verbs _send_, _drive_ and _travel_. _Somewhere_ may refer to "(the path) to New York", "(the path) from Boston to New York" or "(the path) across the country", but not "page 109" (somewhere in a book), "on August 31, 1992" (somewhere in time), "the hand" somewhere on a person, etc. This limitation on the possible referents of _somewhere_ shows by Selection Rule 1 that these phrases are complements of _send_, _drive_ and _travel_. In contrast, the _Ps from New York, to Boston_ and _across the country_ in (3.50) do not narrow down the meaning of _did it_ to verbs of the travel/transfer variety. In fact (3.50) favors a meaning incompatible with (3.48) in which some activity _it_ is done repeatedly during a trip. The verb pronoun collocation _did it_ in (3.50) can mean just about anything, e.g., did a jig, did their homework, made faces at people, etc. Therefore, _Ps headed by from, to and across_ do not select the head verbs in (3.48).

(3.49) (a) Mary sent a letter somewhere

(b) They drove/traveled somewhere

(3.50) (a) They did it from New York to Boston

(b) They did it across the country

In isolation, the _Ps_ in (3.51a) and (3.51b) represent a temporal range and an ideological range. However, in forced readings of these semantically anomalous sentences, the _Ps_ can only specify the path of the letter or the path of the vehicle driven. Our assumptions about the _Ps_ change, not our assumptions about the head verbs. For example, (3.51a) makes sense in a fantasy world in which it is possible to open a line of communication between characters in the beginning of the book and characters at the end—under this interpretation the phrases _from the beginning of the story, and to the end_ are interpreted as the path of the letter. In (3.51b), it is easier to interpret the phrase _the political spectrum_ as a place than as a range of ideologies. However forced readings are not possible in which our assumptions about driving and sending letters change. (The only available interpretations in which the meanings of the verbs change are based on metaphor, e.g., if driving is used as a metaphor for
changing ones political position.) Therefore, these tests show that *send* and *drive* select $P$s headed by *to, from* and *across* among their possible complement types.

(3.51) (a) %Mary sent a letter from the beginning of the story to the end

(b) %They drove across the political spectrum

In (3.52), the phrases *by truck* and *by means of a truck* appear to be modifiers, not complements. The most natural interpretation of (3.52b,d) are that *did it* means "complete some journey," "sent some package," or some other activity completable by means of a truck. These meanings are due to presuppositions associated with *by (means of a) truck*. In contrast, *by some means* has the same set of possible referents in (3.52e) and (3.52f)—the verb complement collocations *drive/travel to New York* and *eat the sandwich* impose no presuppositions on these generic *by means of* phrases. These facts show that *by truck* and *by (means of a) truck* imposing selection restrictions on the head verb in (3.52a,b,c,d) and that the verb does not select for the *by means of phrases* in (3.52e,f). According to Selection Rule 1, this means that *by truck* and *by means of truck* are modifiers.

(3.52) (a) John drove/traveled to New York by truck

(b) John did it by truck

(c) John drove to New York by means of a truck

(d) John did it by means of a truck

(e) John drove/traveled to New York by some means

(f) John ate the sandwich by some means

(3.53a-d) are semantically anomalous. In forced readings of these phrases our assumptions about the action depicted by the verb complement collocations change rather than our assumptions about the *by means of* phrase. For example, (3.53a) and (3.53c) are well-formed in an implausible setting in which John is standing in back of a truck containing a sandwich and the truck keeps backing up so John can take a bite of the sandwich and then advancing slightly forward while John chews. In this example in which John is literally eating his sandwich by truck, we must change our assumptions about what is involved in eating a sandwich, but we don't have to change our assumptions about the possible uses of a truck. Therefore according to Selection Rule 2, these *by means of* and *by + noun
phrases are both modifiers, not complements.

(3.53) (a) %John ate the sandwich by truck
(b) %John did his laundry by truck
(c) %John ate the sandwich by means of a truck
(d) %John did his laundry by truck

However, *by truck*, and other phrases consisting of *by* and a singular determiner-less noun referring to a transportation device, behave like complements when they co-occur with transportation verbs. For example, the most likely forced interpretation of (3.54a) entails that trucks can fly, i.e., our assumptions about the ability of trucks to fly changes, rather than our assumptions about flying. In this respect, (3.54a) is a paraphrase of (3.54b), in which the *truck* is clearly an $N$ complement.\(^\text{39}\) It appears that this complement interpretation of the *by* phrase is highly dependent on the object of *by*, as shown by (3.54c-f). The most likely forced interpretation of (3.54c) and (3.54e) is that the nature of flying changes, e.g., flying is simulated on a computer or walking on air constitutes flying. (3.54c) and (3.54e) are unlikely (though possible) paraphrases of (3.54d) and (3.54f). The *by means of* phrases however appear to be modifiers, since the examples in (3.55) do not lend themselves to the interpretations of flying trucks and flying computers.

(3.54) (a) %John flew to New York by truck
(b) %John flew the truck to New York
(c) %John flew to New York by foot
(d) %John flew his feet to New York
(e) %John flew to New York by computer
(f) %John flew the computer to New York

(3.55) (a) %John flew to New York by means of a truck
(b) %John flew to New York by means of a computer

In our analysis of sentences like (3.52a), the relation Modification(*by truck, drive to New York*) holds and the relation Theta(*drive, by truck*) holds. In GBUG, this analysis is represented by making
the phrase by truck the value of the features Modifier, Theta, Complement and Adjunct in a FS where the verb is the value of the Head-Proj arc. In previous versions of GB theory, it would be difficult to represent this analysis.\textsuperscript{40}

Other types of phrases which typically behave like adjuncts can also have this dual complement/adjunct status when they occur very frequently with a particular type of verb, particularly when these verbs and adjuncts participate in some direct object/adjunct or subject/adjunct alternation. For example, with the verbs cook, cut, slice, write, an $N$ can have an instrumental interpretation either as the subject of the verb or as the object of the preposition with, as shown in (3.56a,b).\textsuperscript{41} Sentences

\begin{enumerate}
  \item (a) Mary sliced the bread with the serrated knife
  \item (b) That serrated knife sliced the bread
  \item (c) Mary did something with a serrated knife
  \item (d) %Mary sliced the bread with a hammer
  \item (e) %Mary learned math with a hammer
\end{enumerate}

like (315.15c), depicting unspecified acts done with a knife, tend to be interpreted as acts which can be achieved by means of using a knife, including e.g., cutting and slicing. This suggests a modification relation. However, in forced interpretations of semantically anomalous uses of the instrumental $\bar{P}$s, we must change our assumptions about the object of the $\bar{P}$s, not about the verb. This suggests a theta (complement) relation. For example, the most reasonable forced interpretation of (3.56d) is one in which our assumptions about hammers change, e.g., this hammer has a sharp edge. In contrast, in forced interpretations of semantically anomalous sentences containing instrumental $\bar{P}$s and main verbs which do not participate in this alternation, we must change our assumptions about the meanings of the verbs. For example, the most reasonable forced interpretation of (3.56e) is one in which our assumptions about learning math change, e.g., in some fantasy world, pounding mathematical facts written on sheets of paper into one’s head with a hammer is a way to learn math. These data suggest that these instrumentals are simultaneously both complements and adjuncts.
Previous work in generative grammar uses different criteria than we do for distinguishing between complements and adjuncts. The most popular tests, involve extraction out of ISLANDS. Island, introduced in Ross (1967), include: noun complements, e.g. (3.57a), (3.51a); complements of modifiers, e.g. (3.57b) and (3.51b); and wh phrase complements, e.g. (3.57c) and (3.51c). Huang (1982) proposes that extraction of complements out of islands is better than extraction of other types of phrases including adjuncts and subjects (specifiers). Huang (1982) and subsequent work in this area (e.g., Lasnik and Saito (1984, 1992), Chomsky (1986b), Rizzi (1990)) attempt to motivate these conditions on extraction in terms of a constraint called the empty category principle (ECP) (cf. the sources cited here).

(3.57) (a) ?What did John misrepresent the fact that Mary ate t? 
(b) ?What did Mary talk to her mother before she ate t? 
(c) ?What did you wonder whether Mary ate t? 

(3.58) (a) *How did John misrepresent the fact that Mary read the text t? 
(b) *How did Mary talk to her mother before she read the text t? 
(c) *How did you wonder whether Mary read the text t? 

The questions in (3.57) all ask what Mary ate, i.e., what is extracted from the position in the sentence marked with t. According to the ECP literature, the fact that such extraction is better than the adjunct cases is evidence that the N canonically following eat is a complement. The examples in (3.58) cannot be interpreted as questions about the manner in which Mary read the text, i.e., the manner adverbials cannot be extracted from the positions marked with t. According to the ECP literature, this is evidence that manner adverbials like slowly are adjuncts. Therefore, the island constraints corroborate the results of our our selection tests with respect to the adjunct and complement status of these phrases in (3.28) and (3.41).

(3.59) (a) *Where did John misrepresent the fact that Mary drove/traveled (to/from) t? 
(b) *Where did Mary talk to her mother before she drove/traveled (to/from) t?
(c) *Where_i did you wonder whether Mary drove/traveled (to/from) t_i?*

(3.60) (a) ?By which means_i did John misrepresent the fact that Mary traveled/sent the letter t_i?

(b) By which means_i would Mary think a lot before she traveled/sent the letter t_i?

(c) By which means_i did you wonder whether Mary traveled/sent the letter t_i?

(3.61) (a) ?With what_i did John misrepresent the fact that Mary sliced/cut the bread t_i?

(b) *With what_i did Mary talk to her mother before she sliced/cut the bread t_i?

(c) *With what_i did you wonder whether Mary sliced/cut the bread t_i?

(3.62) (a) ?Who(m)_i did John misrepresent the fact that Mary gave the book to t_i?

(b) What_i did John misrepresent the fact that Mary gave t_i to Bill?

(c) What_i did John misrepresent the fact that Mary gave Bill t_i?

(d) ?Who_i did John misrepresent the fact that Mary gave t_i the book?

(e) *Who(m)_i did Mary eat lunch before she gave the book to t_i?

(f) *What_i did Mary eat lunch before she gave t_i to Bill?

(g) *What_i did Mary eat lunch before she gave Bill t_i?

(h) *Who_i did Mary eat lunch before she gave t_i the book?

(i) *Who(m)_i did you wonder whether Mary gave the book to t_i?

(j) *What_i did you wonder whether Mary gave t_i to Bill?

(k) *What_i did you wonder whether Mary gave Bill t_i?

(l) *Who_i did you wonder whether Mary gave t_i the book?

(3.63) (a) ?Whom_i did John misrepresent the fact that Mary laughed at t_i?

(b) *Whom_i did Mary say nasty things before she laughed at t_i?

(c) ?Whom_i did you wonder whether Mary laughed at t_i?
(d) At whom did John misrepresent the fact that Mary laughed?

(e) *At whom did Mary say nasty things before she laughed?

(f) *At whom did you wonder whether Mary laughed?

The examples in (3.59) to (3.63) show that Huang’s diagnostic does not always accurately distinguish between complements and adjuncts. For example, \( \bar{P} \) complements of travel verbs headed by \( \text{from} \) and \( \text{to} \) cannot be extracted out of islands, as evidenced by the examples in (3.59).\(^{43}\) By means of phrases can sometimes be extracted grammatically out of islands, as evidenced by (3.60). The questions in (3.60) can: (1) be interpreted as modifier extraction, e.g., they can be answered with \( \text{by means of the CIA} \); or (2) they can be interpreted as extraction of \( \text{by} + \) noun phrases and have the mixed complement/modifier reading discussed above, e.g., they can be answered with \( \text{by plane} \). The instrumental \( \text{with} \) phrases in (3.61) cannot be grammatically extracted out of \( \text{wh} \) islands (although extraction from noun complements doesn’t sound so bad to me)? It is not clear why these phrases which have a dual complement/adjunct status just like the \( \text{by} \) plus noun phrases should not be able to be extracted out of islands. Complements of \( \text{give} \) sometimes can and sometimes cannot be extracted from islands as shown in (3.62).\(^{44}\) Complements of \( \text{angry} \) cannot be extracted from islands as shown in (3.63). Therefore Huang’s approach is not accurate for distinguishing complements from adjunct modifiers.

The fact that not all adjuncts and complements fit into the Huang’s generalization shows that the ECP needs to be fine tuned. The ECP literature seeks to explain extraction phenomena (including islands), not necessarily to provide a means of distinguishing complements and adjuncts (though this would be a nice consequence). Therefore we have shown that the ECP does not provide a sufficient means for distinguishing between adjuncts and complements, but we have not provided a reason for rejecting the ECP.

This section compared theta and modification relations. We showed that related instances of theta licensing and modification licensing were constrained by the same selection restrictions. We discussed a diagnostic for distinguishing complements from modifiers based on the fact that modifiers select heads, but heads select complements. We showed that our diagnostic was both more flexible and more accurate that previous approaches based on the ECP.
3.3.6 Degree

DEGREE(X,Y) is a predicate relation such that the degree word X specifies the degree as to which the trait characterized by the adverb or adjective Y applies in a given sentence. This section shows that degree words select the adjectival/adverbial phrases they form constituents with. Specifically, we show that degree words select for adjectival/adverbial phrases which can vary in extent.

On the basis of Selection Rule 1, we show that degree words select for adjectives and adverbs, but adjectives and adverbs do not select for degree words. In (3.57a,b), something is interpreted as a gradable adjective like witty or clever, not an adjective like dead, asleep, friendless, blue-eyed, etc. which cannot be attributed in degrees. In (3.57c,d), that way is interpreted as a gradable adverbial like slowly, clumsily or flirtatiously, but not non-gradable adverbials including unquestioningly, totally, completely, yesterday, or over there. In contrast the phrase to some extent is interpreted as being just as vague a specification of degree regardless of the trait, as exemplified in (3.57e,f). Therefore according to Selection Rule 1, degree words select gradable adjectives and adverbs, but adjectives and adverbs do not select degree words.

(3.64) (a) Mary was too/very something
(b) Mary was more something than Fred
(c) John danced too/very much that way
(d) John danced more/less that way than Mary
(e) John was, to some extent, angry
(f) John danced, to some extent, clumsily

By Selection Rule 2, the examples in (3.65) are evidence that degree words select the adjectives/adverbs they co-occur with. (3.65a,b) are well-formed because, different politicians have different degrees of greediness. However, (3.65c-e) are semantically odd because, dead refers to an absolute state, a life form is either dead or not. However, in a world where there are degrees of deadness, it may be possible for doctors and veterinarians to bring the slightly dead back to life. A similar distribution is found with degree words and adverbs, as shown in (3.66). In forced interpretations of
(3.66c,d), completely is taken to mean "almost completely" and unquestioningly is taken to mean "nearly without question" instead of "without question", i.e., these words are used for emphasizing the extent of the acceptance. The adjectives/adverbs have no influence on the meaning of the degree words in forced readings of these semantically anomalous sentences, e.g., very in (3.66c) cannot mean almost.

(3.65) (a) The politician was too/very greedy
   (b) The politician was more/less greedy than her predecessor
   (c) %The goldfish was very dead/asleep
   (d) %The goldfish was too dead/asleep to revive/wake up
   (e) %Some goldfish are more/less dead/asleep than others

(3.66) (a) The politician’s explanation was accepted too/very quickly
   (b) The politician’s explanation was accepted more/less quickly than her predecessor’s explanation
   (c) %The politician’s explanation was accepted very/too unquestioningly/completely
   (d) %The politician’s explanation was accepted more/less unquestioningly/completely than her predecessor’s explanation.

Therefore our selection restriction tests show that degree words determine the distribution of co-occurring adverbs/adjectives. We represent this fact with predication relation Degree in our lexical entries for degree words.

3.3.7 Quantification

In our grammar, the relation Quantification(Q,N) holds within an \( \overline{N} \) consisting of a determiner/quantifier Q and N an N\(^0 \), \( \overline{N} \) or \( \overline{\overline{N}} \).\(^{45} \) We show that Q selects N, by elaborating on the proposal from Section 3.2 that Q restricts the value of the feature Quant-Value of N. Our arguments based on selection restrictions differ from previous sections in a number of ways. Arguments based on Selection Rule 1 which show that Q selects N are not possible because pronouns do not form \( \overline{N} \)s with determiners and generic nouns like thing are common nouns, subject to all the same determiner noun co-
occurrence restrictions as other nouns. Arguments based on Selection Rule 2 that Q selects N are possible in some cases. However, most Qs restrict Ns differently than the purely selection restriction cases discussed above in that ill-formed examples are ungrammatical, not merely anomalous in meaning. Thus quantification does not always lend itself to Selection Rule 2 type arguments because we cannot get forced readings of ungrammatical examples, e.g. the ungrammatical examples in (3.67). We argue that Q selects N in these instances based on our analyses of the Q N collocations in which it is possible to get forced readings.

(3.67) (a) the (one) dog, the (three) dogs, *the (the/a) dog/dogs
     (b) this dog, this one dog, *this a/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 each dog, *a the dog, *a one dog
     (f) every dog; every three dogs, *every one dog, *every the/a dog

In Section 3.2 (cf. (3.10)), we showed that determiners the, this and these each require that its quantifiee satisfy Figure 3.34, as exemplified in (3.67a,b,c). We attributed the ungrammaticality of *the a dog and *this a dog to the fact that the determiner a cannot occur in a phonetically stressed position and, based on Perlmutter (1970), the determiner between the and the head noun must be stressed. We also showed that the determiners each and a only quantify Ns, i.e., nominals without specifiers, as evidenced by (3.67e,g). With the exception of Perlmutter’s requirement based on phonological stress, we assume that these restrictions are all selection restrictions imposed by these determiners on their nominal heads. We show that the distribution of grammatical violations of these restrictions supports this assumption.
Most quantifiers/determiners select nominals with common noun heads, as is captured in Figures 3.34 and 3.35 by means of the restrictions on the value of the feature \textit{Quant-Value}. Violations of the common noun requirement allow forced interpretations. For example, Weinreich (1966, p. 463) shows that a non-noun like \textit{if} will be interpreted as a common noun when preceded by a determiner as in (3.68a) (Weinreich’s no. 110i). (3.68b) show that other parts of speech will also be interpreted as common nouns when preceded by determiners. Forced interpretations of noncommon nouns like full names, pronouns, dates, names of planets, etc. are also possible, as shown in (3.68d,e,f). In each of these instances, the determiner’s meaning remains constant and the item following the determiner is interpreted as a common noun (or projection thereof). According to Selection Rule 2, this shows that Q selects N, but N does not select Q.

(3.68) (a) %Scientists study the if

(b) %I want to buy an angry/an only/a the/one give/an on/the is

(c) %There is an Adam Meyers in Manhattan and another one in Queens.

(d) %I know one she who owns a credenza.

(e) %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.)

(f) %A new Neptune was found by the astronomer every day for a year.

\( \bar{\text{Ns}} \) consisting of the determiner every and some N are always singular \( \bar{\text{Ns}} \) agreeing with singular verbal inflection as shown in (3.68). However, in (3.69c) every quantifies over a numerically quantified plural nominal. Even though this nominal is plural it has a singular interpretation, e.g., every three dogs in (3.69c) means something like `every group of three dogs'. This group interpretation is clearly due to the determiner every. Although (3.69c) is perfectly well-formed, the singular interpretation of the plural nominal is similar to a forced interpretation in that a particular aspect of meaning of the nominal is attributable to the determiner. Therefore this appears to be evidence that the determiner selects the nominal.

(3.69) (a) Every dog is/*are a good dog
(b) *Every dogs are good dogs
(c) Every three dogs is a good (group of) three dogs
(d) *Every three dogs are good dogs

Weinreich notes that the count mass distinction can be affected by the selectional properties of determiners:

\( \cdots \), any \([\text{-Count}] \) noun X, when used with the indefinite determiner a, functions as a \([\text{+Count}] \) noun meaning ‘a kind of X’: a water, a wine, a blood. Moreover, any \([\text{+Count}] \) noun used with the mass determiner Null amounts to a \([\text{-Count}] \) noun signifying ‘the substance Y’: I prefer brick. This conversion becomes especially effective with the use of the partitive determiner some (unstressed), or with other quantitative expressions: move over and give me some pillow; leave me a little piece of garden; etc. (Weinreich 1966, p. 435)

Following Weinreich, the count/mass characteristic of nominals changes in forced interpretations of the semantically anomalous examples in (3.70). According to Selection Rule 2, this is further evidence that Q selects N.46,47

(3.70) (a) %The cannibal ate too much man
(b) %The giant intellectual book worm from Mars ate some book prepared by one of the finest chefs
(c) %The breathing expert likes an air that has circulated properly before entering his
(d) %Most vampires prefer a blood they find themselves to a blood that is purchased from a blood bank.

The above data show that the relation Quantification(Q,N) differs from the other predication relations discussed above. For the most part combinations of Q and N are either grammatical or ungrammatical. However, in every instance of a semantically anomalous Q N collocation we are aware of, our assumptions about N change in a forced interpretation. According to Selection Rule 2, this shows that Q selects N.

3.3.8 Possessives

In this section, we show that in an N consisting of a possessive specifier P and a head H, there exist two relations: Quantifier(P,H) and some other relation Pred which varies considerably. We discuss the following possible instances of Pred: (1) Theta(H,P); (2) A predicate relation Inalienable(H,P), representing that H is a part or attribute of P; or (3) an ownership relation Own(H,P) which may not be a type of predicate licensing relation, where P must be a sentient being and H is some object P owns (e.g., legally). For all P and H, P selects quantificational aspects of H and H may select other aspects of P. The data we present about possessives are difficult to account for without our predicate licensing mechanism or some equivalent. Therefore possessives provide crucial evidence for our analysis of predication relations.

Possessives place limitations on the quantificational properties of the head. As evidenced in (3.71) and (3.72), singular possessives and definite plural possessives pattern with the determiners the, this and these (cf. Section 3.2) in that they: (1) select nominals satisfying Figure 3.34; and (2) are followed by a phonologically stressed position which cannot be occupied by the unstressed determiner a. Bare plural possessives pattern with the determiner a in requiring nominal heads which satisfy Figure 3.35, as shown in (3.73). As with quantification relations, forced readings are only possible with a subset of selection restriction violations, e.g., the examples in (3.74). Most other sentences containing this type of selection restriction violation are ungrammatical as shown in (3.71) to (3.73). These examples show that the relation Quantification(Possessive, Head) holds in the Ns containing possessive
specifiers.

(3.71) (a) Mary’s/the woman’s one/only/*a close friend was Bruno.

(b) A woman’s one/only/*a pet should be a lizard.

(c) A/the man’s three favorite flavors of ice cream are rum raisin, black raspberry and beef barley.

(d) Mary’s three books are on the shelf.

(e) (The) water’s one/only quality is wetness.

(f) (The) water’s (only) three qualities are wetness, clarity of color and drinkability.

(g) The man’s/A man’s/John’s one (only) examination by a doctor should be interesting.

(3.72) (a) The soldiers’ (one/only/*a) group examination by the doctor was interesting.

(b) The women’s (one/only/*a) friend in common was Bruno.

(c) The soldiers’ (two) group examinations by doctors tended to be uncomfortable.

(d) The women’s (two) friends in common were Bruno and Gertrude.

(3.73) (a) *Soldiers’ (one/only/a) group examination by the doctor tends to be uncomfortable.

(b) *Women’s (one/only/a) friend in common was Bruno.

(c) Soldiers’ (*two) (only) examinations by doctors tend to be uncomfortable.

(d) Women’s (*two) ears tend to be long.

(3.74) (a) %John’s if was a particularly interesting if

(b) %New York’s Adam Meyers is taller than New Jersey’s.

(c) %That woman’s May 18, 1993 was uneventful, but on this woman’s May 18, 1993, the earth exploded.

Williams (1982, p. 283) observes that the relation Pred between a possessive P and the head of the \( \overline{N} \) containing P varies to a much greater extent than the relation between a verb and its subject. We discuss three possible instantiations of Pred: Theta(Head,Possessive), Inalienable(Head,Possessive) and Own(Head,Possessive). We show that: (1) the head selects the possessive in inalienable and pos-
cessive relations; and (2) possessives in the own relation are subject to presuppositions about ownership, which are independent of the head nominal. We show that the theta and inalienable relations are predicate licensing relations, but the own relation does not appear to be a type of predicate licensing relation.

As evidenced by the examples in (3.75) and (3.76), Pred can (at least) be a theta relation, an inalienable (inalienable possession) relation or an own (ownership) relation. (3.75a) is three ways ambiguous: John performs the examination as in (3.75b); he is examined as in (3.75c); or he owns the examination process as in (3.75d). In the last reading, Pred is a relation Own(Head, Possessive). In the other two readings, Pred is some type of relation Theta(Head, Possessive). (3.76a) has a theta relation interpretation similar to the \(N\) The picture of the gem (assuming picture assigns a theta role to the gem.) (3.76b-e) represent relations of the sort Inalienable(Head, Possessive): (3.76b,c) represent part-of relations, e.g., the facets are parts of the gem; and the tires are parts of the car; (3.76d,e) represent attribute-of relations, e.g., the color is an attribute of the gem and the clarity is an attribute of the speech.

(3.75) (a) John’s examination was a success.
(b) John’s examination of the doctor
(c) John’s examination by the doctor
(d) John’s examination was patented. Now he receives a royalty every time a doctor examines a patient using his technique.

(3.76) (a) The gem’s picture
(b) The gem’s facets
(c) The car’s tires
(d) The gem’s color
(e) The speech’s clarity

Theta(Head, Possessive) is a special case of the predicate relation theta discussed above. In nominalizations, the head noun has basically the same selection restrictions associated with its theta
roles as the corresponding verb. Each of these theta roles may be assigned to the possessive specifier (as observed by Williams). For example in (3.75), the possessive may be interpreted as either the examiner or the examinee (the external and internal theta roles of the verb examine). Theta roles assigned to prepositional complements of the noun are not available, e.g., in (3.75b), John cannot be the one examined; and in (3.75c), John cannot be the one doing the examining. In Ns that are not nominalizations, possessive specifiers may have a similar interpretation to complements of the head noun. For example (3.77a) can mean ‘the picture John created,’ ‘the picture John owns,’ or ‘the picture of John.’ This last reading is unavailable for (3.77b). Similarly (3.77c) can mean ‘the biography Mary authored,’ ‘the biography Mary owns,’ or ‘the biography of Mary.’ The last reading is unavailable for (3.77d).

(3.77) (a) John’s picture
(b) John’s picture of Mary
(c) Mary’s biography
(d) Mary’s biography of John

The examples in (3.78) show that, in readings of possessive nouns in which they are theta licen-ses, the head selects the possessive. Its in (3.78a,b,c) may refer to things that can examine, be examined by doctors, take pictures and be photographed, e.g., a smell or color are not good candidates. However, forced interpretations of (3.78de) allow us to change our assumptions about smells and colors. Therefore both Selection Rules 1 and 2 support our claim that the heads select the possessives in these examples.

(3.78) (a) Its examination of/by the doctor
(b) Its picture
(c) Its picture of John
(d) %The color blue’s examination (of/by the doctor)
(e) %The smell’s picture (of John)

The examples in (3.79) show that in inalienable possessive readings, the head selects the possessive. Its in (3.79a,b,c) must be interpreted as entities which respectively can have arms, a color and a
baby. Thus, the examples in (3.79a,b,c) normally refer to things like humans, balloons and animals, but not lemons, ideas and cars. When readings of the semantically anomalous (3.79d,e,f) are forced, it is the possessor, not the possessed which changes meaning. These examples are well-formed in a world where the parts of a lemon include arms, ideas come in colors and cars have babies. Thus both Selection Rules 1 and 2 support our claim that heads select the possessives in these examples.

(3.79) (a) Its arms
  (b) Its color
  (c) Its baby
  (d) %The lemon’s arms
  (e) %The idea’s color
  (f) %The car’s baby

An own reading is only allowed for possessive forms of sentient nouns (excluding fantasy worlds). For example, an own reading is available for the examples in (3.80), but not (3.81). Its in (3.80d) must refer to a sentient being if it has an own reading, e.g., a pet or other life form which people assume is sentient. Forced own readings of (3.81) all involve personification of the possessive. We are unaware of any possible selection restrictions on the owned item, i.e., people can own absolutely anything. (3.80e) is perhaps the oddest example in (3.80). However, a forced interpretation of (3.80e) changes the nature of ownership, not the nature of Cecil or gravity. Thus Selection Rules 1 and 2 and these examples appear to support our claims that (1) the head selects the possessive and (2) the possessive does not select the head. However, all nominal heads which are owned appear to impose the same selection restrictions on the possessive. The possessive must always be sentient for ownership to be possible. If the head selected the nominal, we would expect different heads to impose different selection restrictions on the possessive. We propose that the concept of ownership entails the presupposition that the owner be sentient. According to this explanation, the possible referents of its in (3.80d) and the forced readings of (3.81) examples are all based on facts about ownership, not selection restrictions imposed by the head. Therefore Own(Head,Possessive) does not appear to be a predicate licensing relation.
(3.80) (a) The man’s book
(b) Mary’s favorite molecule
(c) The philosopher’s imagination
(d) Its book/toy/imagination
(e) %Cecil’s gravity

(3.81) (a) The book’s cover
(b) The molecule’s density
(c) The table’s leg

Given an $\bar{N}$ N consisting of a possessive $\bar{N}$ P and H, the head of N, such that Specifier(H,P), this section presented evidence that:

1. For all P and H, Quantification(P,H) holds;
2. Theta(H,P), Inalienable(H,P) or Own(H,P) may also hold;
3. Theta and Inalienable relations are predicate relations, but own seems to be a natural language semantic relation based on presuppositions concerning ownership.

We propose that the phrasal entry in Figure 3.25 (repeated below) for own readings of possessive $\bar{N}$s and Figure 3.25’ for theta and inalienable readings.$^{48}$ We leave representation of the own relation to future research. We assume that own is part of the natural language semantics and not the syntactic representation of possessives because, unlike the other possessive relations we discuss above, own does not appear to reflect any selection restrictions imposed by one item on another.
3.3.9 Conclusion

We found there to be considerable advantage to analyzing a wide range of relations between co-occurring constituents as predicate relations. In our grammar, each lexical entry represents a phrase anchored by a particular lexical item. An item anchors all phrases in which it or its projection is a predicate licensor. We have provided a principled basis for identifying predicate relations, predicate licensors, and predicate licensees, thus making it possible to construct a detailed GB-based lexicon for lexical items in a wide variety of syntactic classes.
The interaction between predicate relations should provide a fertile ground for future research. In particular, the instances in which two phrases select each other require further investigation. For example, it may be that the *by* + noun phrases like *by plane* are modifiers in most cases, but complements with travel verbs. In order to make this determination, we must rethink our criteria for identifying complements and modifiers. With most instances of possessives, it is clear that two distinct predicate relations hold, the quantification relation in which the possessive is the quantification licensee and one other relation in which the possessive is a predicate licensor. The one exception is the own relation.

3.4 Agreement Licensing Properties

3.4.1 Introduction

In our theory, AGREEMENT(X,Y) is a licensing relation representing the satisfaction of a set of well-formedness conditions on Predicate(Z,Y), (where X is usually the same as Z). The abstract relation Agreement(X,Y) may be realized as morphological agreement, adjacency, and/or other concrete relations between X and Y which are prerequisites for Y to be a particular type of predicate licensee. We show that the following are examples of agreement licensing relations:

(1) CASE — Case(X,Y) is GB’s abstract case relation, where Y is some N or Gerund and X is some verb, preposition or complementizer. We distinguish at least two case relations: Complement-Case and Specifier-Case according to the constituent licensing relations which they are closely associated with.

(2) N-AGREEMENT— N-Agreement(X,Y) is an agreement licensing relation where X is a determiner or adjective and Y is some projection of the head noun of the N containing X and Y; and

The combinatorial properties, or syntax, of our agreement relation is the syntax of licensing relations proposed above. Thus FSs anchored by a lexical item X include all agreement relations for which X is the agreement licensor. A major consequence of this syntactic property is that, it is theoretically possible for a FS to have different morphological agreement features than its head. For example, a determiner D can anchor a FS representing an \( N \) NB with the following properties: (1) NB has nominative plural morphology, a realization of Specifier-Case(X,NB) for some case licensor X; and (2) N, the head of NB is represented by a FS with genitive singular morphology, a realization of N-agreement(D,N). We show that an analysis based on this type of lexical entry solves some major difficulties with accounting for the agreement properties of Russian \( N \)s. A similar Categorial Grammar based analysis of these same phenomena has been proposed in Timberlake (1988).

The intended interpretation of agreement licensing relations is captured by Figure 3.36, the Agreement Filter, where the inverse paths \( \text{Predicate-Licensee}^{-1} \) and \( \text{Agreement}^{-1} \) describe FSs in which one path terminating in a predicate licensee arc structure shares with one path terminating in an Agreement licensee arc. While the predicate licensor and agreement licensor may be the same phrase they need not be. Every FS representing a well-formed complete phrase which includes some predicate relation must satisfy the agreement filter. The actual set of constraints that a particular agreement relation represents varies among languages, types of predicates and individual words, e.g., see Chapter 4 regarding word order considerations. Since case is a type of agreement relation and theta is a type of predication relation, it follows that Figure 2.18, our version of GB’s case filter (repeated below) is one instance of Figure 3.36, i.e., Figure 3.36 \( \subseteq \) Figure 2.18.49

\[ \text{The Agreement Filter} \]
\[ \text{Predicate-Licensee}^{-1} \: \Lambda \: \text{Agreement}^{-1} \]

\[ \text{Figure 3.36} \]
One of the effects of providing an explicit account of agreement relations, is that we clarify GB’s distinction between morphological agreement and abstract agreement relations (including morphological and abstract case), thus making it possible for previous unification-based accounts of morphological agreement to bear on our research in GB theory. In our account, morphological agreement provides one form of evidence that an abstract agreement licensing relation exists, i.e., morphological agreement is one possible REALIZATION of abstract agreement. Another possible realization of abstract agreement is the requirement that X and Y be adjacent in order for some relation Predicate(W,Y) to be well-formed, e.g., the case adjacency requirement.

In summary, the Agreement(X,Y) is a licensing relation embodying a set of conditions which must be fulfilled in order for a well-formed relation predicate(Z,Y) to hold. This section generalizes previous work from GB’s case theory module to other types of abstract agreement. We show that unification-based accounts of morphological agreement bear on abstract agreement relations; and that the combinatorial properties of licensing relations solve some difficulties in handling morphological agreement phenomena.
3.4.2 Case as an Agreement Relation

In GB theory, all realizations of abstract case are conditions which must be satisfied to permit a theta relation to be well-formed according to the case filter:

\[(3.82) \text{Theta}(X,Y) \text{ iff Case}(Z,Y)\]

The case filter is realized as Figure 2.18 in GL. FSs representing complete phrases containing some relation Theta(X,Y) must satisfy the case filter if Y is an \(\overline{N}\) or \(\overline{\text{Gerund}}\). Morphological case inflection is one of several possible realizations of ABSTRACT CASE, i.e., if a grammatical sentence contains an \(\overline{N}\) or \(\overline{\text{Gerund}}\) Y inflected for a particular morphological case, this constitutes evidence that there exists some licensing relation Case(X,Y). This section shows that morphological case is a morphological agreement property, just like gender, number and person. Other realizations of abstract case, e.g. adjacency, are similar to morphological case in the sense that they are all members of the set of conditions which must be satisfied (in a particular language) for a theta relation to be well-formed.

In GBUG, we use different features to represent abstract and morphological case: (1) the abstract Case relation \(\text{Case}(X,Y)\) is represented by two feature value pairs \(\text{Head-Proj}: X\) and \(\text{Case}: Y\), where X is the Case Licensor and Y is the Case Licensee. We distinguish two types of abstract case relations with the case licensee features: \(\text{Complement-Case}\) and \(\text{Specifier-Case}\), where the following subsumption relations hold: \(\text{Case} \subseteq \text{Complement-Case}\) and \(\text{Case} \subseteq \text{Specifier-Case}\); (2) Morphological case is represented by the feature \(\text{Morph-Case}\) and its value from the set of possible morphological cases in a given language, e.g., Nominative, Accusative, Dative, etc.

We assume that for each relation Complement-Case(X,Y), a relation Complement(X,Y) holds, and for each relation Specifier-Case(X,Y), a relation Specifier(X,Y) also holds. We have chosen this interpretation of this case relation in order to maintain a tight correspondence between the sources and targets of case licensor/licensee feature and some constituent licensor/licensee, as this turns out to be a good research strategy for developing the analysis proposed in Chapter 7. Furthermore, this assumption maintains the binary branching structure which we are currently assuming for both constituent licensing relations and predicate licensing relations. Chomsky (1986b, 1992) and others view each
projection of a head X⁰ as a segment of that head. Therefore stating that the head assigns case is ambiguous between a number of interpretations, each in which a different segment of the head is the case licensor. While it is possible to interpret Chomsky’s view of case assignment to mean that the X⁰ is the Specifier-Case licensor, this is not the only available interpretation.⁵³

Morph-Case is like the agreement features Person, Number and Gender in the following respects: (1) there are co-occurrence restrictions based on morphological case inflection requiring that only nouns of inflectional class X can co-occur with case assigners of inflectional class Y; and (2) there are co-occurrence restrictions based on person, number, gender and morphological case requiring that only words of syntactic category A and inflectional class X can co-occur with words syntactic category B and inflectional class Y.⁵⁴ e.g., the subject finite verb inflection agreement in English and adjective noun agreement in Russian.

In English, finite verb inflection requires that its subject has a morphological case (Morph-Case) feature value compatible with Nominative and has values for the features Number and Person that satisfy some constraint. For example, the finite verb forms in the first column of Figure 3.37 require that their subjects obey the constraints listed in the second column. Based on some joint unpublished research by David Johnson and myself, we assume the following simplification of agreement categories in English: (1) There exists the following set of possible values for the Agreement feature: {First-Person-Singular, Second-Person-Singular, Third-Person-Singular, First-Person-Plural, Second-Person-Plural, Third-Person-Plural, ¬Third-Person-Singular}; and (2) We assert that ¬Third-Person-Singular subsumes all members of this set except Third-Person-Singular. Therefore, I, we, they, you and the people who wear bags on their heads all agree with the verb forms eat and ate below, but not with eats nor am. He, John and the woman holding the gun all agree with eats and ate, but not eat, nor am. I agrees with am, eat and ate, but not eats. By combining person and number features present in other accounts and asserting subsumption relations, we avoid many of the apparent complications to unification assumed in previous accounts (see Nakazawa 1991 and section 2.6 for details.) For example, the atomic FS ¬Third-Person-Singular may be viewed as an ordinary atomic FS, rather than a negative constraint. This is because the set of values for the feature Agreement is a small set, and all
subsumption relations between the members of that set are known. This is basically the same approach taken in Carpenter (1992, pp. 14-17).

<table>
<thead>
<tr>
<th>Verb Form</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>eats</td>
<td>Morph-Case(Nominative) ∧ Morph-Agr(Third-Person-Singular)</td>
</tr>
<tr>
<td>eat</td>
<td>Morph-Case(Nominative) ∧ Morph-Agr(¬Third-Person-Singular)</td>
</tr>
<tr>
<td>am</td>
<td>Morph-Case(Nominative) ∧ Morph-Agr(First-Person-Singular)</td>
</tr>
<tr>
<td>ate</td>
<td>Morph-Case(Nominative)</td>
</tr>
</tbody>
</table>

Figure 3.37

Case relations in English require that the case licensee has a specific Morph-Case feature value, but places no restrictions on other feature values. We distinguish four morphological cases for English: (1) NOMINATIVE, e.g. the italicized phrases in (3.83); (2) ACCUSATIVE, e.g., the italicized phrases in (3.84); (3) POSSESSIVE, e.g., the italicized phrases in (3.85); and (4) GENITIVE, e.g., the italicized phrases in (3.86). What we call possessive and genitive case are not often distinguished. Possessive case is an inflection on any $\bar{N}$ in the relation Specifier($N_2$, $N_1$) as in (3.84), where $N_2$ can be either an $N^0$ or an $\bar{N}$. Genitive case inflects $\bar{N}s$ and carries the meaning of ownership. It is found in a number of syntactic positions including the complement of the preposition of, e.g., (3.84a) and as the predicate (head) of a small clause, e.g., (3.84b,c). Nonpronomouns generally have two morphological realizations of case: (1) Nominative | Accusative and (2) Possessive | Genitive. Some pronouns have one phonological form for each morphological case, e.g., I, me, my, mine and others have disjunctive Morph-Case values, e.g., the feminine third person singular pronoun has nominative (she), accusative | possessive (her) and genitive (hers) forms.

(3.83) I/you/he/she/we/they/John/the baby/the kids laughed.

(3.84) (a) John saw me/you/him/her/us/them/the book/the books;
(b) John looked at me/you/him/her/us/them/the book/the books
(c) They consider me/you/him/her/us/them/the book/the books to be a threat to the security of the nation.

(3.85) My/your/his/her/our/their/John’s/the baby’s/the kids’ books are on the table.

(3.86) (a) Those books of mine/yours/his/hers/ours/their/John’s/the baby’s/those kids’ are very interesting.

(b) Those books are mine/yours/his/hers/ours/their/John’s/the baby’s/the kids’.

(c) I consider those books mine/yours/his/hers/ours/their/John’s/the baby’s/the kids’.

Languages vary with respect to how abstract case is realized. Baker (1988, pp. 111-124) states that (at least) the following possibilities exist:

(1) The requirement that the licensor and licensee be in adjacent positions in a string and (in some languages) follow a particular word order. For example, English N must be adjacent and to the right of verbal and prepositional case assigners, as exemplified in (3.87). Chomsky (1981, pp. 94-95) suggests that the case adjacency requirement is parameterized: the case adjacency parameter is set to positive in English and negative in many free word order languages.

(3.87) (a) John (really) saw the book.

(b) *John saw really the book.

(c) The book was put (all the way) on the shelf.

(d) *The book was put on all the way the shelf.

(e) Mary (really) believed Sally to have stolen the records.

(f) *Mary believed really Sally to have stolen the records.

(2) Morphological case, the inflection of Ns which restricts abstract case assignment in some languages. For example, (3.83) to (3.86) and the discussion above show how morphological case is realized in English. Languages with richer case morphology like Russian, German, Latin and Finnish typically allow more variation in word order. This supports the idea that case morphology and adjacency conditions serve a similar function. We represent this similarity by
maintaining that both are realizations of the same abstract relation.

(3) Person/Number/Gender Agreement with the case licensor or one of its projections. See for example Figure 3.37 and the discussion above regarding subject agreement in English. Verbs in languages like Tuscarora (see Williams 1976) agree with their objects.

(4) Noun incorporation is a process by which the noun attaches to the verb. According to Baker (1988, pp. 105-129) this process like case assignment, makes it possible for the noun to receive a theta role. This phenomenon occurs in a variety of languages including Mohawk (See Postal 1962).

Under Baker’s analysis, each of the above relations satisfies a different VISIBILITY condition, the case filter being only one possibility. Under our analysis, each of these relations is a type of abstract case assignment and there is only one case filter. We find this to be a clearer way to formulate these ideas.

The case adjacency requirement in English is clearest for examples like (3.87) for the Complement-Case relation. However this requirement also applies to Specifier-Case relations, since subjects of finite clauses and possessive $\bar{N}$s must occur to the immediate left of the $\bar{X}$ case licensors ($\bar{I}$ or $\bar{N}$), as exemplified in (3.88). An $\bar{N}$ A and a possessive $\bar{N}$ B in the relation Specifier-Case(A,B) must be adjacent. The only lexical items which may occur between the the head $A'$ of A and B are modifiers of $A'$ which together with $A'$ form the $\bar{N}$ A. Adjacency is maintained between the $\bar{I}$ X and the subject Y in the relation Specifier-Case(X,Y) provided that all adverbs between the head of X and Y are part of the constituent X (at S-structure). Sentential adverbs like probably and suddenly modify the whole sentence (including the subject) and verbal adverbs like graciously and hungrily modify the $\bar{V}$. (3.88c) illustrates that conjuncts which do not contain the subject may contain either type of adverbs. This shows that adverbs occurring between a finite verb and its subject form a constituent with the finite verb which does not include the subject. This constituent is the $\bar{I}$ X, the case licensor. Therefore the requirement that the case licensor and licensee be adjacent holds for subjects and finite $\bar{I}$s in English.

(3.88) (a) Mary’s filthy loud dirty irritating cocker spaniel

(b) Mary probably/suddenly/graciously/hungrily ate the raw turnips.
(c) Mary [probably/suddenly/graciously/hungrily ate the raw turnips] and [will probably be sick later]

GB Theory assumes that a case recipient must be GOVERNED by its case assigner (e.g., see Chomsky 1986a, p.187), where government is a structural relation defined in Section 3.7. For now, it suffices to list the set of government relations: (1) heads govern their complements; (2) heads govern their specifiers; and (3) heads govern the specifiers of their complements (exceptional case marking). Therefore, the relation Case(X,Y) requires one of the following relations to hold: (1) Specifier(X,Y); (2) Complement(X,Y); and (3) Complement(X,Z) Λ Specifier(Z,Y). If this is correct, then the government relation provides one basis for claiming that the various realizations of abstract case form a natural class—the class of realizations of abstract case.

In summary, GB theory claims that the abstract case relation has a number of possible morphological/phonological realizations. These realizations are similar in that: (1) each language requires case licensors and case licensees to be in relations based on a subset of the possible realizations of abstract case; and (2) for all realizations of abstract case, the case licensee must be the governor of some government relation.

3.4.3 N-Agreement

In our grammar, N-AGREEMENT(X,Y) is an agreement licensing relation, such that X is a quantifier/determiner or adjective and Y is some projection of the head of an $\overline{N}$. N-Agreement may be realized in at least two ways: (1) as morphological restrictions, i.e., restrictions on the values of Morph-Agr and Morph-Case features; and (2) adjacency requirements. The fact that both N-Agreement and Case licensing relations have these same potential realizations follows from our assumption that they are both types of Agreement relations.

The N-Agreement Filter in Figure 3.38 provides a way of interpreting the N-Agreement relation. The N-Agreement Filter is the particular instantiation of the Agreement Filter which holds for the predicate relations: adjective noun (or $\overline{N}$) modification and noun (or $\overline{N}$) quantification. In the overwhelming majority of instances, the more restrictive version of the N-Agreement Filter in Figure
3.38' holds, i.e., in most cases the predicate licensor and N-Agreement licensor are the same. However in Section 3.4.4, we discuss one case of N-Agreement in Russian in which only the less restrictive Figure 3.38 holds.

The N-Agreement Filter

Predicate-Licensee\(^{-1}\) A N-Agreement\(^{-1}\)

Figure 3.38

The abstract N-Agreement relation is realized in English as the requirement that adjective modifiers are adjacent to the nominal heads they modify and quantifiers are adjacent to the nominal heads they quantify. Each adjective/quantifier X predicates of an \(\overline{N}\) or \(N^0\) left adjacent to X, e.g., each adjective \(A_n\) modifies the following \(N_n\) in (3.89), \(N_n\) an \(\overline{N}\) or \(N^0^{,58}\) and each determiner \(D_n\) quantifies the following \(N_n\). Violations of this left adjacency result in ungrammaticality as shown in (3.90).
In Spanish N-Agreement is realized as the additional condition that determiners, nouns and adjectives must agree in gender, e.g., the examples in (3.91). In Russian, N-Agreement is typically realized as morphological agreement of case, gender and number between determiners, adjectives and nouns as exemplified in (3.92). However, Russian examples like (3.93) require a more complex explanation. The fact that (3.93) is perfectly grammatical shows that the abstract N-Agreement relations between the quantifiers, the adjectives and the nouns are well-formed even though the specific nature of the morphological realizations of these N-Agreement relations are not obvious. Below we account for (3.93) using two distinct N-Agreement relations, and a GB-style trace analysis.
In our GBUG based account of morphological agreement, FS representations of morphological agreement are part of the value of agreement licensee features. Thus in our system, morphological agreement is literally a realization of agreement licensing relations (or abstract agreement) because morphological agreement explicitly constrains the possible values of agreement licensee features, i.e., only FSs which satisfy morphological agreement conditions can unify with the value of the agreement licensee feature.

Shieber (1986, pp. 26-27) provides a simple example of subject verb (i.e. subject INFL) agreement which we have modified for GB theory modeled in GBUG. Figure 3.39 unifies with the shared value of the Specifier and Specifier-Case features in Figure 3.40 to form Figure 3.42. However unification of Figure 3.39 with the shared value of the Specifier and Specifier-Case features of Figure 3.41 fails due to the conflict in the value of the feature Morph-Agr, i.e., the feature value Third-Person-Singular does not unify with ¬Third-Person-Singular. Most instances of adjective noun agreement can be accounted for in much the same way. For example, Spanish adjectives and nouns with the same genders are compatible. Therefore, Figure 3.43 representing the masculine noun libro ‘book’ may unify with the shared value of the Head-Proj and Agreement features in Figure 3.44 representing the masculine adjective rojo ‘red’ yielding Figure 3.46, the $\overline{\text{N}}$ libro rojo ‘red book’ as in $\overline{\text{N}}$ un libro rojo ‘a red book’. However, Figure 3.43 cannot unify with the shared value of Head-Proj and Agreement features in Figure 3.46 representing the feminine adjective roja ‘red’. This assumes the following:

1. Third-Person-Singular $\subseteq$ Masculine-Third-Person-Singular
2. Third-Person-Singular $\subseteq$ Feminine-Third-Person-Singular
3. Masculine $\subseteq$ Masculine-Third-Person-Singular
(4)  Feminine $\subseteq$ Feminine-Third-Person-Singular

However the feature values Feminine and Masculine are incompatible.

**Figure 3.39**

**Lexical entry for**

*Uther*

Category: Noun
Phonology: Uther
Quant-Value: Def
Morph-Agr: Third-Person-Singular

**Figure 3.40**

**Lexical Entry for**

*sleeps*

Category: INFL
Tense: Present
Specifer: Category: Noun
   Morph-Agr: Third-Person-Singular
   Morph-Case: Nominative
   Antecedent: [I]
Head-Proj: Category: Noun
   Tense: Present
   Head-Proj: Category: Noun
   Tense: Present
   Complement: Category: Finite-Verb
   Specifer: Category: NP-trace
   Same-Reference: [I]
   Head-Proj: Category: Noun
   Phonology: sleeps
   External-Theta: [O]
   Internal-Theta: [O]
Specifer-Case: [K]
Lexical Entry for the finite entry of sleep

Category: INFL
Tense: Present
Specifier: Category: Noun
   Morph-Agr: Third-Person-Singular
   Morph-Case: Nominative
   Antecedent: 
Head-Proj: Category: Finite-Verb
   Tense: 
   Head-Proj: Category: Complement
   Tense: 
   Complement: Category: NP-trace
      Specifier: Category: NP-trace
      Same-Reference: 
      Head-Proj: Category: 
      Phonology: sleep
      Internal-Theta: 
      External-Theta: 
      Specifier-Case: 

Figure 3.41
FS representing

*Uther sleeps*

<table>
<thead>
<tr>
<th>Category</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tense</td>
<td>Present</td>
</tr>
<tr>
<td>Specifier</td>
<td>Noun</td>
</tr>
<tr>
<td>Phonology</td>
<td>Uther</td>
</tr>
<tr>
<td>Morph-Agr</td>
<td>Third-Person-Singular</td>
</tr>
<tr>
<td>Morph-Case</td>
<td>Nominative</td>
</tr>
<tr>
<td>Quant-Value</td>
<td>Def</td>
</tr>
<tr>
<td>Antecedent</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Head-Proj</th>
<th>Category</th>
<th>Tense</th>
<th>Head-Proj</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Complement</th>
<th>Category</th>
<th>Finite-Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specifier</th>
<th>Category</th>
<th>NP-trace</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Same-Reference | 1 |
| Internal-Theta | 0 |
| Specifier-Case | [K] |

**Figure 3.42**

**Lexical entry for**

*libro*

<table>
<thead>
<tr>
<th>Category</th>
<th>Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonology</td>
<td>libro</td>
</tr>
<tr>
<td>Morph-Agr</td>
<td>Masculine-Third-Person-Singular</td>
</tr>
</tbody>
</table>

**Figure 3.43**
**Lexical entry for**

*rojo*

| Category: | Noun |
| Quant-Value: | [j] |
| Morph-Agr: | [k] Masculine |
| Head-Proj: | [l] Category: [l] |
| Quant-Value: | [j] |
| Morph-Agr: | [k] |
| Adjunct: | [m] Category: |
| Phonology: | rojo |
| Modifier: | [m] |
| N-Agreement-Licensor: | [m] |
| N-Agreement: | [l] |

**Figure 3.44**

---

**Lexical entry for**

*roja*

| Category: | Noun |
| Quant-Value: | [j] |
| Morph-Agr: | [k] Feminine |
| Head-Proj: | [l] Category: [l] |
| Quant-Value: | [j] |
| Morph-Agr: | [k] |
| Adjunct: | [m] Category: |
| Phonology: | roja |
| Modifier: | [m] |
| N-Agreement-Licensor: | [m] |
| N-Agreement: | [l] |

**Figure 3.45**
Figure 3.46

In Figures 3.40, 3.41 and 3.42, the feature value pairs Head-Proj: X and Specifier-Case: Y, modeled as arcs with the same source, represent a abstract case relation (a type of agreement relation) Specifier-Case(X,Y). In Figures 3.44, 3.45 and 3.46, the feature value pairs N-Agreement-Licensor: X and N-Agreement: Y, modeled as arcs with the same source, represent the abstract agreement relation N-Agreement(X,Y) between the adjective X and the head of the phrase Y. The value of the Specifier-Case and N-Agreement features includes Morph-Agr and Morph-Case features and their values. Therefore Y must satisfy the morphological constraints represented by these morphological agreement features in order for the abstract agreement relations to be well-formed.

3.4.5 Morphological Agreement and Agreement Licensing

This section provides GBUG-based solutions to some descriptive problems posed by morphological agreement in Russian. The licensing relations N-Agreement and Case play a crucial role in our solutions.

Josef Fioretta (1993) demonstrates that Russian agreement within and external to numerically quantified Ns cannot be described by unification-based analyses in which each projection of the head and the maximal projection share all morphological agreement features. (3.94) consists of nominative
plural Ns which are quantified by numerals and agree with plural forms of finite verbs. Different numerals require different number and case markings depending on the final digit. Russian nominative Ns with numeral determiners ending in 1 are relatively unproblematic: like most Russian Ns, the numeral determiner ending in 1, any adjectives and the head noun in case, gender and number. As noted above, Ns 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 numbers ending in other digits take genitive plural head nouns. Furthermore in Ns quantified by numbers ending in 2, 3 or 4, masculine nouns co-occur with genitive plural adjectives and feminine 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 Morph-Agr and Morph-Case for each projection of a given head noun must be shared with the accompanying adjectives and determiners. Figure 3.47 summarizes this distribution of (nominative) numerals, nouns and adjectives. First of all, the Ns in (3.94) all agree with plural finite verbs, even though the head nouns of (3.94a,b) are singular. Secondly, internal to the N, the noun adjective co-occurrence patterns are too complex. Corbett (1983: 215-240, 1988: 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.

(3.94) (a) dva novyx studenta
       2 (MASC-NOM) new (GEN-PLUR) student (MASC-GEN-SING)
       ‘2 new students’
(b) dvé novyx/novye studentki
       2 new (GEN-PLUR/NOM-PLUR) student (FEM-GEN-SING)
       ‘2 new (female) students’
(c) pjat’ novyx studentov/studentok
       5 new (GEN-PLUR) students (MASC/FEM-GEN-PLUR)
       ‘5 new students’
In our analysis, nominative numerals ending in digits from the set \{5,6,7,8,9,0\} anchor nominative plural \(N\)s, i.e., an \(N\) \(Y\) anchored by one of these numerals satisfies certain morphological conditions which make the relation \(\text{Case}(X,Y)\) possible for the subset of \(X\) requiring nominative \(\bar{N}\)s (or accusative as per note 59). A determiner \(D\) from this class requires its head \(H\) to be genitive plural in order for the relation \(\text{N-Agreement}(D,H)\) to hold. For example, we propose the FS in Figure 3.48 as the lexical entry for \(pjat'\) the nominative form of ‘five’. All nominative Russian numerals ending in a digit from the set \{5,6,7,8,9,0\} have basically the same lexical entry as \(pjat'\), except of course for the value of the path \(\text{Head-Proj Phonology}\). \(Pjat'\) anchors a nominative plural \(\bar{N}\) and requires that its head be genitive plural. Adjectives combine with nouns agreeing in case, number and gender to form \(N\)s. Thus any genitive plural nominative will unify with the shared value of the features \(\text{Head-Proj}\) and \(\text{N-Agreement}\) in Figure 3.48. By differentiating \(\text{N-Agreement}\) and \(\text{Case}\) we allow for the possibility that agreement properties internal and external to the \(\bar{N}\) are different. Timberlake (1988) has taken a similar categorial grammar based approach to case phenomena in Lithuanian. In Timberlake’s approach these determiners could be viewed as belonging to the category \(N-\text{Nom-Plur}/N-\text{Gen-Plur}\), a category which combines with a category \(N-\text{Gen-Plur}\) to yield a category \(N-\text{Nom-Plur}\).\(^62\)

<table>
<thead>
<tr>
<th>Final Digit (Masc/Fem)</th>
<th>Adjective</th>
<th>Noun</th>
<th>(\bar{N}) containing number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</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>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>

Figure 3.47

\(^{61}\) In Timberlake’s approach these determiners could be viewed as belonging to the category \(N-\text{Nom-Plur}/N-\text{Gen-Plur}\), a category which combines with a category \(N-\text{Gen-Plur}\) to yield a category \(N-\text{Nom-Plur}\).\(^62\)
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 $A_1, \ldots, A_k$ which can occur between N and D differ in their morphological agreement features. On our analysis, sentences like (3.94b) are viewed as instance of antecedent trace relations as shown in (3.95):

$$(3.95) \ [N \text{dve} \ [t1_i \ [\text{novyx} \ [t2_i \ \text{studentki}]])]]$$

We show that the lexical entries in Figures 3.49 to 3.51 and the following agreement relations account for the distribution of the morphological properties of all constituents in phrases like (3.94):

1. N-Agreement1(X,N1), where X is a determiner or adjective and N1 is an adjacent nominal constituent that contains one or more adjective modifiers. The relation N-Agreement1 is optional since not all nominal constituents contain adjectives.

2. N-Agreement2(D,N2), where D is a determiner or a determiner trace and N2 is the head noun.
Lexical entry for
dve

Category: [I] Noun
Morph-Agr: [I] Fem-Plural
Morph-Case: [K] Nominative
Specifier: [I] Category: Determiner
  Phonology: dve
  Morph-Agr: [I]
  Morph-Case: [K]
  Antecedent: [M]
Head-Proj: [I] Category: [I]
  Specifier: [O] Category: Determiner-trace
  Same-Reference: [M]
  Head-Proj: Category: [I]
  Adjunct: Category: Adjective
    Morph-Agr: [I]
    Morph-Case: Genitive | Nominative
N-Agreement-Licensor: [O]
N-Agreement2: [P] Category: Noun
  Morph-Agr: Fem-Sing
  Morph-Case: Genitive
Quantifier: [I]
N-Agreement-Licensor: [I]
N-Agreement1: [M]
N-Agreement2: [P]

Figure 3.49
Lexical entry for

novyx

Category: Noun
Specifiers: Category: Determiner-trace
Antecedent: Specifiers: Category: Noun
Head-Proj: Category: Head-Proj
Adjunct: Category: Adjective
Phonology: novyx
Morph-Agr: Fem-Plural
Morph-Case: Genitive
Head-Proj: Specifiers: Category: Noun
Same-Reference: Specifiers: Category: Noun
Head-Proj: Category: Head-Proj
Adjunct: Category: Noun
Morph-Agr: Fem-Sing
Morph-Case: Genitive
N-Agreement1: Specifiers: Category: Noun
N-Agreement2: Specifiers: Category: Noun
N-Agreement-Licensor: Specifiers: Category: Noun
N-Agreement2: Specifiers: Category: Noun

Figure 3.50

Lexical entry for

studentki

Category: Noun
Specifiers: Category: Determiner-trace
Head-Proj: Specifiers: Category: Noun

Phonology: studentki
Morph-Agr: Fem-Sing
Morph-Case: Genitive
N-Agreement-Licensor: Specifiers: Category: Noun
N-Agreement2: Specifiers: Category: Noun

Figure 3.51

The agreement properties of (3.94b) derive from: (1) the unification of Figure 3.51 with the value of the path Head-Proj Head-Proj in Figure 3.50 to yield Figure 3.52; and (2) the unification of Figure 3.52 with the value of the path Head-Proj in Figure 3.49 to yield Figure 3.53. Properties of the lexical entries ensure the proper result. The lexical entry for dve: (1) states two separate N-agreement
relations, N-Agreement1(dve,Adj) and N-Agreement2(dve,Head-Noun) along with the Morph-Case and Morph-Agr realizations of these relations; and (2) contains a determiner trace t bound to dve, which enters into the relation N-Agreement2(t,Head-Noun), i.e., the path equation $N\text{-Agreement}2 = Specifier N\text{-Agreement}2$ holds. The lexical entry for novyx percolates the agreement properties down to the head noun through any intervening adjectives with like structure. The adjective entry states that: (1) the relation N-Agreement1(novyx, adj) holds for any adjective adj which shares all morphological agreement features with novyx; (2) the determiner trace t1, the value of the path Specifier binds the determiner trace t2, the value of the path Head-Proj Head-Proj Specifier; and (3) the relations N-Agreement2(t1,X) and N-Agreement(t2,X) hold for the same X. The lexical entry for the noun studentki includes the trace which unifies the determiner trace in dve in the phrase dve studentki or t2 in the entry for novyx in dve novyx studentki. T, the determiner trace in the lexical entry for dve unifies with t1. Thus successive binding of traces through unification and the structure sharing between the N-Agreement2 features throughout Figure 3.53, the FS resulting from unification, ensure that the head noun is restricted in the way stated in the lexical entry for the determiner.
FS representing

novyx studentki

Category: Noun
Specifier: Category: Determiner-trace
Antecedent: [j]
Head-Proj: Category: [i]
Adjunct: [k] Category: Adjective
Phonology: novyx
Morph-Agr: [l] Fem-Plural
Morph-Case: [m] Genitive
Head-Proj: [n] Category: [i]
Specifier: [o] Category: Determiner-Trace
Same-Reference: [i]
Head-Proj: Category: [i]
Adjunct: Morph-Agr: [l]
Morph-Case: [m]
Head-Proj: [j] Category: [i]
Phonology: studentki
Morph-Agr: Fem-Sing
Morph-Case: Genitive
N-Agreement-Licensor: [o]
N-Agreement2: [p]
Modifier: [k]
N-Agreement-Licensor: [k]
N-Agreement1: [n]
N-Agreement2: [p]

Figure 3.52
Figure 3.53 contains a chain [D t1 t2] consisting of determiner traces and their antecedent D. In our analysis, determiner chains connect agreement licensor positions and are well-formed only if the
initial and final position in the chain both have the following special property:

LOCAL AGREEMENT — Given a relation Agreement(X,Y) represented by agreement licensor arc A and agreement licensee arc B, Agreement(X,Y) is local iff there exists constituent arcs Q and R such that the source and targets of A match the source and targets of Q and the source and targets of R match the source and targets of B.

The chain [D t1 t2] is local because the relations N-Agreement1(dve, novyx) and N-Agreement2(t2, studentki) are local. The adjectives in our analysis of Russian Ns function much the same as infinitival subject to subject raising predicates, e.g. appear, seem, be likely in sentences like The birds appear to seem to be likely to fly—the adjectives and raising predicates are respectively like catalysts for the agreement relation between dve and studentki and the predicate relation between fly and birds. Like our adjectives, infinitival raising predicates link pairs of traces.63,64

3.4.6 Limitations of Unification

Unification is an extremely useful mechanism for representing the matching of compatible morphological agreement features and their values. However, we show below that unification cannot handle all cases of morphological agreement. For example:

(1) The determination of the morphological agreement properties of coordinate phrases of requires (language specific) rules which are sensitive to word order considerations, particular morphological feature values and particular coordinate conjunctions. These rules are called RESOLUTION RULES (cf. Givón 1970, Corbett 1983:176-214, 1988: 26; cf. Vanek 1970: 45-6 regarding ‘feature computation rules’, an equivalent term).65

(2) Examples from Ingria (1990) in which some item X marked with agreement morphology A | B can agree simultaneously with some item Y marked with morphology A and some item Z marked with morphology B, e.g., the German relative pronoun was, which is morphologically nominative | accusative, can agree simultaneously with the accusative requirements of a verb in a relative clause and the nominative requirements of a matrix verb.
Instances in which the agreement properties of an $N$ is based on its meaning rather than its morphological properties. For example, English singular nouns which can have plural meanings can agree with plural finite verbs and an English plural $N$ like *chicken and rice* agrees with a singular finite verb when it is interpreted as a mass $N$.

(1), (2) and (3) are similar in that they are examples in which morphological agreement must be handled by some mechanism other than unification. These examples illustrate that we must develop a wide range of operations on FSs to handle the full range of morphological agreement phenomena.

We show that the plural agreement properties of coordinate phrases in English and Russian follow from one or both of two resolution rules:

**RESOLUTION RULE 1:** The coordinate phrase is plural, even if some or all of its conjuncts are singular

**RESOLUTION RULE 2:** The Verb agrees with the conjunct closest to the verb

We model Resolution Rule 1 as a conditional rule which is part of the lexical entry for the conjunction, i.e., in the FS representing the coordinate phrase anchored by the conjunction, the value of $\text{Morph-Agr}$ is subsumed by *plural* if the coordinates are of category noun. We model Resolution Rule 2 as a condition based on linear precedence requiring that the agreement properties of a coordinate phrase $X$ is determined by the conjunct adjacent to the agreement licensor $Y$ of the relation $\text{Agreement}(Y,X)$.

In English, all conjunctions of nouns with the coordinate conjunction *and* are plural, e.g. (3.96a). However, coordinate phrases with *or* can be singular or plural depending on whether the coordinates are singular or plural as shown in (3.96b,c) (from Morgan (1972, p.283). If a coordinate phrase with *or* contains both singular and plural coordinates, then the coordinate closest to the agreeing finite verb determines the agreement properties of the whole phrase, as evidenced in (3.96d,e). Therefore English coordinations with *and* follow Resolution Rule 1 and coordinations with *or* follow Resolution Rule 2.

(3.96) (a) *[The woman and the man] *walks/walk to work everyday.*

(b) *[Either) The woman or the man] is/*are a botanist.*

(c) *[Either) All the women or all the men] walk/*walks to work.*
(d) [(Either) Harry or his parents] *is/are coming.

(e) [(Either) Harry’s parents or his wife] is/*are coming.

Russian \( N \) consisting of \( N \)s and the conjunction \( i \) ‘and’ agree with either plural or singular finite verbs, as exemplified in (3.97) (our examples (3.97a,b,d,e) are from Corbett (1983, pp. 95-99).\(^{66}\)). When one of these \( N \)s \( N \) agrees with a singular finite verb \( V \), the conjunct in \( N \) which is closest to \( V \) must agree in gender. Thus the first conjunct in (3.97d) agrees with byla and the last conjunct in (3.97e) agrees with ogorčilo, udručilo i napugalo. Therefore either Resolution Rule 1 or 2 can determine the agreement properties of Russian \( N \)s using the coordinate conjunction \( i \).\(^{67}\)

(3.97) (a) Prepodavalas’ matematika i fizika
teach (pass-fem-sing) math (fem-sing) and physics (fem-sing)
‘Math and physics were taught’

(b) Prepodavalis’ matematika i fizika
teach (pass-plural) math (fem-sing) and physics (fem-sing)
‘Math and physics were taught’

(c) Matematika i fizika prepodavalis’
math (fem-sing) and physics (fem-sing) teach (pass-plural)
‘Math and physics were taught’

(d) Byla v nej i skromnost’
be (past-fem-sing) in her and modesty (fem-sing)
i izjaščesto, i dostoinstvo
and elegance (neut-sing) and dignity (neut-sing)
‘She had modesty, elegance and dignity’

(e) Otkaz katoržan ot rabskoj
rejection (masc-sing) (of the) convicts of slave-like
raboty, vozmuščenje resjawkami i rasstrelami
work, resentment (neut-sing) at the bars and shootings
ogorčilo, udručilo i napugalo
grieved (neut-sing) depressed (neut-sing) and frightened (neut-sing)
pokornyx lagernyx kommunistov
(the) meek (neut-sing) camp communists
‘The convicts’ rejection of slave-like work and the resentment
at the bars and the shootings grieved, depressed and
frightened the camp communists’

Corbett (1983, pp. 183-207) observes that gender of coordinate phrases is always either masculine or feminine in Slovene, Old Church Slavic and Serbo-Croat. Coordinate phrases in these
languages consisting of only feminine conjuncts agree with feminine finite verbs, and all other coordinate phrases agree with masculine finite verbs. This distribution is illustrated by the Serbo-Croat examples in (3.98) taken from Corbett (1983, pp. 187-188).68,69 See Corbett (1983) for related analyses of gender agreement in Czech, Slovak, Sorbian and Polish.

(3.98) (a) Plać i kletve roditelja i sva crying (masc) and curses (fem-plur) of parents and all nastojanja sarajevskih fratara nisu efforts (neut-plur) of Sarajevan monks did not ništa pomagali nothing help (masc-plur) ‘The crying and curses of the parents and all the efforts of the Sarajaevan monks did not help anything’

(b) Znanje i intuicija su kod knowledge (neut) and intuition (fem) have in njega saradivali i dopunjavali him worked together (masc-plur) and supplemented (masc-plur) se . . . each other ‘His knowledge and intuition worked together and supplemented each other’

(c) Njegovo mesto u razvitku kasabe His place (neut) in development of town i njegovo značenje u životu kasabalija and his importance (neut) in life of inhabitants bili su onakvi kako smo ih were (masc-plur) such (masc-plur) as (we) have them napred ukratko opisali before briefly described ‘His place in the development of the town and his importance to the life of the inhabitants were as we have briefly described them before’

(d) Sve snage i sva pažnja biće All powers (fem-plur) and all attention (fem-sing) will be posvećene toj borbi . . . dedicated (fem-plur) to this struggle ‘All powers and all attention will be dedicated to this struggle’

Our account of gender agreement for these Serbo-Croat coordinate Ns is based on the following:
Resolution Rule 3: A coordinate $\overline{N}$ is feminine plural if the feminine part of the Morph-Agr feature value does not conflict with any conjunct.

Resolution Rule 4: Resolution of conflicts is satisfied by r-unification, where R-SUBSUMPTION and R-UNIFICATION are subsumption and unification relations which only apply in the domain of agreement resolution rules.

R-Subsumption Assertion 1: feminine $r$-subsumes masculine

R-Subsumption Assertion 2: neuter $r$-subsumes masculine

Following Resolution Rule 1, each coordinate phrase assumes feminine gender as a default. This default is $r$-unified with the gender values of each of the conjuncts of the phrase as shown in Figure 3.54 to derive the correct distribution of gender in (3.98), e.g., feminine $r$-unify neuter $r$-unify neuter $\rightarrow$ masculine because masculine is the only gender $r$-subsumed by both feminine and neuter.

<table>
<thead>
<tr>
<th>Example</th>
<th>Unification Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3.98a)</td>
<td>feminine $r$-unify masculine $r$-unify feminine $r$-unify neuter $\rightarrow$ masculine</td>
</tr>
<tr>
<td>(3.98b)</td>
<td>feminine $r$-unify feminine $r$-unify neuter $\rightarrow$ masculine</td>
</tr>
<tr>
<td>(3.98c)</td>
<td>feminine $r$-unify neuter $r$-unify neuter $\rightarrow$ masculine</td>
</tr>
<tr>
<td>(3.98d)</td>
<td>feminine $r$-unify feminine $r$-unify feminine $\rightarrow$ feminine</td>
</tr>
</tbody>
</table>

Figure 3.54

In our account, we use the domain specific relations $r$-subsumption and $r$-unification and not normal subsumption and unification to model these relations because otherwise an assertion like feminine $\subseteq$ masculine would reek havoc, e.g., singular feminine $\overline{N}$s could agree with singular masculine finite verbs. Therefore although these agreement phenomena cannot be handled by an ordinary unification based approach, they are do not really represent a great difficulty. An interesting topic for future research would be to further investigate the use of separate sets of defined subsumption relations for unification in different subdomains, e.g., the hierarchical lexicon, agreement relations, building phrases (e.g. in parsing), and other domains.
Ingria (1990) presents evidence that agreement should be handled by a "nondistinctness" check instead of unification. In Ingria’s system, A agrees with B in Feature F if the values of feature F in A and B do not conflict. Ingria shows that agreement in languages like German, French and Hungarian pose significant problems for purely unification-based analyses, but these problems can be solved using a non-distinctness check.

In each of Ingria’s examples, there is an item X which can agree with items Y and Z simultaneously, even though Y and Z conflict in their agreement properties. For example, the German relative pronoun was may agree simultaneously with the accusative requirements of the verb in a free relative and the nominative requirements of the matrix verb, as shown in (3.99a) (Ingria’s no. 6.) The case agreement feature of was are nominative | accusative. The relevant case feature of war is nominative and the relevant case feature of gegessen is accusative. In Ingria’s approach, the grammar needs only to check that Nominative does not conflict with Nominative | Accusative and that Accusative does not conflict with Nominative | Accusative. In contrast, in a unification based approach the morph-case features of was, war and gegessen would be unified together. Since accusative cannot unify with nominative, the unification-based analysis is untenable. (3.99b) (Ingria’s no. 35) is a similar example involving person agreement in German verb ellipsis. Ingria also provides similar examples for WH phrases and topicalization in Hungarian, and conjoined verbs in French (see Ingria 1990 for details.)

Zaenen and Karttunen (1984) and Pullum and Zwicky (1986) have consider similar data, but their data is limited to conjunction, whereas Ingria shows agreement phenomena to be problematic for unification in general.

(3.99) (a) Ich habe gegessen was noch übrig war.
    I have eaten (acc) what (nom | acc) still left was (nom)
    ‘I ate what was left’
(b) ...weil wir das Haus und die Muellers den Garten kaufen.
    because we (1st) the house and the Muellers (3rd) the garden
    buy (1st | 3rd)
    ‘...because we buy the house and the Muellers, the garden.’

A challenge to accounts of morphological agreement is posed by polysemous Ns whose morphological agreement properties depend on which meaning applies in a given context. In our account,
each of these $N$s has a Morph-Agr value of third-singular | third-plural and particular instances can be disambiguated by selection restrictions. Unfortunately, it is not clear how selection restriction violations should be represented in a unification-based approach. We are not interested in violations which result in semantic anomaly, because we are only trying to account for syntactic well-formedness. However, the fact that selection restriction violations with quantifiers can result in ungrammaticality means that selection restriction violations cannot be simply ignored.

For example, the bracketed $N$s in (3.100) all have both plural and singular senses which have more or less plausibility depending on selection restrictions, as indicated by % and *. ((3.100a,b) are from Fitzpatrick and Sager 1981, p.329 and (3.100e-j) are from Morgan 1972, pp. 279-281. I have replaced some of Morgan’s stars with % to indicate that those examples are semantically anomalous rather than ungrammatical.):

(3.100) (a) [The group] has changed its mind

(b) [The group] have changed their minds

c) [The trio] play/plays jazz.

d) [The public] love/loves fast food.

e) [Pickles and strawberries] taste good.

(f) [Pickles and strawberries] tastes good.

g) %[Pickles and ice cream] is sitting on the table.

(h) [Pickles and ice cream] is a favorite of pregnant women.

(i) [40 acres] are/%is ready for John to plow.

(j) [40 acres] is/*are a lot to plow.

$N$s referring to unified groups of people can freely agree with either singular or plural finite verbs depending on whether they are given a singular or plural reading, e.g., the examples in (3.100a-d) as well as similar sentences with group, committee, organization, couple. If you think of the phrase the group in these sentences as a single unit, it must agree with singular finite verbs, but if you think of the group as a set of members, each predicated of the finite verb separately, then the group takes plural

A coordinate phrase can be interpreted as a singular $\bar{N}$ meaning some sort of mixture if all its conjuncts are types of food, chemicals or other substances that can be mixed together. Thus the bracketed $\bar{N}$s in (3.100e-h) are all grammatical assuming that *pickles and strawberries and pickles and ice cream* can either be interpreted as normal plural $\bar{N}$s or as as culinary delicacies like, e.g., the plural $\bar{N}$ *spaghetti and meatballs*. The matrix predicates in (3.100f) and (3.100h) impose selection restrictions on their subjects are sympathetic to the subject being a food, thus making these sentences slightly more plausible than (3.100g).

(3.100i) is ill-formed with *is* because: (1) *40 acres* must have a singular interpretation in order to agree with *is*; (2) the only available singular interpretation of *40 acres* is a quantity; and (3) our assumptions about the world are that one can only plow concrete objects, not abstractions like quantities. However, one could imagine a fantasy world called Math-World in which farmers plowed numbers and quantities instead of concrete objects. In Math-World, both versions of (3.100i) are well-formed. Additionally, the following example is perfectly okay with either *is* or *are*:

(3.101) [The 40 acres] are/is ready for John to plow.

This is a consequence of the fact that *the* allows the $\bar{N}$ *the 40 acres* to easily have the singular interpretation ‘a plot of land measuring 40 acres’. This singular count noun interpretation was unavailable in the absence of the determiner *the*.

(3.100j) with *are* is ungrammatical because *a lot* can only predicate of a singular quantity and, to our knowledge, the only plural interpretation of the phrase *40 acres* is ‘40 one acre plots of land’. The fact that the selection restrictions associated with *a lot* cannot be violated probably relates to the fact that selection restrictions imposed by some determiner/quantifiers also cannot be violated (cf. Section 3.3.7 above), especially since both *a lot* and these determiners have quantificational meanings. Therefore it is not surprising that the phrases *40 is a lot of acres* and *40 acres is a lot* have very similar meanings, since the collocation *a lot of* acts like a determiner.
The unusual instances of agreement above required a variety of different devices to determine the Morph-Agr and Morph-Case feature values of the various Ns:

(1) We propose resolution rules for conjoined Ns based on word order constraints like adjacency; inherent features of conjunctions, e.g., plural and feminine plural; and r-unification, a domain-specific version of unification.

(2) We cited Ingria’s account of data which could be accounted for by a nondistinctness check, but not unification.

(3) We discussed some examples in which selection restrictions restricted agreement.

Following, Ingria (1990) among others, we have demonstrated that a wide variety of operations on FSs are required to solve the problem of morphological agreement, unification and subsumption being only two of them.

3.4.7 Summary

We assume a GB-style view of abstract Agreement relations including abstract case and abstract N-Agreement. Based on previous GB analyses involving the case filter (Figure 2.18), we have proposed a more general filter called the Agreement filter (Figure 3.36) which constrains a wider range of Predicate relations. In GB theory, morphological agreement and morphological case are phenomena which constrains the application of abstract case. However, the intricacies of handling these morphological phenomena have not previously been studied in GB theory. We have provided explicit means for representing morphological agreement and morphological case relations based on previous work on other unification-based theories. Our incorporation of research from non-GB frameworks is facilitated by our unification-based formalism.

3.5 A-Chains

The main point of this section is to describe passive and raising constructions in terms of the interaction between case, theta and constituent relations. Following GB theory, we characterize this interaction using an empty category approach, even though we show, in Chapter 6, that these empty categories should be replaced by structure sharing arcs. For the sake of discussion we revise the case
filter to accommodate an empty category analysis, even though Chapter 6’s approach allows us to maintain the version of the case filter in Figure 2.18. Chapter 6 also provides an alternative view of A-chains to the one presented here.

In GB theory, the well-formedness of raising and passive constructions depends on a version of the case filter which allows theta roles and case to be assigned to two different syntactic positions which are part of the same A-Chain (argument chain)\(^71\), where an A-CHAIN is a set consisting of A, \(t_n\), and zero or more NP-traces \(t_1, \ldots, t_{n-1}\) between A and \(t_n\), such that A is the antecedent of \(t_1\), \(t_1\) is the antecedent of \(t_2\), \(\ldots\), \(t_{n-1}\) is the antecedent of \(t_n\) (cf. Chomsky 1981, 1986a, Jaeggli 1986 among others for GB accounts of passive and raising.) We call A the C-HEAD of the A-chain and \(t_n\) the C-TAIL of the A-chain.\(^72\) A-chains provide a means of extending the case filter to passive and subject to subject raising sentences, e.g., the examples in Figures 3.55 and 3.56. Applied to A-chains, the case filter must be reformulated as follows:

\[
(3.102) \text{For every } C, \text{ an A-Chain, the C-tail of } C \text{ must be theta marked and the C-head of } C \text{ must be case marked.}
\]

For example the chains \([\text{John, NP-trace}_1, \text{NP-trace}_2]\) and \([\text{Mary, NP-trace}_1, \text{NP-trace}_2]\) in Figures 3.55 and 3.56 obey this version of the case filter.\(^73\) When an \(\overrightarrow{N}\) is in a constituent position which is assigned both case and a theta role, that \(\overrightarrow{N}\) comprises an A-chain by itself, i.e., the \(\overrightarrow{N}\) is both the C-head and the C-tail of an A-Chain. We call such an \(\overrightarrow{N}\), a MINIMAL A-CHAIN (MAC). Thus (3.102) accounts for the well-formedness of MACs, e.g., the \(\overrightarrow{N} \text{ the book}\) in \textit{John read the book}, as well as A-chains in raising/passive constructions.
A FS representing an EC analysis of

*John was tickled by the gorilla*

Category: INFL
Tense: Past
Specifier: Category: Noun
  Phonology: John
  Antecedent: I
Head-Proj: Category: I
  Tense: I
  Head-Proj: Category: I
    Tense: I
    Phonology: was
    Complement: Category: NP-trace
      Specifier: Category: NP-trace
        Same-reference: I
        Antecedent: o
      Head-Proj: Category: n
        Head-Proj: Category: n
          Phonology: tickled
          Complement1: Category: NP-trace
            Same-reference: o
            Antecedent: o
          Head-Proj: Category: n
            Phonology: by
            Complement: Category: Noun
              Specifier: Category: Determiner
              Phonology: the
              Head-Proj: Category: n
                Phonology: gorilla
                Complement-Case: I
                Internal-Theta: p
                Suppressed-Theta: r
          Specifier-Case: I

Figure 3.55
C-heads and C-tails are defined in terms of pairs of structure sharing arcs:

1. The C-head is the shared value of one case arc and one constituent licensee arc. We say that the C-head occupies a CASE POSITION because it is in a constituent position which is assigned case.

2. The C-tail is the shared value of one theta arc and one constituent licensee arc. We say that the C-tail occupies a THETA POSITION because it is in a constituent position which is assigned a theta role.

The A-chain begins with A (the case position) and ends with $t_n$ (the theta position), where the A-chain is viewed as a sequence, the first member A being the antecedent of the second member $t_n$, $t_1$ being the
antecedent of the third member \( t_2 \), \( \cdots \), and the penultimate member \( t_{n-1} \) being the antecedent of the final member \( t_n \). In GBUG, the binding of each trace is represented as a combination of two structure sharing arcs: one labeled *Antecedent* and one labeled *Same-Reference*. The FS rooted at the source of the Antecedent arc binds (is the antecedent of) the FS rooted at the source of the Same-Reference arc. The binding of each NP trace \( t_i \), \( 1 \leq i \leq n \), is determined in the lexicon as shown below.

In our view the A-chain represents that, in effect, the same \( \vec{N} \) occupies each syntactic position represented in the A-chain. We generate our representation of a sentence containing an A-chain by unifying together representations of each lexical entry. As appropriate, the different words in the sentence add pieces of the A-chain. For example, the A-chain in Figure 3.55 is derived by unifying Figure 3.57 with the value of the feature *Complement* in Figure 3.58. The subject and by-phrase are unified in the obvious places in order to yield Figure 3.55. In contrast, previous GB analyses of raising and passive constructions assume that each trace \( t_i \), \( 1 \leq i \leq n \), represents the constituent position of the antecedent of the trace in one stage of the derivation of the sentence. For example in Figure 3.55, \( t_2 \) would represent the initial position, \( t_1 \) the subsequent position *John*, the final position. The term *movement* is used to describe these analyses as a metaphor for the change in position of the \( \vec{N} \) in subsequent derivations. Since, we have only one syntactic level of representation, this metaphor is not appropriate for us.\(^{75}\) The parts of our analysis analogous to movement all occur in the lexicon. Therefore, we reinterpret previously discussed constraints on movement as constraints on lexical entries. See Chapter 6 for details. The generation of empty categories without reference to previous derivations has been discussed previously with in EST and GB frameworks in Chomsky (1973, p.157, 1981, p.90), Koster (1978, 1984), Correa (1987, 1988, 1991), Leacock (1991), among others. GPSG (e.g. Gazdar, et. al. 1985, p. 139) also assumes base generated empty categories or [+ Null] categories.
In summary, this section showed how case, theta and constituent relations interact to license clause bounded dependencies. The same basic principles shown here allow raising and passive lexical entries could combine to produce longer A-chains, e.g., in sentences like:
John seemed \( t_1 \) to \( t_2 \) be \( t_3 \) expected \( t_4 \) to \( t_5 \) be \( t_6 \) seen \( t_7 \) by Mary.

A revised version of the case filter was presented as a condition on A-chains in order to account for this type of phenomena, pending further discussion in Chapter 6, where we substitute many of the ec analyses above with structure sharing analyses along the lines of previous work in APG and other frameworks.

### 3.6 Verb Chains

In GB theory, a finite verbal inflectional affix \( I \) is the head of \( \bar{I} \), GB notation for a sentence, and the base verb \( V \) to which \( I \) is attached is the head of the \( \bar{V} \) complement of \( I \) (cf. Stowell 1981). It is hardly controversial that \( V \) and \( I \) form a single phonological unit (in English). Nevertheless characterizing how \( V \) and \( I \) combine is not straightforward under many previous GB accounts. Other contemporary linguistic theories (e.g. CG, GPSG, HPSG, LFG, RG, etc.) assume that the inflected verb is a single syntactic unit, and (for example) finite inflection is a set of morphological features on the verb (including tense, agreement, etc.). In these theories, the determination of how to combine \( V \) and \( I \) is a topic of derivational morphology, not syntax.

In our account, a finite verb \( V-I \) is a single syntactic unit, an amalgam of verbal and finite inflectional features—the finite inflection is the source of the INFL features and the base verb is the source of the verb features of \( V-I \). We assume that \( V-I \) plays two distinct syntactic roles with respect to the subject of the sentence: due to its verbal features, \( V-I \) is a Theta licensor, i.e., the relation theta(\( V-I \), subject) holds; and due to its INFL features, \( V-I \) is a case licensor, case being realized, in English, as nominative morph-case and some morph-agr restrictions, i.e., the relation Specifier-Case(\( V-I \), subject) holds. In this section, we model this dual role of \( V-I \) as follows. We propose that: (1) the inflected verb \( V-I \) is the head of \( \bar{I} \); (2) a verb-trace \( t \) is the head of \( \bar{V} \); (3) \( V-I \) is the antecedent of \( t \); (4) the relation Specifier-Case(\( V-I \), subject) holds; and (5) the relation Theta(\( t \), subject) holds. (For the purposes of this discussion, we assume that if a theta role is assigned to an NP-trace, it is in effect assigned do the antecedent of that trace.)
The relation between inflected verbs and verb traces is modeled as a sequence of \([V-I, t_1, t_2, \ldots, t_n]\) called a V-CHAIN, where V-I, is the C-head of the V-chain and \(t_n\) is the C-tail of the V-chain and each item in the V-chain is the antecedent of the following item. Based on Koopman (1983), the following condition holds for V-chains:

(3.104) The C-head must be a case licensor position and the C-tail must be a theta licensor position

As with our account of A-chains, this account of V-chains will be improved in chapter 6 when we eliminate V-traces in favor of structure sharing Head-Proj arcs with V-I as the value.

Figure 3.59 diagrams an empty category based analysis of the finite inflected verb *eats* as in *Brenda eats lots of cheese every day*. The significant features of our analysis are: (1) *eats*, the head of \(\overline{I}\), is the antecedent of Verb-trace, the value of the path Head-Proj Complement Head-Proj; (2) The subject of *eats*, i.e. the value of the path Specifier, is the antecedent of NP-trace, the value of the path Head-Proj Complement Specifier. (3) the relation Specifier-Case(\(\overline{V}\), subject) holds; and (4) the relation External-Theta(\(\overline{V}\), NP-trace) holds. In our analysis, the two antecedent trace relations in Figure 3.59 represent that the same item is playing two roles in syntax: the inflected verb is acting like the head of \(\overline{I}\) and \(\overline{V}\); and the subject is acting like the specifier of both \(\overline{I}\) and \(\overline{V}\).
Koopman (1983, pp. 137-151) proposes an analysis of finite verbs which differs from our account in at least two respects: (1) Koopman argues that the verbal element, not INFL, assigns case and (2) the verb must "move" to INFL (and combine with INFL) in order to assign case.

Koopman (1983, pp. 188-193) specifically argues against case being assigned by INFL, which she divides into AGR (agreement) and Tense. She summarizes her arguments as follows:

...the Turkish and Italian data illustrate the problems with the assumption that AGR assigns nominative Case. First, in Turkish gerunds which only contain AGR, genitive Case is assigned, and secondly, in certain infinitival clauses in Italian, nominative Case may be assigned in the absence of AGR and Tense. Furthermore, the analysis proposed in Vata and Gbadi seems to carry over straightforwardly to Portuguese, where the language specific property of Portuguese consists of allowing for an INFL with AGR in certain infinitival (i.e. non-tensed) complements. This forces V-movement into INFL, a pro-
cess which, in a way, mirrors affix-hopping. (Koopman 1983, p. 192)

We find Koopman’s position inconsistent with the following facts: (1) both Koopman and I assume that morphological agreement is one realization of the case relation for languages in which subject verb agreement holds; and (2) in most languages, the morphological case of the subject depends on the inflection of the verb—finite verbs typically require nominative case and (apparently) Turkish gerunds require genitive morph-case. Therefore these morphological agreement and morphological case features of the verb must be part of verbal inflection. It would be inconsistent for anything other than INFL to be the source of the case assigning feature of the verb INFL amalgam. Otherwise, GB would have no explanation for why non-finite verbs in most languages do not assign case.

Koopman’s assumption that the verb "moves" to INFL assumes multiple stages of derivation of a sentence, i.e., the verb moves in the sense that it is in one "position" at one stage of the derivation and in another position at another stage. This is in conflict with our assumption that what has previously been modeled as movement operations belongs in the lexicon, i.e., this difference between Koopman’s account and our own stems from different theoretical goals and assumptions.

Chomsky (1986b, 1992) offer approaches to how and why verbs and INFLs combine. Chomsky assumes that the verb starts out as a verb and "moves" to the inflectional position where it amalgamates with the affix. In the Chomsky (1992, pp. 39-42) analysis there is an Agreement relation between INFL and the verb which Chomsky characterizes as a form of \( \phi \)-feature checking. He describes a mechanism for "checking" \( \phi \)-features by which the verb "moves" to the inflectional position where the \( \phi \)-features are assigned and removed if the features of the verb are compatible with these features. Other \( \phi \)-features are located in the specifier positions of INFL for \( N \)s to move to so they can be assigned case. In Chomsky’s system, the position \( t_n \) marks the position from which \( V \) assigns any theta roles before moving to later positions in the derivation. We find this formulation problematic because: (1) the agreement mechanism is destructive. \( \phi \)-features are erased during the derivational process so that the final representation only reflects whether \( \phi \)-feature checking is successful, but does not identify these \( \phi \)-features; and (2) the mechanism assumes "movement" operations effecting multiple stages of a derivation of a sentence. As noted above, movement analyses violate our assumptions
about the role of the lexicon. The mechanism we propose is unproblematic in these respects because:
(1) unification preserves the information contained in the input; and (2) affixation occurs in the lexicon.78,79

The jury is still out on whether finite verbal inflection should be treated as syntactically distinct
from the verb it inflects. This section provided a coherent lexically-based account of finite verbs from
GB’s perspective that the INFL portion and the verbal portion of an inflected verb play separate roles
in syntax. We showed that our account is preferred to previous accounts by Koopman and Chomsky
due, to a large extent, to our unification-based model.

3.7 Command Relations

3.7.1 Introduction

COMMAND RELATIONS are binary relations defined on pairs of subconstituents of a phrase P,
Based on the relative positions of P’s subconstituents in a graph-based representation. Command rela-
tions are crucial in GB, APG, HPSG and LFG (among other) accounts of pro-form coreference and
disjoint reference. GB theory also uses command relations to define well-formedness constraints on
case/theta relations, empty categories, etc. This section provides general means for representing com-
mand relations on FSs in order to: (1) model GB’s command relations in GBUG; and (2) compare key
aspects of command relations from different linguistic frameworks. Towards these goals, we model
GB’s c-command, m-command and government relations; LFG’s f-command relation; and HPSG’s o-
command and local o-command relations.

In GBUG, a command relation is a type of binary relation between two arcs A and B from the
set of licensor and licensee arcs in some FS X. All command relations are defined in terms of the rela-
tive positions of A and B in the hierarchical structure of X.30 We assume the following structural
definitions:
**The Dominates Relation:** (1) A node $a$ DOMINATES a node $b$ if there exists a path with source $a$ and target $b$; (2) A node $a$ DOMINATES an arc $B$ if there exists a path $P$, such that $a$ is the source of $P$ and the target of $B$ is the target of $P$; (3) An arc $A$ DOMINATES a node $b$ if the source of $A$ dominates $b$; (4) 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 for our purposes. cf. Section 3.2.)

**X-BRANCHING NODE:** An X-branching node is a node which is the source of more than one arc of type $X$.

We define three basic command relations:

1. **Arc A X-COMMANDS Arc B if:**
   - (a) $A$ and $B$ are arcs of type $X$
   - (b) $A$ and $B$ are part of the same FS
   - (c) every X-branching node that dominates $A$ also dominates $B$
   - (d) $A$ does not dominate $B$
   - (e) $B$ does not dominate $A$

2. **A MAXIMAL X-COMMANDS B if** conditions (a), (c), (d) and (e) of X-command are met, plus condition (b'):
   - (b') every node which is a root of a maximal projection that dominates $A$ also dominates $B$.

3. **A X-GOVERNS B iff:** A Maximal X-commands B and B Maximal X-commands A.

The $X$ in X-command, maximal X-command and X-government is a variable over the types of arcs considered for these relations. We represent the main differences between the command relations used in different theoretical frameworks as differences in the value of $X$. 

---

186
For expository convenience, we will say that a path A X-commands (maximal X-commands, X-governs) a path B to mean that the final arc in the path A X-commands (maximal X-commands, X-governs) the final arc in the path B, e.g., "path X Y Z X-commands path A B C" means that arc Z X-commands arc C.

3.7.2 GB Command Relations

In GBUG, the GB relations c-command (constituent-command) and m-command (maximal constituent-command) are formulated as X-command and Maximal X-command respectively, with X equal to the set of all constituent licensor and constituent licensee arcs. Lasnik and Saito’s (1984: 240) version of GB’s government (constituent-government) relation is formulated as our X-government relation with X equal to the set of all constituent licensor/licensee arcs. For example, in Figure 2.9 repeated below: (1) the path Specifier c-commands the path Head-Proj Complement Complement; and (2) the path Head-Proj Complement Head-Proj both m-commands and governs the path Head-Proj Complement Complement.

Previous GB definitions assume a constituent structure tree model. For example, the following definition of c-command, a term originating in Reinhart (1976), is taken from Lasnik and Uriagereka (1988):

For A, B nodes in a tree, A c-commands B iff every branching node dominating A dominates B and neither A nor B dominates the other. (Lasnik and Uriagereka 1988, p. 32)

Following Aoun and Sportiche (1983), m-command is defined as follows when assuming a constituent structure tree model:

For A, B nodes in a tree, A m-commands B iff every maximal projection dominating A dominates B and neither A nor B dominates the other.

Lasnik and Saito define government as follows:

\( \alpha \) governs \( \beta \) if every maximal projection dominating \( \alpha \) also dominates \( \beta \) and conversely. (Lasnik and Saito 1984, p. 240)

Clearly, the same constituents c-command, m-command and govern each other in Figure 3.60 assuming these definitions as in Figure 2.9 assuming our FS based definitions.
A FS representation of

*John saw Mary*

Category: [I] INFL
Tense: [I] Past
Specifier: [K] Category: Noun
Phonology: John
Antecedent: [I]
Head-Proj: Category: [I]
Tense: [I]
Head-Proj: Category: [I]
Tense: [I]
Complement: [M] Category: [N] Verb
Specifier: [O] Category: NP-trace
Same-reference: [I]
Head-Proj: Category: [N]
Head-Proj: Category: [N]
Phonology: saw
Complement: [P] Category: Noun
Phonology: Mary
Internal-Theta: [O]
Complement-Case: [P]
External-Theta: [O]
Internal-Theta: [M]
Specifier-Case: [K]

**Figure 2.9**
Chomsky (1981, 1986a), and van Riemsdijk and Williams (1986) provide definitions of government which are basically equivalent to Lasnik and Saito’s definition, except that they place the following restrictions on the governor: (1) they require that the governor be the head of a maximal projection containing the governee; and (2) they restrict the potential governors to lexical X\(^0\)s and INFL, where a "lexical" category is a category from the set \{noun, verb, adjective, preposition\}. In GBUG, these restrictions translate as: (1) only Head-Proj arcs may be governors; (2) the category feature value of the FS rooted at the target of the governor arc is constrained to members of the set \{INFL, noun, verb, adjective, preposition\}. Government with these restrictions is often referred to as LEXICAL GOVERNMENT (government by a lexical head).
Chomsky (1986b), Rizzi (1990) and Chomsky and Lasnik (1991), among others assume more complex definitions of government than Lasnik and Saito (1984). For example, Chomsky defines government as follows:

\[ \alpha \text{ governs } \beta \text{ iff } \alpha \text{ m-commands } \beta \text{ and there is no } \gamma \text{ a barrier for } \beta, \text{ such that } \gamma \text{ excludes } \alpha. \]

The term \textit{excludes} is defined:

\[ \alpha \text{ excludes } \beta \text{ if no segment of } \alpha \text{ dominates } \beta \text{ (Chomsky 1986b, p.9)} \]

The term segment refers to projections of the same head \( \alpha \). The term \textit{barrier} and the terms which Chomsky uses to define barrier are complex and vary considerably. The idea basically is that \( \alpha \) does not govern \( \beta \) if there is some \( \gamma \) which is ‘closer’ to \( \beta \) such that \( \gamma \) governs \( \beta \). We leave a concise GB formulation of these newer definitions of government as well as other GB relations like \textit{proper government} to future research. We note in passing that many of the complexities of the later definitions of government are due to so-called exceptional case marking (ECM) analyses of sentences like:

\begin{align*}
(3.105) \quad & (a) \text{ John believes } [\exists_{\text{I}} \text{Mary to be a linguist}] \\
& (b) \text{ John considers } [\exists_{\text{N}} \text{Mary } [\exists_{\text{N}} \text{a linguist}]]
\end{align*}

In Chapter 7, we reject the ECM analysis altogether in favor of the raising to object analysis. We expect that this, and perhaps other, simplifications of GB theory will make it possible to maintain essentially the Lasnik and Saito (1984) definition of government.

3.7.3 F-command and O-command

In this section, we review f-command and o-command, command relations used to constrain pro-form coreference and disjoint reference in the LFG and HPSG frameworks. This section shows that our command relation predicate-command, X-command with X equal to the set of all predicate licensor/licensee arcs approximates the f-command relation. Furthermore, we show that the main the difference between o-command and f-command is the use by o-command of what HPSG calls the obliqueness hierarchy, also known as the relational hierarchy and the noun phrase accessibility hierarchy:
Strictly speaking, the elements in the obliqueness hierarchy do not exist as primitives of GB theory. However, equivalent terms or combinations of terms exist, sufficient to make it possible to model o-command in GBUG. This is a major step towards a comparison of LFG, HPSG and GB theories of pro-form coreference/disjoint reference and control of the PRO subject of non-finite clauses.

Bresnan (1982b) 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 $\beta$. (Bresnan 1982b, p. 386)

Figure 3.61 is a FS consisting only of predicate licensor/licensee arcs. This predicate structure would be equivalent in both function and structure to an LFG f-structure of this same sentence if we abstract away from: (1) the differences between the grammatical relations assumed in LFG and the predicate licensing relations assumed in GB, e.g., if we assume that the terms external-theta and subject were interchangeable; (2) the differences between the syntactic categories assumed in the two frameworks, e.g., if we assume that LFG’s category $S$ and GB’s category $I$ are equivalent; and (3) If, contra Marantz (1981, 1984), we assume that the external-theta licensor and internal-theta-licensor are the same, i.e., they are both V$^0$ for transitive verbs. Figure 3.61 may be derived from Figure 2.9 by: (1) eliminating all licensor/licensee arcs other than predicate licensor/licensee arcs; (2) replacing all NP-traces, Wh-traces and obligatorily controlled PRO (e.g., subject control PRO, object control PRO, but not arbitrary control PRO) with their antecedents (cf. Chapter 6); and (3) eliminating nonbranching structure from the result of (1) and (2), i.e., eliminating the topmost projections of the INFL and the verb. We call the resulting FS, a PREDICATE STRUCTURE, since, the only licensor/licensee arcs contained therein are predicate licensor/licensee arcs, just like a FS representing a constituent structure would only contain constituent licensor/licensee arcs. The advantage of a predicate structure over a complete feature structure is purely expositional—we eliminate all licensor/licensee arcs which do not concern us for purposes of predicate command and we eliminate all non-predicate-branching nodes, since they will not effect the predicate command relation. The result is a simpler representation for studying predicate argument relations. Our predicate structure plays the same role in our analyses as
L-graphs in APG and Predicate Argument Graphs in SFG.

We define \( f\text{-command}^\text{GB} \), a variant on \( f\text{-command} \) as follows:

For any occurrences of the functions \( \alpha, \beta \) in a predicate structure \( P \), \( \alpha f\text{-commands} \beta \) if and only if \( \alpha \) does not contain \( \beta \) and every predicate structure of \( P \) that contains \( \alpha \) contains \( \beta \).

Clearly \( f\text{-command}^\text{GB} \) is equivalent to our command relation:

**PREDICATE COMMAND** — A predicate commands \( B \) if \( A X\text{-commands} B \), where \( X \) equals the set of predicate licensor/licensee arcs.

---

**A Predicate Structure representation of**

*John saw Mary*

<table>
<thead>
<tr>
<th>Category: [n] INFL</th>
<th>Tense: [j] Past</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head-Proj: Category: [n]</td>
<td>Head-Proj: Category: [n]</td>
</tr>
<tr>
<td>Tense: [j] Verb</td>
<td>Internal-Theta: Category: Noun</td>
</tr>
<tr>
<td></td>
<td>External-Theta: Category: Noun</td>
</tr>
<tr>
<td></td>
<td>Phonology: John</td>
</tr>
<tr>
<td></td>
<td>Phonology: saw</td>
</tr>
<tr>
<td></td>
<td>Internal-Theta: Category: Noun</td>
</tr>
<tr>
<td></td>
<td>Phonology: Mary</td>
</tr>
</tbody>
</table>

*Figure 3.61*

Figure 3.62 is an LFG \( f\)-structure of *People who know John discuss working too hard* taken from Bresnan (1982b, p. 387). Bresnan accounts for the possible coreferents of the PRO subject of *working too hard* by her *Universal Condition on Anaphoric Control* requiring all possible antecedents of PRO to \( f\)-command it. According to Bresnan, the \( \overline{N} \) *people who know John* can be coreferential with PRO, but not the \( \overline{N} \) *John*, because the \( \text{SUBJ}_{\alpha_1} \) feature \( f\)-commands the subject feature with PRO as a value, but the \( \text{OBJ}_{\alpha_2} \) feature does not.
An LFG F-structure representing
People who know John discuss working too hard

<table>
<thead>
<tr>
<th>SUBJ</th>
<th>PRED: ‘PEOPLE’</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJUNCT:</td>
<td>TOP: <img src="#" alt="PRED: (relative) ‘PRO’" /></td>
</tr>
<tr>
<td></td>
<td>SUBJ: <img src="#" alt="PRED: ‘KNOW &lt;(subj)(obj)&gt;’" /></td>
</tr>
<tr>
<td>OBJ</td>
<td>PRED: ‘DISCUSS &lt;(SUBJ)(OBJ)&gt;’</td>
</tr>
<tr>
<td>OBJ: SUBJ</td>
<td>PRED: ‘PRO’</td>
</tr>
<tr>
<td>U: +</td>
<td>PRED: ‘WORK&lt;(SUBJ)&gt;’</td>
</tr>
<tr>
<td>ADJUNCT:</td>
<td>&quot;too hard&quot;</td>
</tr>
</tbody>
</table>

Figure 3.62

Figure 3.63\(^9\) is a GBUG predicate structure representation of Figure 3.62. Our predicate command relation (i.e., f-command\(\text{_{GB}}\)) gives the same result as Bresnan’s f-command. The fact that people who know John, but not John may be coreferential with PRO can be accounted for since the path dominating people who know John predicate commands the path dominating PRO, but the path dominating John does not predicate command the path dominating PRO, i.e., Internal-Theta External-Theta predicate commands the path Internal-Theta Head-Proj Internal-Theta External-Theta, but the path Internal-Theta External-Theta Modifier Predicate-Licensee Internal-Theta Head-Proj Internal-Theta does not predicate command the path Internal-Theta Head-Proj Internal-Theta External-Theta.
Assuming that there is sufficient overlap in GB and LFG assumptions of predicate argument structure, f-command and predicate command should always yield the same results. We will not investigate the use of predicate command further here since the goal of this section is to facilitate research in this direction and a full account of LFG’s control theory will take us to far afield.

Pollard and Sag (1992, pp. 285-325) propose a theory of pro-form coreference/disjoint reference (a Binding Theory) based on the relations local o-command and o-command. Pollard and Sag define local o-command as follows:
Let Y and Z be referential *synsem* objects with distinct LOCAL values. 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 1992, p. 321)

According to this definition, X locally o-commands Y, if X and Y are syntacto-semantic FS representations (e.g., not phonetic representations) of *Ns* which are not expletives (and are hence referential), and X is earlier in the relational hierarchy in (3.106) than Y. The requirement that X and Y have distinct *local* features rules out Y being the HPSG equivalent of a Wh-trace and of X being its antecedent, e.g. in (3.107), taken from Pollard and Sag (1992, p.296). *John* and *e* structure share their *local* values, i.e., they share most syntactic and semantic features. Thus *John* does not locally o-command *e* in either sentence.

(3.107) (a) John, [I like e].

(b) *John, [he said that you like e]*

Pollard and Sag define o-command as follows:

Let Y and Z be referential *synsem* objects with distinct LOCAL values. 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 1992, p. 322)

Clauses (ii) and (iii) in this definition have essentially the effect of f-command or predicate command discussed above, i.e., they identify hierarchical paths of predicate licensor/licensee arcs. O-command differs from predicate command mainly in its use of the obliqueness hierarchy.

Pollard and Sag define o-binding and o-free as follows:

Y (locally) o-binds Z just in case Y and Z are coindexed and Y (locally) o-commands Z. If Z is not (locally) o-bound, then it is said to be (locally) o-free. (Pollard and Sag 1992, p. 292)
Pollard and Sag define HPSG’s Binding Theory as follows:

HPSG Binding Theory

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

(Pollard and Sag 1992, p. 393)

Pollard and Sag’s binding theory correctly predicts the well-formedness or ill-formedness of the following examples from Pollard and Sag (1992, pp. 292-312). It is important to keep in mind that Pollard and Sag do not require non-locally-o-commanded anaphors to be o-bound. Hence they claim that coreference in (3.108j,k,l) are determined by discourse factors.

(3.108) (a) *John$_i$ likes him$_i$
            (b) [The children]$_i$ like their$_i$ friends
            (c) *John$_i$ depends on him$_i$
            (d) *Mary$_i$ described Bill$_j$ to her$_i$/him$_j$
            (e) John$_i$ likes himself$_i$
            (f) *John$_i$ knows Bill$_j$ likes himself$_i$
            (g) John$_i$ depends on himself$_i$
            (h) *John$_i$ knows Bill$_j$ depends on himself$_i$
            (i) Mary$_i$ described Bill$_j$ to herself$_i$/himself$_j$
            (j) The children$_i$ like [each other’s]$_i$ friends.
            (k) Why are [John and Mary]$_i$ letting the honey drip on each other’s$_i$ feet?
            (l) John$_i$ was going to get even with Mary. That picture of himself$_i$ in the paper would really annoy her, as would the other stunts he planned.

We model HPSG’s binding theory in GBUG in two logically equivalent ways:

(1) We use predicate command in place of o-command and derive the effect of the obliqueness hierarchy from the hierarchical organization of the predicate structure, e.g., the external-theta licensee arc predicate commands the internal-theta licensee arc.
(2) We make two basic changes in our assumptions in order to accommodate an HPSG stype analysis: (1) we assume that the objects of prepositional complements of verbs are indirectly assigned theta roles by the verb, i.e., the preposition is not part of the predicate structure and the relation Indirect-Theta(Verb,N) holds (cf. George 1980 regarding this proposal in the EST framework; cf. Johonson and Postal 1980 for a similar proposal in the APG framework); and (2) we assume that the same head is both the external-theta licensor and the internal-theta licensor. We use basically the same definitions of o-command, local o-command and Binding Theory as HPSG, except that: (1) o-command is defined on arcs in a predicate structure and (2) Obliqueness is defined in terms of the hierarchy:

\[(3.109) \text{External-Theta} < \text{Internal-Theta} < \text{Indirect-Theta} < \text{Object in Modifier Phrase}^{90,91}\]

While these approaches seem superficially very different, they make all the same predictions. The trade-off is that the GB style analysis has a slightly larger predicate structure (one additional Head-Proj arc), but does not require (3.109). For example, the fact that John is coreferential with himself in Figures 3.64 and Figure 3.64' follows from either approach because the path Internal-Theta External-Theta predicate-commands the path Internal-Theta Head-Proj Internal-Theta in Figure 3.64 and the path Internal-Theta External-Theta o-commands Internal-Theta Internal-Theta in Figure 3.64'. Similarly the fact that direct objects can (cannot) be coreferential with indirect object anaphors (personal pronouns) as in (3.108d,i) follows from either approach. Thus the Internal-Theta arc dominating Bill predicate commands or o-commands the Internal-Theta arc dominating himself in both Figures 3.65 and 3.65'.^{92,93}
A Predicate Structure representation of
John likes himself

Category: INFL
Tense: Present
Head-Proj: Category: Verb
  Tense: 
  Internal-Theta: Category: Verb
    External-Theta: Category: Noun
      Phonology: John
      Antecedent: 0
    Head-Proj: Category: Noun
    Head-Proj: Category: Noun
      Phonology: likes
      Internal-Theta: Category: Noun
      Phonology: himself
      Same-Reference: 0

Figure 3.64

A Predicate Structure representation of
John likes himself

Category: INFL
Tense: Present
Head-Proj: Category: Verb
  Tense: 
  Internal-Theta: Category: Verb
    External-Theta: Category: Noun
      Phonology: John
      Antecedent: 0
    Head-Proj: Category: Noun
    Head-Proj: Category: Noun
      Phonology: likes
      Internal-Theta: Category: Noun
      Phonology: himself
      Same-Reference: 0

Figure 3.64'
A Predicate Structure representation of
Mary described Bill to himself

Category: [INFL] INFL
Tense: [Past] Past
Head-Proj: Category: [I]
Tense: [Past] Past
Internal-Theta: Category: [Verb] Verb
  External-Theta: Category: Noun
  Phonology: Mary
  Head-Proj: Category: [N]
  Head-Proj: Category: [described]
  Phonology: described
  Internal-Theta1: Category: Noun
  Phonology: Bill
  Antecedent: [o]
  Internal-Theta2: Category: [Preposition] Preposition
  Head-Proj: Category: [to]
  Phonology: to
  Internal-Theta: Category: Noun
  Phonology: himself
  Same-Reference: [o]

Figure 3.65
A Predicate Structure representation of
Mary described Bill to himself

Category: INFL
Tense: Past
Head-Proj: Category: Verb
   Tense: Past
   Internal-Theta: Category: Noun
      Head-Proj: Category: Noun
      Head-Proj: Category: Noun
      Phonology: described
   External-Theta: Category: Noun
      Phonology: Mary
   Internal-Theta: Category: Noun
      Phonology: Bill
      Antecedent: [I]
   Indirect-Theta: Category: Noun
      Phonology: himself
      Same-Reference: [I]

Figure 3.65'

In summary, predicate command, f-command and o-command are mostly equivalent relations if we abstract away from differences between the syntactic categories assumed, the basic relations assumed, and whether a relative ordering of the prominence of the basic relations is defined in the predicate argument structure itself or explicitly as a hierarchical ordering between relations. The relationship between o-command and predicate command is obscured by the use of the obliqueness hierarchy in HPSG. However, it is possible to translate this hierarchy into predicate structure so that predicate command and o-command are equivalent. The only concrete argument we are aware of for using a hierarchical predicate structure and predicate command rather than o-command and the obliqueness hierarchy is the evidence that the verb object collocation selects for the subject of the verb as discussed in Section 3.3 and Marantz (1981:48-51, 1984:23-31).

3.7.4 Conclusion

This section has mapped a set of relations between nodes in trees into a set of relations between arcs on FSSs. In so doing, we have raised the question of whether non-constituent based command relations bear on GB analyses. This section focused on constituent based and predicate based command
relations. However, there is no reason that agreement based command relations should not be investigated as well, e.g., X-command with X equal to the set of agreement licensor/licensee arcs.

We have demonstrated that LFG and HPSG f-command and o-command based analyses of control and binding theory should be viewed in GB in terms of the predicate command relation. Previously, comparing GB and non-GB analyses was mainly on the level of relative descriptive adequacy. Typically there were certain examples which each framework handled better. However, by casting these non-GB analyses into GB we have removed further barriers to comparison. It is now possible to more clearly see what bearing a particular LFG/HP SG analysis has on GB theory and perhaps combine insights of both approaches. An actual comparison of predicate command based analyses with previous GB constituent command relation based analyses is a topic for future research.

For an account of the mathematical properties of command relations defined on constituent structure trees, see Kracht (1993). 94

3.8 Conclusion

This chapter formulated the underpinnings of GB theory in terms of binary relations called licensing relations. In GBUG, a licensing relation is modeled as a pair of two arcs with the same source: one licensor arc and licensee arc. Licensor and licensee arcs are the basis of our model of constituent relations, predicate relations and agreement relations, as well as, perhaps, types of licensing relations which we have not explored here.

We provide a principled basis for formulating \( \bar{X} \) theory, GB’s theory of constituent structure in terms of one set of constituent relations and one set of constraints. Our formulation is both descriptively adequate and captures the main linguistic generalizations represented by previous versions of \( \bar{X} \) theory, e.g., the relation between a head and its projection, the differences between complete phrases and intermediate projections of heads, etc. We showed that: (1) previous GB specifications of \( \bar{X} \) theory were inadequate for describing the full range of constituent structures assumed in GB theory; and (2) if \( \bar{X} \) theory was extended along previously proposed lines of research, the result would be a system riddled with ambiguity and inconsistencies.
We provided an detailed explication of predicate and agreement relations. Predicate relations were defined in terms of selection restrictions. This explicit definition provided the basis for extending previous GB accounts of theta relations to an account that encompassed a wider range of predicate relations. Agreement relations were defined in terms of so-called agreement filters—sets of constraints on predicate relations, which include morphological agreement and adjacency, among others. One important result is that we have facilitated the incorporation of a large body of research on morphological agreement into GB theory by clearly differentiating morphological and so-called abstract agreement relations, where the former is merely one case of the latter.

This chapter also dealt with various uses for licensing relations in GB theory, including: (1) Chains of constituent licensor/licensee arcs, where the first and last node in the chain is the target of some predicate and/or agreement arc. We have discussed three types of chains, each with different sets of properties: Determiner chains, Verb chains and A-chains; and (2) Command relations, constraints on configurations of licensor and licensee arcs are useful for constraining various modules of grammar including binding and control, among other things, in a variety of linguistic frameworks. By providing a means for representing a wide variety of GB command relations, we made it possible to compare analyses of binding and control from different linguistic frameworks.

One glaring omission from this chapter is an explication of what properties of human language constituent relations represent. Due to the complex history of constituent relations in generative grammar as a whole and GB theory in particular, and the overlap in meaning between constituent relations and other licensing relations, we cannot take up this line of inquiry until Chapter 7. The purpose of much of the research presented prior to Chapter 7 is to make a (partial) explication of constituent relations possible.
Notes

1 We assume that the following three statements all have the interpretation that X and Y form a constituent together (possibly with other items as well.)

(i) X is the complement of Y

(ii) X is the specifier of Y

(iii) X is the adjunct of Y

However, the GB literature is not consistent in this respect. For example, the statement the $N$ is the complement of the verb means that the verb and the $N$ are part of the same constituent, but the statement the determiner is the specifier of the $N$ means that the determiner is part of the $N$.

2 Our usage of the term licensing property does not fit the mathematical usage of the term property. However, we find this term useful for characterizing GB’s notion of assignment of some X, a member of {case, theta role, etc.} which has no clear analog in mathematics.

Discussing GB analyses in terms of assignment of various licensing properties reflects current terminology better than discussion in terms of licensing relations, even though the latter is more mathematically precise.

3 Jackendoff (1977) develops an alternative version of $\overline{X}$ theory in which there are three bar levels instead of two. To its credit, Jackendoff’s version of $\overline{X}$ theory covers a wider range of structures than the Chomsky (1970) version. However, Jackendoff’s system is distinct from the version of $\overline{X}$ theory assumed in most GB analyses which we discuss. A comparison between our version of $\overline{X}$ and Jackendoff’s is outside the scope of this dissertation.

4 Ray Dougherty (personal communication) suggests that completion constraints may be parameterized. An interesting topic for future research may be to determine the distribution of completion constraints across syntactic categories/languages in order to provide an explicit analysis in terms of this parameter.
5 Radford (1988, p.207) shows that nominal adjunct modifiers of nouns must be $N$s. His argument hinges on the fact that nominal modifiers cannot contain determiners, as evidenced by the grammatical examples (ia-b) and the ungrammatical examples (ic,d). This is one environment in which an $X$ can, but an $\bar{X}$ cannot occur as an adjunct.

(i) (a) $\bar{N}$ the [$N$ eighteenth century philosophy] [$N$ book]]
(b) $\bar{N}$ the [$N$ big time] [$N$ band leader]]
(c) *$\bar{N}$ the [$N$ $\bar{N}$ the eighteenth century philosophy] [$N$ book]]
(d) *$\bar{N}$ the [$N$ $\bar{N}$ Mary’s philosophy] [$N$ book]]

6 Consider GB’ a theory of grammar like GB in every respect except that no empty categories are used, so that $\bar{V}$s and $\bar{I}$s need not contain specifiers. For GB’, we propose no completion constraints for $\bar{V}$ and $\bar{I}$s.

7 David Johnson (personal communication) notes that more than one instance of the degree word *very* can occur with adverbs and adjectives as in (i). This also holds for other intensifiers like *really* in (ii). However, it is usually not permissible to mix degree words as shown in (iii). We propose that repetitions of the same degree word for emphasis should be viewed as a single constituent, e.g., *very very very very* is a constituent in (i).

(i) She was very very very very happy.
(ii) The podiatrist was really really really involved in her work.
(iii) *They were very quite happy.

One apparent counter-example to our claim that degree words do not alternate is found in (iv). Note, however, that *really* and *very* cannot alternate, as shown in (v). The sequence of *very* must follow the sequence of *really* as per (vi) and (vii). In addition, it appears that the meaning of *really* is adverbial (sort of like ‘actually’) when it cooccurs with *very*. Otherwise, *really* can have a meaning more similar to *very*. Further study is needed to fully understand this interaction.
They were really very happy.

*They were very really very happy.

They were really really really very very very happy.

*They were very very very really really really happy.

8We assume that degree words like *very* and words like *right* which convey emphasis are specifiers in examples (3.2b,e,g) and the temporal adverb *instantly* in (3.2c) is a modifier. Since each $\bar{X}$ of these types can take only one phrase of type adjunct or specifier, we do not know of any formal syntactic means of determining whether *very, right* and *instantly* in these contexts are adjuncts or specifiers.

Intuitively, *very* and *right* seem to behave as types of quantifiers. Therefore we propose that they are specifiers on the analogy of quantifiers in $\bar{Ns}$. Additionally, the comparative *more* can quantify over both the degree of a gradable adjective as in (i) or the number of an $\bar{N}$ as in (ii). Examples like (iii) are ambiguous between these two readings (similar ambiguities are often found with quantifiers). The most natural assumption is that both these positions are specifier positions.

(i) This book was more expensive than that one.

(ii) John had more books than Mary.

(iii) Mary had more expensive books than John.

*Instantly*, in contrast with these degree words, is an adjunct modifier in every other environment it occurs in. The intuitive meaning of *instantly* in (3.2c) concur with this interpretation.

9In Chapter 7 we adopt a raising to object analysis of these constructions. The result is a small clause with an NP-trace specifier, according to an empty category analysis as in (i) or alternatively in the structure sharing analysis, the same $\bar{N}$ is both the complement of the verb and the specifier of the small clause.

(i) I consider $\bar{N}$ the toothbrush $\bar{\bar{N}}$ NP-trace $\bar{\bar{\bar{N}}}$ very $\bar{\bar{\bar{\bar{N}}}}$ expensive$\bar{\bar{\bar{\bar{\bar{N}}}}}$]
Jespersen (1924), Emonds (1970) and Jackendoff (1973), among others propose that verbal particles and adverbs are intransitive prepositions, prepositions which do not take complements. This suggests that like verbs, only some prepositions take complements. Under this view, a completion condition explicitly requiring $P$s to contain complements would be inaccurate. (This assumes that this type of completion condition is posited in spite of our arguments above to the contrary.)

We emphasize that completion conditions are language specific. For example, all Russian $N$s satisfy the completion conditions for that language. Therefore bare nouns, or nouns modified by adjectives may occur as complements of verbs, etc. For example:

(i) eto kniga
   This book
   ‘This is (a/the) book’

(ii) eto xoroshaya kniga
   This good book
   ‘This is a good book’

Perlmutter (1970, pp. 241-3) claims most proper nouns with articles are grammatical, but only when modified by relative clauses. The examples below suggest that this view is two narrow. It appears that other types of modification are sufficient to allow articles and certain types of quantification, e.g., numerals, also permit articles.

(i) There are three Johns in my class
(ii) The Cecil with the red hair is a friend of mine, the other Cecils are not.
(iii) The Paris that is in Texas, not the one in France
(iv) The three Marys arrived at the hotel simultaneously.

Perlmutter also points out that certain proper nouns which occur obligatorily with determiners, cannot take relative clauses, as shown:
However, we observe a different distribution for the $\bar{N}$ *The Bronx*, as exemplified in (vi) to (x) below. The proper noun *Bronx* must occur with the determiner *the*. The only modifiers which can occur between *the* and *Bronx* are expletives. Modifiers following the $\bar{N}$ are however, acceptable.

(vi) *A Bronx*

(vii) *The very large Bronx*

(viii) (We had to drive all the way to) the bloody/fuckin’ Bronx

(ix) The Bronx that I know is very different than the one you know.

(x) The Bronx in New York City is very different from the Bronx colony on Mars.

Sometimes proper names may behave like $\bar{N}$s when parts of nicknames, as shown.

(i) Little John, Fish-fingered Clyde, Silly Sally

The feature *Morph-Agr* takes morphological agreement properties as values, e.g., *Singular, Plural, Third-Person-Singular*, See Section 3.4 for details.

As shown in Figure 3.14, mass nouns like *money* are singular (in English) in the sense that they agree with third person singular verbs as shown by the grammaticality of (i) and the ungrammaticality of (ii).

(i) Money causes lots of suffering in the world.

(ii) *Money cause lots of suffering in the world.

The determiners *a* and *the* each anchor $\bar{N}$s with a specified value for the head feature *Quant-Value*. Notice that even though a determiner is not the head of the phrase it anchors, the determiner determines one of the head features of that phrase.

Possessive $\bar{N}$s may receive theta roles from the head of the phrase. For example in (i), *John* can be interpreted as the one being examined or the one conducting the examination. See Section 3.3 for details.
(i) John’s examination of/by the doctor

18 We are discussing the quantificational sense of *only*, not the adjectival sense. The adjectival sense of *only* found in *an only child* is different than the sense of *only* in the examples discussed in the text.

19 (3.18) is based on Chomsky (1975a, p.246). See Larson (1988) and Chomsky (1992, 19) for a variant on this approach which is becoming popular in recent GB and Minimalist literature. See Chapter 7 for further details.

20 If we do not assume the binary branching analysis, multiple complements are sisters to the same head. This requires that Rule (3) be generalized to Rule (3'):

\[(3') \bar{X} \rightarrow \text{Head-Proj: } X^0 + (\text{Complement: } \bar{X})^*\]

The Kleene star in (3') means that a head can have any number of complements. In practice, there is probably an upward bound on the number of possible complements—in English, the upward bound is probably three, e.g.,

(i) I bet [John] [five dollars] [that he couldn’t eat three gallons of ice cream in one hour]

(ii) I traded [John] [a red jelly bean] [for a blue one]

In order to maintain (3') and the Kay’s functionality requirement on FSs, we must number complements (use the features \textit{Complement1, Complement2,} \ldots) as we have been doing with theta roles.

It is not within the scope of this dissertation to chose between the flat structure for complements imposed by rule (3') and the binary branching structure imposed by rule (6). It appears that adopting rule (3') would be a stronger view for advocates of the \(\bar{X} \) configuration hypothesis. However, recent literature (e.g., Chomsky 1992) appears to favor the binary branching analysis of multiple complements.

21 Section 3.2.2 argues that N-x is an \(\bar{N} \). However, this choice by no means follows from any previous descriptions of \(\bar{X} \) rules.
The GB literature (e.g., Chomsky 1986b) often assumes that adjuncts are never sisters to $X^0$s and complements are always sisters to $X^0$. This provides a basis for claims that configurational distinctions underly analyses of certain syntactic phenomena rather than distinctions between complements and adjuncts. For example, (i), but not (ii) allows wh extraction from the clause *the fact that Mary fixed the car*. (In (ii) *How* cannot modify the way the car was fixed. The answer to the question could not be with a wrench.) Previous accounts define the difference between these two instances of extraction configurationally, i.e., sisters to $X^0$ can be extracted in this environment, but not sisters to $\bar{X}$ (see for example, the discussions of proper government in Chomsky 1986b). Assuming the binary branching analysis of multiple complement constructions like (iii), this claim is problematic. However, if the relation *Complement* was the determining factor, these facts would be accounted for.

(i) What did John misrepresent the fact that Mary fixed the car?

(ii) *How did John misrepresent the fact that Mary fixed the car?*

(iii) What did John misrepresent the fact that Mary gave Bill the car?

Chomsky and Lasnik (1991, p.29) note that $Y$ is ambiguous in the rule:

$$\bar{X} \rightarrow \bar{X} Y$$

However, they do not comment on the extent of ambiguity throughout the $\bar{X}$ system they assume.

In the approach in which nonbranching nodes are generated: (1) both complements and adjuncts can be in the position sister to $\bar{X}/$daughter of $\bar{X}$; and (2) Both adjuncts and specifiers can be in the position sister to $\bar{X}/$daughter to $\bar{X}$. That approach disambiguates some of the inconsistent definitions discussed above by strictly enforcing the three bar level distinctions and requiring all $\bar{X}$s to have all three bar levels. One can imagine a similar ad hoc system in which all definitions could be made configurationally. In this system, instead of labeling arcs with specifier, complement and adjunct features, extra nonbranching structure would be added and category types would be changed so that adjuncts would always be daughters of $\bar{X}$ and sisters of $\bar{X}$, complements would always be sisters of $X^0$ and specifiers would always be daughters of $\bar{X}$.
In categorial grammar when a category of type X/Y or X\Y (the functor) is combined with a category Y, the result is X. Informally, the combining operation works much like the multiplication of fractions. Therefore (in one variant of Categorial Grammar) an intransitive verb (or verb phrase) would be of type S\NP; a transitive verb would be of type (S\NP)/NP; an adverb that modifies a VP would be of type (S\NP)/(S\NP), etc.

Categorial grammar is a framework based on Leśniewski (1929), Ajdukiewicz (1935) and Bar-Hillel (1953, 1964), among others. See Oehrle, Bach and Wheeler (1988) for some recent work.

This analysis of pronouns is based on discussions from Katz and Postal (1964, p.83) and Katz (1972, p.89) regarding how the meanings of pronouns are effected by selection restrictions.

Selection restrictions are useful for both sense disambiguation and placing limits on possible referents of phrases. However, only the latter provides evidence that one item selects another. For example, the most felicitous readings of (i) are derived by testing the compatibility of the various senses of read/tease and the phrase the oracle. In the phrase tease the oracle, tease is used in the sense of "taunt" rather than the sense "comb", and the oracle is used to refer to a person rather than a prophecy. Each word in the verb complement collocations in (3.23c) appears to rule out potential meanings of the other. Furthermore, it is not obvious from this example that presuppositions associated with either word add any meaning to the felicitous sense of the other word. Thus this does not provide any clear evidence that either word selects the other.

(i) Cecil read/teased the oracle.

This approach of disambiguating word combinations from selection restrictions comes from a body of literature including Katz and Fodor (1964), Katz and Postal (1964), Katz (1964a, 1964b, 1966, 1972), Weinreich (1966), among others.

The distinction between semantic anomaly and ungrammaticality is made clear by Chomsky’s (1965, p.149) famous grammatical, but semantically odd sentence Colorless green ideas sleep furiously. Interestingly this distinction is thousands of years old as noted by Weinreich:

But the problem exemplified by grammatically faultless, yet semantically odd expressions such as
Colorless green ideas has an old history. For two thousand years linguists have striven to limit the accountability of grammar vis-à-vis abnormal constructions of some kinds. Apollonios Dyskolos struggled with the question in 2nd century Alexandria; Bhartr hari, in 9th century India, argued that barren woman’s son, despite its semantic abnormality, is a syntactically well-formed expression. His near-contemporary in Iraq, Sibawaihi, distinguished semantic deviance (e.g., in I carried the mountain, I came to you tomorrow) from grammatical deviance in quad Zaidum qām for qad qām Zaidun ‘Zaid rose’ (the particle quad must be immediately followed by the verb). The medieval grammarians in Western Europe likewise conceded that the expression cappa categorica ‘categorial cloak’ is linguistically faultless (congrua), so that its impropriety must be elsewhere than in grammar.

... [the following text is a footnote to the text above] On Bhartr hari, see Chakravarti (1933: 117f.) and Sastri (1959: 245); unfortunately, neither source does justice to the subject. The Indian argument goes back at least to Patanjali (2nd century B. C.). See also Sibawaihi (1895:10f); Thomas of Erfut (c. 1350: 47). (Weinreich 1966, p. 414)

30See McIntosh (1961), Ziff (1964), Weinreich (1966), G. Lakoff (1969) and McCawley (1968a, 1968b, 1970), Katz (1972), among others regarding the fact that selection restrictions may be violated in various types of fantasy settings. 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 colocations. Weinreich (1966, p. 465) notes that some types of violations of selection restrictions have become conventionalized in literature, e.g., personification. Also Ralph Grishman (personal communication) was very helpful in clarifying my views on this topic.

31Ray Dougherty (personal communication) alerted me to the ‘name of a boat’ interpretation.

32Following Rouveret and Vergnaud (1980, 130-132), we assume that the noun digestion takes an N complement. The preposition of is required in order for the complement to be case marked.

33See Halliday (1967), Williams (1980), Rothstein (1985), Hornstein and Lightfoot (1987), Rappaport (1991), among others regarding sentences in which phrases appear to modify nonheads. For example, in (i) the adjective nude modifies the subject of the sentence and in (ii) the adjective raw modifies the object.

(i) Cecil only eats vegetables nude.

(ii) Cecil only eats vegetables raw.

Rappaport (1991) observes that these modifiers modify the action depicted by the main verbs as well as the NP they modify. For example (i) has the interpretation that Cecil only eats vegetables when
he (Cecil) is nude. Following Rappaport, our account of these constructions would involve two predicate relations: one modification relation and one external theta relation. The adjectives in these constructions both modify the verb (or its projection) and assign external theta roles to the noun, either directly (cf. Williams 1980 and Rothstein 1985) or by means of PRO (cf. Hornstein and Lightfoot 1987).

We treat verbs like be, seem and remain as a subject to subject raising verbs following Postal (1976, p.158, note 18), Pullum and Wilson (1977), Stowell (1978), among others. Under the GB interpretation of subject to subject raising, the adjectival, nominal or prepositional complement assigns an external theta role to its NP-trace subject and this subject is bound to the matrix subject.

Adjectives which can only occur attributively like former and leading in (i) and (ii) only anchor FSs in which the relation Modification(adjective,X) holds, where X is a nominal constituent. Adjectives which can only occur predicatively like ajar and alive in (iii) and (iv) only anchor FSs in which the relation External-Theta(Adjective, X) holds.

(i) The former mayor; *the mayor was former

(ii) The leading lady; *The lady was leading

(iii) The door was ajar; *The ajar door

(iv) The victims were alive; *The alive victims

Leslie Barrett (personal communication) notes that adjectives licensing the tough construction license that construction regardless of whether they occur as prenominal modifiers or predicates, as illustrated in (i) and (ii). This follows from our assumption that adjectives anchor both types of FSs containing their predicate licensees. Presumably, in (iia), person is a modifiee and to talk to is an internal-theta licensee.

(i) (a) Hildegard is difficult to talk to.

(b) *Hildegard is unfair to talk to.

(ii) (a) Hildegard is a difficult person to talk to.
(b) *Hildegard is an unfair person to talk to.

See Chapter 7 note 8 regarding the implications of (ii) for Chomsky’s (1981) definition of the projection principle.

See Postal and Pullum (1988), Stroik (1990) and Chapter 7 of this dissertation regarding expletive pronouns which occur in complement positions, e.g., *it in (i). The verb selects these pronouns in some sense, but it is not clear that, strictly speaking, they are assigned theta roles.

(i) I would hate it for you to leave without tasting these mashed potatoes.

(3.49b) has a possible readings in which somewhere is a locative modifier indicating the location in which the activity of "driving" takes place. This is much clearer in (i) a progressive variant of (3.49b), e.g., somewhere can refer to "in France" or "in the parking lot", etc. This possibility is the result of multiple valencies of the verb drive and the multiple senses of the locative adverbial somewhere. If somewhere is preceded by a preposition, like to or from, the modification reading is ruled out, as in (ii) and (iii).

(i) They were driving somewhere

(ii) They drove to/from somewhere

(iii) They were driving to/from somewhere

The relevant reading of (3.54b) is the anomalous one in which the truck is flying. The well-formed reading in which John is sending a truck to New York is not relevant here. (Catherine Macleod, personal communication)

Grimshaw (1990, pp. 109 - 110) claims that passive by-phrases have both complement-like traits and adjunct-like traits. (See Chapter 2, note 28.) However according to our selection tests, by-phrases clearly pattern with complements. For example eaten requires, the object of the by-phrase to be animate. Therefore (i) is well-formed, the object of (ii) is interpreted as being animate and the deviance of (iii) may be resolved in the context of a fantasy world where rocks are animate.
(i) The food was eaten by the alligator.
(ii) The food was eaten by something.
(iii) %The food was eaten by the rock.


42Related proposals are made in Cattell (1976), Hornstein and Weinberg (1981) and Kayne (1981). Cattell notes that complements, but not subjects can be extracted out of Islands. Both Hornstein/Weinberg and Kayne distinguish objects of prepositions which can be extracted from those that cannot. This turns out to be essentially a complement/adjunct distinction.

43All the examples in (3.59) are ungrammatical. However, the examples below which contain both \textit{from} phrases and \textit{to} phrases sound worse than the single complement examples.

(i) (a) *Where did John misrepresent the fact that Maria drove from \textit{t} to New York?
(b) *Where did Mary talk her mother before she drove from \textit{t} to New York?
(c) *Where did you wonder whether Mary drove from \textit{t} to New York?
(d) *Where did John misrepresent the fact that Maria drove from Boston to \textit{t}?
(e) *Where did Mary talk her mother before she drove from Boston to \textit{t}?
(f) *Where did you wonder whether Mary drove from Boston to \textit{t}?
(g) *Where did John misrepresent the fact that Maria drove to New York from \textit{t}?
(h) *Where did Mary talk her mother before she drove from \textit{t} to New York?
(i) *Where did you wonder whether Mary drove to New York from \textit{t}?

44In informal surveys of our acquaintances, we have found there to be considerable variation with respect to the acceptability of these and similar examples.

45We abstract away from such issues as quantifier scope and from detailed accounts of quantification other than reference to the feature \textit{Quant-Value} and the possible values proposed in Section 3.2.
We do not currently assume ‘null determiners’ in $\bar{N}$s which lack phonological determiners. However, given Weinreich’s examples, it appears that the absence of a determiner forces a mass interpretation of singular count nouns. One way of capturing this would be to claim that a null determiner selects for a plural count or a mass nominal in examples like (i). We account for this distribution differently. We claim that verbs select for complements that are complete $\bar{N}$s and thus *book* in (i) is interpreted as an $\bar{N}$, a mass noun being the most likely interpretation.

(i) The giant intellectual book worm from Mars ate book for lunch

Weinreich specifically speaks out against analyses using ‘selection conditions’. These arguments were directed against specific proposals made in the literature of the time. In that literature, the interaction between semantic features and selection restrictions was much more articulated than in the present account. We have not formulated a specific enough formulation of natural language semantics for these Weinreich’s objections to carry over. In fact, our account is probably equally compatible with Weinreich’s work and the work he was arguing against.

Note that Theta and Inalienable-Licensee properties must be head features. Modifiers and other items which can occur between a head and a specifier percolate these features to the specifier position, via structure sharing. See for example our treatment of the Quant-Value feature in Section 3.2.

Lapointe (1988) proposes that morphological agreement relations only hold between logically connected (L-connected) elements, where he defines L-connected as relations between determiners and nouns, verbs and nouns, adjectives and nouns, i.e., essentially the same pairs of elements as we define predicate relations. In our terms, Lapointe’s proposal would be something like

(i) Agreement(X,Y) iff Predicate(X,Y)

One defect of Lapointe’s proposal is that it does not allow for instances in which some Y is in predicate and agreement relations with different licensors, e.g., exceptional case marking or raising to object type analyses of *believe*-type verbs. Therefore sentences like (ii) would violate his analysis since the relation Case(*believes, her*) holds, but the relation Theta(*believes, her*) does not.
(ii) I believe her to be a linguist

Another difference between Lapointe’s proposal and our own is that Lapointe proposes a condition on morphological agreement rather than on the predicate relation. Whether agreement phenomena are licensed by predication or vice versa may in fact be a chicken and egg type problem. However, morphological agreement and the other realizations of abstract agreement are a nonuniform set of syntactic phenomena not logically necessary for understanding, whereas predicate relations appear to be a single phenomenon, one which is essential for understanding. We believe that it is a more promising research strategy to view the disparate set of syntactic phenomena grouped together under the umbrella of agreement as conditions on predicate relations, rather than the other way around.

In Meyers (1991), I argue that small clause $N$s do not require case. For example, the complements of the verbs $be$ and $consider$ in (i) and (ii) are both $\bar{N}$. Following Postal (1976, p.158, note 18), Pullum and Wilson (1977) and Stowell (1978), $be$ is a subject to subject raising verb. $Be$ does not assign case to its complement. The specifier of its complement must raise to subject position to receive case. I argue that neither small clause $\bar{N} t a$ good person nor her friends $good people$ receive case because both are headed by $\bar{N}$ predicate nominals ($\bar{N}$s) and only nonpredicative $\bar{N}$s require case. In Chapter 7, we adopt a raising to object account. However, similar arguments apply to both types of analyses. (In the raising to object account, the predicate nominal in (ii) is $t$ good people.)

(i) Mary was $\bar{N} t a$ good person]]

(ii) Mary considered $\bar{N}$ her friends $\bar{N} good people]]$

$That$ clauses may require case as evidenced by the fact that they passivize, as shown in (i) to (iv). In GB theory, the objects of actives correspond to the subjects of passives due to case considerations, as discussed in Chapter 2. There would be no reason for the $that$-clauses in (i) to (iv) to MOVE to subject position, unless they required case according to, for example, Chomsky’s (1991) last resort principle. See Postal (1974, p.56) for an argument that a passive rule cannot be formulated unless these $that$-clauses are NPs ($\bar{N}$s.) The argument presented here is similar in suggesting that $that$-clauses must be like $\bar{N}$ in that they require case.
(i) John assumed that Mary would win the prize.

(ii) That Mary would win the prize was assumed (by John).

(iii) John decided (three days ago) that Mary would win the prize.

(iv) That Mary would win the prize was decided (three days ago) (by John).

The distribution of subjects of raising verbs/adjectives in (v) to (viii) provide further evidence that (at least some) *that*-clauses require case.

(v) It is very likely that Mary will win the prize.

(vi) That Mary will win the prize is very likely.

(vii) It seems that Mary will win the prize is a true statement.

(viii) That Mary will win the prize seems to be a true statement.

Our Specifier-Case/Complement-Case distinction approximates a distinction made in GB theory (e.g., Chomsky 1981, 1986a,b, 1992, etc.) between nominative and accusative abstract case, although our Specifier-Case would include, previously noted instances of Genitive abstract case. We will assume these terms instead of the previous ones in order to avoid confusion between abstract case and morphological case terminology.

Ideally, we would find evidence that, e.g., a transitive verb together with its object is the case licensor for the Specifier-Case relation. For example, if there existed a hypothetical finite verb V taking two different sets of objects O1 and O2, such that V had different agreement properties depending on whether object O1 or O2, then V would provide the kind of evidence we are looking for.

Based on Pollock (1989), Chomsky (1991, 1992)/ Chomsky and Lasnik (1991) replace $\tilde{T}$ with the more complicated structure:

$$
\begin{array}{c}
\text{AGR-S} \\
\text{T-AGR-O} \\
\text{V} \\
\end{array}
$$
In the Chomsky/Chomsky and Lasnik approach: (1) each AGR element contains case and a set of \( \phi \)-features which include gender, person, number; and (2) Case assignment occurs at LF when the case recipient raises to the specifier of the appropriate AGR element to check that the \( \phi \)-features and case agree: the subject raises to AGR-S and the object raises to AGR-O.

We propose that case is a feature in Chomsky’s terms. We see no reason to differentiate it from gender, person and number. Chomsky (1992, p.10) suggests that case depends on properties of Tense. The reason that \( \bar{N} \)s receive case at AGR is that Tense raises to AGR (and combines with it). We find his arguments unclear, particularly with respect to: (1) whether the case assigned to both subjects and objects depends on tense, e.g., in languages with both subject and object agreement; and (2) why does case depend on properties of tense instead of, for example, agreement properties.

This section adopts the basic insight that case may be viewed as a type of agreement relation, but does not adopt the phrase structure features of the Chomsky/Chomsky and Lasnik analysis for reasons discussed in Chapter 7.

55 \( \bar{N} \) heads of small clauses need not be case marked as discussed in note 51.

56 Ray Dougherty (personal communication) notes that genitive case cannot represent items owned by inalienable possession, only items owned by the own relation. Thus, the following examples are ill-formed on the inalienable possession reading, but well-formed on the own reading.

(i) The arm of mine/John’s is next to me.

(ii) That arm is mine/John’s.

(iii) I consider that arm mine/John’s.

57 In GB literature, special considerations enter into the definition of adjacency for purposes of case assignment, i.e., CASE ADJACENCY.

Chomsky (1975a, 1992) and Larson (1988) maintain that sentences like *John gave Bob some money* have approximately the following structure:

\[
[\bar{\bar{\bar{\bar{N}}}} \bar{I} [\bar{\bar{\bar{\bar{I}}} N F L} [\bar{\bar{\bar{\bar{I}}} y} gave [\bar{\bar{\bar{\bar{N}}} Bob}]] [\bar{\bar{\bar{\bar{N}}} some money}]]]]
\]
The \( \bar{V} \) gave Bob assigns case and a theta role to the adjacent \( \bar{N} \) some money. See Chapter 7 for details.

58 The \( \bar{N} \) the big strong durable plastic bag is structurally ambiguous between a left recursion analysis like (3.89) and various conjunctive analyses like (i) or (ii). In the left recursive analysis, each adjective modifies a constituent to its left. In a conjunctive analysis an \( \bar{ADJ} \) consisting of the conjunction of some set of adjectives modifies the constituent to the left.

(i) \[ \bar{N} \left( \bar{N} \left( \bar{ADJ} \text{big, strong, durable, plastic bag} \right) \right) \]

(ii) \[ \bar{N} \left( \bar{N} \left( \bar{ADJ} \text{big, strong, durable} \left( \bar{N} \text{plastic bag} \right) \right) \right) \]

59 The analysis proposed here also applies to inanimate accusative \( \bar{N} \)s which contain these numerals probably because inanimate nouns have the same morphological form in the nominative and accusative. For purposes of exposition, we do not discuss this in the text.

60 The plural \( \bar{N} \)s containing numerals are actually ambiguously plural or singular. We could have represented this fact by stating a disjunctive value under the root Morph-Agr feature in the FSs above. However, we ignored this fact for the sake of simplicity. For analyses of the dual plural/singular role of these \( \bar{N} \)s see Pesetsky (1977, pp. 74-90), Corbett (1988), among others.

61 Josef Fioretta (personal communication) notes that the nominative plural adjective is more common, although both are acceptable.

62 Babby (1988) proposes an alternative analysis of this phenomena. He proposes that case is assigned to the \( \bar{N} \) and percolates down to the head. However, the following precedence hierarchy resolves morphological feature conflicts:

\[
\text{Lexical Case} > \text{Genitive} > \text{Nom/Acc}
\]

In GBUG, we could model this as subsumption relations between various morphological cases. The problem is that these subsumption relations could not hold generally, since, e.g., instrumental \( \bar{N} \)s cannot occur wherever nominative \( \bar{N} \)s can.
Another problem we find with Babby’s approach is that it does not attempt to account for the
conflicts in number agreement for numerals ending in 2 to 4. We deal with these conflicts below.

Note that the crucial part of our analysis for handling the data is the set of structure sharing N-Agreement2 arcs. These arcs percolate the morphological requirements the determiner imposes on the
noun down to the noun. Some theories/analyses would no doubt simplify our diagrams, removing, the
specifier and N-Agreement Licensor arcs with determiner trace values. In our account, these arcs and
their determiner-traces are required for explicitly stating which element is the N-Agreement licensor
(i.e. the controller of morphological agreement or the N-Agreement assigner) at each level of phrase
structure.

of feature percolation. Correa’s account suggests that chains of traces and sequences of percolation of
features (e.g. slash categories) may in fact be equivalent.

We use the term resolution rules generically to refer to any rules which determine the morpho-
logical agreement property of a conjunction of \( N \)s. In the work cited, resolution rules cover instances
in which all conjuncts are in the singular or in which clashes in morphological agreement features are
resolved in favor of defaults. We have extended the term to rules which resolve agreement clashes
based on the relative word order of conjuncts with respect to the agreement licensor.

Corbett cites Pravda, 7.11.1966 as the original source of our (3.97d) and he takes this example
from Graudina et. al. (1976, p.31). Corbet cites our (3.97e) from Solženicyn’s Arxipelag Galag.

Corbett (1983, p.100) cites (i), a Slovene example from Lenc’ek (1972, 59) as an instance in
which agreement conflicts are resolved in favor of the properties of the first conjunct in an \( N \) coor-
dinate phrase (another possible resolution rule). Apparently this is a relatively rare phenomenon in slavic
languages.
Corbett observes that masculine plural agreement is optionally found with coordinate phrases containing feminine conjuncts from the 0 declension class, e.g., the following example from Corbett (1983, pp. 189-190):

(i) Vezirova potpuna ukočenost, njegova ispravnost
vizar’s complete inflexibility (fem) his correctness (fem)
u novčanim pitanjima, i zatim Tahir-begova
in financial questions and then Tahir-beg’s
umešnost, umerenost i širokgrudnost
skillfulness (fem) moderation (fem) and generosity (fem)
u upravljanju zemljom, stvarali su . . .
in administering country created (masc-plur)
‘The vizir’s complete inflexibility, his correctness in financial questions and Tahir-beg’s skillfulness, moderation and generosity in administering the country created . . .’

In order to account for this fact in the approach discussed below, we allow 0 declension feminine nouns two options: (1) they can be treated as ordinary feminine nouns; (2) they can be treated the same as neuter nouns.

Corbett takes the examples in (3.98) from various literary work by Ivo Andrić.

Much can only be a determiner for mass nouns and many can only be a determiner for count nouns. Therefore, the examples in (i), (ii) and (iii) illustrate that conjoined food/chemical/etc. nouns are mass nouns, rather than count nouns. Either determiners is available in (i) because, either a mass or count N can occur in this context. However, the relative clauses in (ii) and (iii) force plural count noun and singular mass noun interpretations, respectively. Therefore the subject of (ii) can only agree with are, and the subject of (iii) can only agree with is.

(i) I ate too much/many pickles and strawberries.

(ii) The pickles and strawberries which I ate too many of are/*is in my stomach.
(iii) The pickles and strawberries which I ate too much of *are/is in my stomach.

71 We ignore here the distinction made between chains and CHAINS in Chomsky (1981,1986a) (among others). This term (capitalized) CHAIN subsumes the term chain. CHAIN also includes expletive argument pairs.

72 In GB literature the term head is used alternatively for ‘head of a phrase’ and ‘head of a chain’. To avoid this ambiguity we use the term C-head to refer to the head of a chain. For consistency, we use the term C-tail to refer to the tail of a chain.

73 For purposes of exposition, we combine the words to laugh into a single unit of category INFL-Verb. See Chapter 6 for a more thorough analysis of infinitives. Adopting that analysis would involve a bigger diagram in which: (1) to is an INFL element and laugh is a Verb element; (2) NP-trace$_2$ would be in the relation Specifier(to, NP-trace$_2$); (3) NP-trace$_2$ would bind NP-trace$_3$, the C-tail of the chain; (4) NP-trace$_4$ would be in the relations Specifier(laugh, NP-trace$_3$) and External-Theta(laugh, NP-trace$_3$).

74 In GB theory, the phrase “occupies a position” is a metaphor for participation in some constituent relation, i.e, location refers to a position in a constituent structure tree. Thus a case/theta position is a position in a constituent structure tree occupied by the licensee of some case/theta relation.

75 We do not investigate so-called LF movement in this dissertation. Characterizing a combined LF/S-structure representation without movement rules is an interesting topic for future research.

76 GB also assumes that modals and auxiliary verbs are instances of I, the head of I. However, in this section we only use I to refer to verbal inflection affixes.

77 Koopman distinguishes two types of V-chains: (1) V-chains, which are motivated by case considerations; and (2) V-chains, which handle long distance dependencies between verbs and verb traces. V-chains are motivated by similar considerations as A-chains, chains linking wh-traces and antecedents. We only discuss V-chains in this thesis. Most subsequent GB accounts about "verb movement" focus on V-chains, e.g., Chomsky 1986b, 1992, Rizzi 1990, Baltin 1991. (Also cf. Koopman
1987 regarding this topic.)

Based on Pollock (1989), Chomsky (1992) assumes that INFL is segmented into more than one phrasal projection representing different aspects of verbal inflection, and the verb moves to each one amalgamating along the way. Each agreement phrasal projection contains a specifier position so it can assign case to some \( \overline{N} \). We ignore the details of this analysis that relate to this segmentation in this section. We discuss Chomsky’s approach in more detail in Chapter 7.

In an earlier draft, we experimented with the idea of viewing INFL and V as distinct elements such that predicate and agreement relations held between INFL and V. Our intention was to motivate V-chains like A-chains as a linking between an agreement licensee position and a theta licensee position.

Affixes in a particular verbal paradigm only combine with a particular class of verbs. Violations of this requirement are ungrammatical in much the same way that agreement violations are ungrammatical. For example, English has at least two verbal paradigms for past tense inflection: the suffix \(-ed\) and a morphological rule changing the diphthong /ay/ to /ow/ (e.g. the past tense forms drove, rode, rose, wrote, corresponding to the verbs drive, ride, rise, wrote). (i) and (ii) are synonymous in meaning, although (i) is ill-formed. The type of ill-formedness of (i) is much the same as (iii), an agreement violation.

(i) (a) *Mary drived to work yesterday.

(ii) Mary drove to work yesterday.

(iii) *I are a good botanist.

This suggests a parallel between the fact that certain verbal inflections match verbs in certain paradigms and the fact that certain finite verbs agree with certain finite \( \overline{N} \)s.

Our selection restriction tests failed to provide a principled reason for believing that inflectional affixes select, i.e., predicate of the phrases they attach to. Attaching a verbal affix to a nonverb, as in (iv), appears to provide such an argument. A new word appears to be coined in (iv) by this affixation, forcing an interpretations of (iv) in which Mary gave everyone in the room an aspirin, coated their
bodies in aspirin powder, etc. However, as shown in (v) - (vii), it is always possible to force some interpretation on a real nonsense word. In these other environments, it is possible for the noun to force the affix to be part of a noun.

(i) %Mary aspirined everyone in the room.

(v) %Mary bought an aspirined

(vi) %The aspirined entered the room

(vii) %He seems aspirined

In summary, this approach failed to provide a principled reason for treating English verbs and verbal inflections as distinct syntactic elements.

80Original work in Arc-Pair Grammar, a framework which has since been renamed Meta-Graph Grammar, assumes a variety of relations defined on arcs, including Arc-command, a command relation defined between two arcs: a nominal arc and a predecessor of that arc (see Johnson and Postal 1980, p.257). LFG’s f-command (cf. Bresnan 1982a,b) is also a command relation defined on features (i.e., arcs).

81In this thesis, we assume that non-X-branching nodes do not exist, where X = all constituent licensor/licensee arcs. Condition (c) is necessary for this definition so that GBUG can describe (1) command relations in theories which allow nonbranching nodes in paths consisting of constituent licensor/licensee arcs; and (2) command relations with X set to other sets of arcs.

82Assuming a raising to object analysis, the $\overline{N}$ following the matrix verbs in (3.105) is governed by the verb under the Lasnik and Saito definition. The following infinitival complement in (3.105a) is an $\overline{I}$, just like complements of subject to subject raising verbs, as noted in Pullum and Postal (1988). The specifier of the following $\overline{X}$ in each example is an NP-trace, which is antecedent governed rather than head governed as per the definition of proper government in Lasnik and Saito (1984), Chomsky (1986b), etc.

Our decision to define command relations between arcs instead of between nodes as in previous GB accounts was necessary to model the analysis proposed in Chapter 6. Given the role of constituent licensor/licensee features in the GBUG model of GB theory, this decision was unproblematic for modeling other GB analyses. Furthermore, f-command is defined in LFG as a relation between features in a FS, i.e., as a relation between arcs. Thus our decision to model command relations as binary relations between arcs was required to model f-command as well.

In Chapter 6, we replace (most) empty categories with structure sharing constituent licensor/licensee arcs due to considerations of economy of representation, i.e. Occam’s razor (cf. Chomsky 1991). For example, in Figure 6.1 (from Chapter 6), our analysis of the passive, the \( \overline{N} \text{John} \) is simultaneously a specifier in one phrase and a complement in another. Assuming that c-command is a relation between arcs, a constraint can require the specifier arc to c-command the complement arc, i.e., the \( \overline{N} \text{John} \) in its role as specifier can c-command the \( \overline{N} \text{John} \) in its role as complement. This constraint would be our formulation of GB’s requirement that an antecedent c-command its trace.
Due to our analysis, defining command relations as relations between nodes would produce undesirable results (Ray Dougherty, personal communication). For example, the root node of the $\overline{N}$ John would c-command the entire sentence, where John represents both the NP-trace and the antecedent (cf. Chapter 6 for details). Defining command relations as relations between arcs allows us to distinguish the two roles John plays in our analysis.

See also Chierchia (1988) for a discussion of a Categorial Grammar command relation called F-command, which we would model as predicate command in GBUG. Chierchia proposes F-command as the basis for analyses of binding theory and contrasts the proposed categorial analyses with GB analyses of the same phenomena. Chierchia’s F-command makes the same predictions as HPSG’s o-command due to the hierarchical structure assumed in Categorial Grammar.
Perlmutter and Postal (1974) and Keenan and Comrie (1977) are credited with this hierarchy, which is assumed throughout the RG literature. See Ross (1974, p. 574) and Johnson (1977, pp. 165-173) for Relational Grammar treatments of reflexivization. Pollard and Sag’s o-command relation draws substantially on this work.

Our use of the term predicate structure differs markedly with its use in the framework of Williams 1980 and subsequent work. For Williams, predicate structure is a "level of representation" like D-structure, S-structure or Logical Form. We use the term predicate structure to refer to a FS with arcs representing predication, just like phrase structure is a FS with arcs representing constituency.

See Chapter 5 regarding the fact that Ns and arbitrary PRO can only be intersecting in reference, not coreferential.

The nature of the relation Predicate(Complementizer, $I$) is far from clear in GB theory. We use the feature Predicate-Licenser in Figure 3.63 in order to remain neutral on this issue. See Chomsky (1986b, pp. 14-15) for some discussion regarding why complementizers may not $\theta$-mark their $I$ complements.

Above, the Suppressed-Theta relation represents the predicate relation between a passive verb and the object of the passive by-phrase. Suppressed-Theta and External-Theta relations are mutually exclusive relations since, i.e., they are the external-theta roles which passive verbs can optionally assign. Examples like (i) could be accounted for if (3.109) is revised as follows:

(3.109)' Suppressed-Theta/External-Theta $<$ Internal-Theta $<$ Indirect-Theta $<$ Object in Modifier Phrase

(i) The pictures of [John and Mary]$_i$ were taken by each other$_i$

Pollard and Sag (1992, p. 317) note that Jackendoff (1972), among others, seek to constrain coreference constraints on reflexives by means of the following hierarchy of types of theta roles:

Agent $<$ Location/Source/Goal $<$Theme
Pollard and Sag claim that this approach should yield many of the same results as their own.

If following George (1980), we assume that verbs indirectly assign theta roles to prepositional objects, we require a different analysis. One possibility is a preposition incorporation analysis along the lines of Baker (1988, p. 254), while maintaining the binary branching structure discussed above. The predicate structure would be 3.65 below.

![A Predicate Structure representation of Mary described Bill to himself]

Following Baker’s analysis, in a full FS we would include a representation of a head chain connecting the preposition’s Incorporation position with its case-assigning position, and an A-chain linking the \( \bar{N} \) \textit{Bill} and an NP-trace, the NP-trace being a complement of the \( \bar{V} \) (where it receives a theta role in the diagram above) and \textit{Bill} as the complement of the \( \bar{V}^0 \) (where it receives case).

Pollard and Sag (1992, pp. 51-55) propose that adjuncts select their heads. In their section on binding, they do not made it explicitly clear how adjuncts fit in, e.g., they do not have any oblique examples for testing o-command. However, since they assume that modifiers do not appear on the
SUBCAT list for lexical items, HPSG binding theory does not effect modifiers containing anaphors.
We extend HPSG’s binding theory to cover modifiers as follows.

We assume that verb phrase modifiers modify and therefore predicate command at least the \( \overline{V} \) and perhaps even the \( \overline{V} \) in some cases. For example, the modifier *with a hammer* can modify the \( \overline{V} \) *break the window*, but not the \( \overline{V} \) *break the date*, as in

(i) (a) John broke the window with a hammer.
(b) %John broke his date with Sally with a hammer.

Examples (iia,b,c) are accounted for: the direct objects cannot be correferential with the objects of the prepositional phrases since the direct objects do not predicate command the modifying \( \overline{P} \)s. Unfortunately, this approach incorrectly rules out (iid). However, even that example is questionable since it is not clear whether the collocation possessive pronoun + *own* is an anaphor. These collocations have a slightly different distribution since they can only occur as the object of the preposition *of* or as the specifier of an *N*. In contrast canonical anaphors can occur as subjects and/or objects of verbs.

(ii) (a) *Hilda spanked the children* *in each other’s rooms*.
(b) *Hilda spanked the children* *next to each other/themselves*.
(c) *Hilda denounced [John and Mary] *in each other’s houses*.
(d) Hilda lectured [Mary and John] *in their own home*.

In contrast, anaphors in modifiers can commonly be coreferential with the subject of the matrix clause, as shown in (iii). This is accounted for assuming that the modifiers modify the \( \overline{V} \), which in turn predicates of the subject.

(iii) (a) Mary and John *were doing their homework* *in their own/each other’s rooms*.
(b) Mary and John *jumped up and down* [for each other’s/their own benefit].
(c) Wally *played the flute* [in his own rhythm].

Kracht provides a mathematical analysis of command relations which are binary relations between nodes in trees. We define command relations as binary relations between arcs in DAGs.
Though we have not studied the mathematical properties of our type of command relations, we strongly suspect that they have the same properties as the ones Kracht studied.

The following is a sketch of a strategy for proving that nodes-in-a-tree based definitions of command relations (node-command relations) and arcs-in-a-DAG based definitions of command relations (arc-command relations) are equivalent:

1. Show that every DAG D can be mapped into a tree T, such that: each arc A is represented as a node n; each pair of arcs A₁ and A₂ in D such that source(A₁) = target(A₂) map into nodes n₁ and n₂ in T such that n₁ is the source of an arc with a target of n₂; subgraphs of D rooted at targets of structure sharing arcs are mapped into two (or more separate) subtrees in T. For example, the DAG in Figure i would be mapped into the tree in Figure 3.60 (repeated below).

2. Show that each possible node-command relation defined on T gives the equivalent result as the corresponding arc-command relation defined on D.
A Constituent Structure FS Representation of
John saw Mary

Figure i
A Constituent Structure Tree Representation of
John saw Mary

Figure 3.60
CHAPTER 4

A Theory of Word Order Based on Predicate Licensing Relations

4.1 Introduction

This chapter describes the relative word order of a predicate licensor and its predicate licensee. Based on the pattern of exceptions to canonical word order, we demonstrate that word order is determined by the predicate licensor of a phrase. For example, modifiers in English canonically modify the constituent to their left as in (4.1a,b). Adjectives are an exception in that they canonically modify constituents to their right, as in (4.1c,d). Some exceptional adjectives modify to their left, as in (4.1e,f). These data show that (nominal) modifiers determine the word order of the phase they occur because, exceptions to canonical word order is caused by a change in modifier, but not by a change in modifiee. Further evidence that the predicate licensor determines word order comes from phrases containing two predicate licensors of the same type which occur on opposite sides of the same licensee. For example, left modifying and right modifying adjectives may occur in the same $\overline{N}$ in Spanish, as shown in (4.2), where prenatal ‘prenatal’ and most adjectives (and modifiers in general) modify to the left in Spanish and a few adjectives like buen ‘good’ are exceptional in that they modify to the right. This could not be accounted for if the noun determined the word order rather than the adjective.

(4.1) (a) a city [in Maine]
(b) the book [that I saw]
(c) the blond gentleman
(d) the blue marble
(e) the president elect
(f) eggplant parmigiana
Johnson, Meyers, and Moss (1993a,b) used a similar approach to word order in constructing an SFG (RG) Based Parsing System. In that project, all word order of a phrase is linked to surface Grammatical Relations represented as arcs. In each such set of grammatical relations, one arc is taken to be the functor, the arc which determines word order. Both that SFG approach to word order and the GB approach proposed here owe a debt to Categorial Grammar, where it is standardly assumed that functors determine word order.

We model our theory of word order in a WORD ORDER HIERARCHY a type subtype hierarchy in which word order parameters are set for types of predicate licensors, syntactic categories and individual lexical items. A word order parameter is set for each element in the hierarchy based on either (a) information associated with that element; or (b) information associated with its superclass. (b) is only checked if (a) is not provided. Thus both canonical and exceptional word order are accounted for at each level of the hierarchy. A word order hierarchy is an example of default inheritance hierarchy which no doubt has all sorts of uses in other areas of study. Default inheritance hierarchies have previously been used to associate linguistic rules to elements in a hierarchy based on parts of speech in Wilensky (1981), Flickinger, Pollard and Wasow (1985), Watanabe and Johnson (1990) and Johnson and Watanabe (1991).

In this section, our main contribution to GB theory are (1) importing the word order properties of the functor of Categorial Grammar to the predicate licensor of GB Theory; and (2) the integration of a default inheritance hierarchy into GB’s theory of parameter setting.
4.2 The Word Order Parameter

Following Johnson, Meyers and Moss (1993a,b), we assume a word order parameter with three possible settings:

(1) Free Order
(2) Predicate First
(3) Predicate Last

For a given predicate relation: the Free Order setting imposes no constraints on word order, the Predicate First Setting requires the predicate licensor to precede the predicate licensee, and the Predicate Last Setting requires the predicate licensor to follow the predicate licensee. For purposes of parameter setting, we 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.

The setting of this parameter translates into a partial order on the set of constituent licensor/licensee arcs representing the surface positions of constituents in active declarative sentences—we call these arcs SURFACE ARCS. The surface arc corresponding to a particular predicate licensor/licensee arc is determined as follows: if the licensor/licensee is an empty category in some chain C, the constituent licensor/licensee arc at the C-head of C; otherwise, the surface arc is a structure sharing constituent licensor/licensee arc.

For example, in Figure 4.1 the path *Adjunct Head-Proj* is ordered before the path *Adjunct Complement* because the Internal-Theta relation is set to Predicate First. Also in Figure 4.1, the root Adjunct arc is ordered before the root Head-Proj arc because the modification relation is set to Predicate First. Each of these constituent licensor/licensee arcs structure share with the predicate licensor/licensee arcs which represent the relevant predicate relation. Similarly, in Figure 4.2 the path *Specifier* is ordered before the path *Head-Proj Complement Head-Proj* because the specifier (the matrix subject) is the C-head of the A-chain \([N, NP\text{-}trace]\), the NP-trace is the External-Theta licensee and the External-Theta relation is set to Predicate Last.
The word order parameter, as presented here, 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 include Sanders (1967, 1974), Johnson (1974), Johnson and Postal (1980) and Gazdar, Klein, Pullum and Sag (1985).
4.3 Predicate Licensors and the Word Order Hierarchy

In this section, we present two types of evidence that the predicate licensor rather than some other constituent determines word order: (1) exceptions to canonical word order results from changing the predicate licensor in a phrase; and (2) nonhead predicate licensors of the same type which have different word order properties may occur in the same phrase, predicing of (different projections of) the same head. These facts are not accounted for if some constituent other than the predicate licensor, e.g., the head of the phrase, determines the word order.

Based on these findings, we derive the word order hierarchy in Figure 4.3. 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 default setting of its superclass is assumed.

<table>
<thead>
<tr>
<th>Word Order Hierarchy for English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default Setting (Predicate First)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Degree</td>
</tr>
<tr>
<td>Quantification</td>
</tr>
<tr>
<td>Predicate Last</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.3**
The Word Order Hierarchy states that English is predominantly a predicate first language, and that, canonically, Internal-Theta, Degree Licensing and Quantification relations follow this default. For example all the Internal-Theta relations in (4.3), the degree relations in (4.4) and the quantifier relations in (4.5) are predicate first. Therefore the word order parameter need not be set for the general classes of items which license these relations. However, as shown in (4.6), (4.7) and (4.8), there are a few specific lexical items which are exceptions to the parameter settings for these licensing relations. Thus the word order parameter must be set for these lexical items as follows: (1) Free Order for the quantifiers More and less; (2) Predicate Last for the degree word enough; (3) Predicate last for the adpositions ago and hence; and (4) either the default setting or Predicate last for the polysemous adpositions before, after and late, depending on the sense intended. We assume that ago, hence, before, after and late are adpositions because phrases consisting of one of these words and an \( \overline{N} \) can occur in environments which \( \overline{P}s \) can occur, as shown in (4.9a-d). Another possible hypothesis is that ago, hence, before, after and late are adjectives modifying the adjacent nouns. However, \( \overline{N}s \) cannot occur in the same environments as these phrases as shown in (4.9e,f); and these phrases cannot occur in the same environments as \( \overline{N}s \), as shown in (4.9g,h). Therefore these lexical items are not adjectives in the context of (4.8) and (4.9) (late is an adjective in other contexts, e.g., the late physicist.

(4.3) (a) Mary [saw John]
(b) Cecil [said that he was a gorilla]
(c) in the house
(d) for five years
(e) Matilda was [angry at Clyde]
(f) Matilda was [annoyed that Clyde ate the last cookie]

(4.4) (a) You can’t be [too angry] with the fish merchants
(b) the [very uncomfortable chair]
(c) He was [quite hungry]
(4.5) (a) The dog
   (b) One dog
   (c) A dog
   (d) five dogs
   (e) Every dog

(4.6) (a) one/an inch more
   (b) one/an inch less
   (b) one/*a more inch
   (b) one/*a less inch

(4.7) (a) You could not be [angry enough] at those fish merchants
   (b) You weren’t [quite angry enough] at those fish merchants

(4.8) (a) five years ago/hence
   (b) before the day of the party
   (c) the day before/after
   (d) *[the day of the party] before/after
   (e) a day late
   (f) two days late

(4.9) (a) John left five years ago
   (b) John will leave five years hence
   (c) John left the day before/after
   (d) John left a day late
   (e) *John left the/a day
   (f) *John left/will leave five years.
   (g) I lived for five years (*ago)
   (h) The baby will be born in three months (*hence)
The word order parameter is Predicate Last for the External-Theta relation in English— to our knowledge there are no exceptions to this setting. This word order parameter must be explicitly set in the grammar for English because Theta has a Predicate First setting and Theta is the immediate superclass of External-Theta in the Word Order Hierarchy. The External-Theta Relation is also set to Predicate Last in Spanish. However, some External-Theta Licensors allow Free Word Order, as shown in (4.10), (4.10’), (4.11) and (4.11’). The examples in (4.11) and (4.11’) only allow free word order with pronominal objects.

(4.10) (a) El baila

He dances

‘He dances/is dancing’

(b) Yo pienso

I think

‘I think’

(4.10’) (a) Baila el

(b) Pienso yo

(4.11) (a) Los libros [me gustan]

The books me please (third-person-plural)

‘Books please me’ or ‘I like the books’

(b) El libro/bailar [me encanta]

The book/dancing me enchants/loves (third-person-plural)

‘The book/dancing enchants me’ or ‘I love the book/dancing’

(c) Los libros [me facinan]

The books me fascinate (third-person-plural)

‘The books fascinate me’
(d) Usted [me trae] el almuerzo

You me bring (third-person-singular) the lunch

‘You bring/are bringing me lunch’

(e) El trabajo [me fatiga]

The work me fatigues (third-person-singular)

‘The work fatigues me’

(f) [Le fatiga] el trabajo a mi hermano.

Him fatigues (third-person-singular) the work to my brother

‘The work fatigues my brother’

(4.11’) (a) [Me gustan] los libros

(b) [Me encanta] el libro/bailar

(c) [Me facinan] los libros

(d) [me trae el almuerzo] Usted

(e) [Me fatiga] el trabajo

(f) A Mi hermano le fatiga el trabajo

The Word Order Parameter must be set to Predicate Last for modification in English, since this setting is in conflict with the default. $\bar{P}$s, relative clauses, purposes clauses and other modifiers follow this default, as shown in (4.12). The Word Order Parameter must be set to: (1) Free Order for Adverbial modifiers of verbs, as shown in (4.13); and (2) Predicate First for adjectives and adverbial modifiers are adjectives, as shown in (4.14).

(4.12) (a) the book $\bar{P}$ with the bright orange cover

(b) The astronaut jumped rope $\bar{P}$ on the moon

(c) the book [that I saw t]

(d) I ran the race [Pro to prove that I am an athlete]

(4.13) (a) I walked slowly; I slowly walked

(b) I ate hungrily; I hungrily ate
(4.14) (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) Predicate First for a few idiomatic $\overline{P} s$, as in (4.15); (2) Free Order for the adjective nearby, as in (4.16); and Predicate Last for the adjectives elect, general, public, parmigiana, almondine, across, above and agog in the particular senses used in (4.17). We assume that the words nearby and above have adjective senses because: (4.18a,b) show that these words and the adjacent noun form $Ns$, as they can occur in environments which require $N$s. (4.18c,d) shows that nearby and above cannot be adpositions when occurring after the noun because the resulting phrase cannot occur as the second complement of put even though it is clearly locative in meaning. 2

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

(4.16) the nearby lake; the lake nearby

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

(4.18) (a) Mary went to the lake (nearby)
(b) John watched the sky (above)
(c) Mary put the fish [above the shelf]/[*the shelf above]
(d) Mary put the fish [near the lake]/[nearby the lake]/[*the lake nearby]
In English, most of the adjectives with exceptional word order properties 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* in (4.19a), but some adjectives must occur predicate first, e.g., *buen, bajo* and *grande* in (4.19b,c,d). Others like *importante, malvada* and *cruel* in (4.19e,f,g,h) can modify the noun from either side.

(4.19)  (a) el libro rojo
        the book red
        ‘the red book’

(b) buen cuidado prenatal
    good care prenatal
    ‘good prenatal care’

(c) Al bajo costo
    at low cost
    ‘at low cost’

(d) el gran autor/monumento/evento
    the great author/monument/event
    ‘the great author/monument/event’

(e) la importante noticia; la noticia importante
    the important notice
    ‘the important notice’

(f) la malvada criminala; la criminala malvada
    the wicked criminal
    ‘the wicked criminal’

(g) el cruel criminal; el criminal cruel
    the cruel criminal
    ‘the cruel criminal’
The data above motivates the Word Order Hierarchy for English which repeat below. The most important result is that exceptional word order properties were associated with the predicate licensor, rather than some other constituent, e.g., the head of the phrase. Exceptions to canonical parameter settings always came about from changing the particular type of predicate licensor, the particular syntactic category of the predicate licensor and even the specific lexical item playing the role of the predicate licensor. Exceptional word order was never associated with the predicate licensee in the predicate licensing relations we discussed above.

<table>
<thead>
<tr>
<th>Word Order Hierarchy for English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default Setting (Predicate First)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Degree</td>
</tr>
<tr>
<td>Quantification</td>
</tr>
<tr>
<td>Predicate Last</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Predicate First</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Free Order</td>
</tr>
</tbody>
</table>

Further evidence that it is the predicate licensor that determines word order come from the examples in (4.20). 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. We conclude that the predicate licensor determines word order.

(4.20) (a) a great president elect (but a mediocre president)

(b) the disgusting veal parmigiana
(c) quite angry enough

(d) three miles more

(e) buen cuidado prenatal (= 4.19b)

In English, we know of at least one potential counterexample to our theory of word order: Compound indefinite pronouns, unlike other nouns, can only be modified by following adjectives, as in (4.21) (David Johnson, personal communication). In these examples, the predicate licensee determines exceptional word order. Therefore these examples appear to be problematic for our account.³

(4.21) (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\}. We assume that the pronouns in (4.21) are simultaneously quantification licensors and modification licensees. As quantifiers, they predicate of the $\tilde{N}$ to the right and as modifiees of adjectives, they would normally occur to the left of the adjective. Since the adjective is part of the quantified $\tilde{N}$, the pronouns must simultaneously occur both before and after the adjective. This is impossible, so one predicate relation must have priority. We assume that the quantification relation takes priority because it is ‘higher’ in the predicate structure, i.e., the quantification-licensee arc predicate commands the modification licensee arc in Figure 4.4.⁴
4.4 Previous GB Accounts of Word Order

Our default inheritance hierarchy (the Word Order Hierarchy) contrasts with the two main ways researchers claim that parameters are set in GB theory. Stowell (1981), Travis (1983), among others claim that a parameter is set for an entire language. This does not permit exceptions. Wexler and Manzini’s (1985) in their work on Binding Theory Domains argue that parameters are set for particular lexical items. This misses 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 (but see above.)

Stowell (1981) 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 our 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 the post-
positions discussed above.

Travis\(^5\) (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 pro-
poses that: (1) the verb must (immediately) follow INFL (the head of \(I\)) 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 (and therefore 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,\(^6\) clearly *-s* follows *laugh*. Therefore Travis’s account provides no explanation
of why the string *Laughs John* is ungrammatical. Furthermore, even if we assume (counterfactually),
that there is some sense in which a verbal affix precedes the verb, this would not explain why some
Spanish verbs allow the subject to follow the finite verb, e.g. the examples in (4.10’) and (4.11’) which
we repeat below. Spanish, like English, requires strict adjacency for case assignment. Therefore, if
Travis’ account were correct, we would expect the following data not to occur.

\[(4.10’)(a)\] Baila el
\[(b)\] Pienso yo

\[(4.11’)(a)\] [Me gustan] los libros
\[(b)\] [Me encanta] el libro/bailar
\[(c)\] [Me facinan] los libros
\[(d)\] [me trae el almuerzo] Usted
\[(e)\] [Me fatiga] el trabajo
\[(f)\] A Mi hermano \(_i\) le \(_i\) fatiga el trabajo
Independent of the above considerations, Travis’ account of word order fails to account for the external-theta relation in SOV, VSO and VOS languages or in languages with no case adjacency requirement.

While we assume that case adjacency is not relevant to subject verb order in English, it may be relevant to the order of complements. For example, we follow GB theory in assuming that if a verb takes one $N$ and one $P$ complement, the $N$ complement occurs first, barring considerations like heavy NP shift, as shown in (4.22).

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

Also, I assume GB accounts of variation from canonical order which result from some morphological or syntactic operation, e.g., passive, heavy NP shift, topicalization, Koopman’s (1983) account of word order variation in languages like Vata, etc.

In Summary, our account provides a principled description of a wider range of word order phenomena than previous GB accounts. We have shown that Travis’ account of the relative order of subjects and finite verbs is 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 Conclusion

Therefore, assuming that word order parameters are set for types of predicate 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 of a constituent can lead to variations in word order, but varying the head does not, unless the head is the also the predicate; and
(2) In phrases with one head and more than one predicate of the same type, the predicates may have different relative orders to the head. The head cannot determine word order in these phrases.

Our position rests on an account of systematic exceptions to canonical settings of word order parameters. In contrast, previous GB theories of parameter setting either: (1) set parameters for the entire language, thus not accounting for any exceptions; and (2) set parameters for individual lexical items, thus not accounting for canonical word order.

Our account has the following implication for the content and structure of the lexicon: the word order parameter should be set for the predicate licensor of a phrase according to the position of that lexical item in the word order hierarchy, unless that word is exceptional with respect to word order.
Notes

1 Johnson, Meyers and Moss (1993a,b) describe a bottom-up chart parsing algorithm which operates on FSs like the ones discussed in this section.

2 Maling (1983) argues that *near* functions syntactically and morphologically like an adjective, even when *near* takes an $\bar{N}$ object like a preposition. As evidenced below, *near* can simultaneously (1) take an $\bar{N}$ object like a preposition and (2) take comparative and superlative suffixes like an adjective or be predicated of by the degree licensors *very, how, so* and *too*, all of which normally only predicate of adjectives.

(i) Mary shot the arrow much nearer the target than John.

(ii) How near the target did Mary shoot the arrow?

(iii) Mary shot the arrow very/too near the target.

(iv) Mary shot the arrow so near the target that everyone fainted.

Maling proposes that *near* is a transitive adjective. Maling also shows that *like* and *worth* have many of these properties as well.

When *nearby* precedes the $\bar{N}$ it has basically the same properties as *near*. However as noted above, *nearby* behaves only like an adjective when it follows the object.

3 Spanish object pronouns, often called clitics, occur to the left of the verb assigning them an internal theta role, as in (4.11) above. The Internal-Theta relation is canonically Predicate First in Spanish. Thus these pronouns appear to be a problem for our theory of word order.

However, previous GB accounts of clitics do not derive surface positions by parameter setting. Indeed, the surface positions of Spanish clitics would violate the head parameter (see below.) Rather, GB 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 there final surface positions depend on a variety of factors. There is a large literature on clitics which we will not attempt to review here. However, we assume compatibility with previously proposed ana-
Therefore our theory of word order is no worse off than previous GB analyses with respect to deriving the relative word order of clitics and verbs.

The fact that these indefinite nouns can only be modified by a single adjective (except for coordinated adjectives) is not surprising on this account. We assume that the lexical entry for these indefinite pronouns includes the position for the adjective modifier.

Travis’ account is based on a 1982 class lecture by Noam Chomsky.

We 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. We do however assume GB’s hierarchical relation between phrases based on the distinction between certain affixes, e.g. INFL and verbal stems.
CHAPTER 5

A GBUG Characterization of Coreference

5.1 Introduction

This chapter develops the GBUG representation of coreference based on an innovation of Pollard and Sag (1992). This is a necessary preliminary to further discussions of Trace theory and Binding Theory.

Previous GB accounts represent coreference in constituent structure trees, by coindexing elements which have the same reference. GB theory also uses "coindexing" notation to represent: morphological agreement (e.g., Chomsky 1981) and predication (e.g., Williams 1980), as shown in (5.1).

(5.1) (a) Mary, promised herself, that she would be a linguist some day. (coreference)

(b) [I, Mary, [I, AGR, promised herself . . . ]] (morphological agreement)

(c) Mary, [promised herself . . . ] (predication)

The fact that these very different phenomena are all represented by coindexing has sometimes obscured the fact that the phrase "is coindexed with" is a disjunction of the relations "is coreferential with", "morphologically agrees with" and/or "is predicated of". Coreference and agreement relations are both integral parts of the definitions for governing category found in Chomsky (1981, 1986a), among others. In these definitions, stating that \( \alpha \) is coindexed with \( \beta \) means that \( \alpha \) and \( \beta \) are either (a) coreferential or (b) are in a morphological agreement relation, namely subject INFL agreement,\(^1\) i.e., it is equivalent to define the term coindexing to mean "either coreference or morphological agreement" or to explicitly state the disjunction. If "coreference" and "agreement" were shown to have some wide range of traits in common, there would be some reason for assigning them to a natural class. However, this has never been shown to obtain for agreement and coreference relations (or for predicate and coreference relations). Statements like "X holds iff A and B are coindexed" are misleading because they give the impression that a generalization has been made even if the writer has merely stated a
disjunction of unrelated properties. In Chapter 3, we have motivated GBUG-based alternatives to coindexing for representing morphological agreement and predicate relations. This chapter motivates an GBUG-based alternative for representing coreference relations.

5.2 Coreference in GBUG

Figure 5.1 represents coreference between two FSs: FS$_1$ with root R$_1$ representing a pro-form or an empty category, and FS$_2$ with root R$_2$, the antecedent of FS$_1$. Coreference is expressed as structure sharing between: (1) an ANTECEDENT arc with source R$_2$ representing the capacity of FS$_2$ to bind or be coreferential with a pro-form or empty category; and (2) a COREFERENCE arc with source R$_1$ representing that FS$_1$ may be coreferential with some antecedent. This mode of representation is based on Pollard and Sag’s (1992) representation of coreference in the linguistic framework HPSG.

<table>
<thead>
<tr>
<th>Coreference in GBUG</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS$_1$ Coreference: [ ]</td>
</tr>
<tr>
<td>•</td>
</tr>
<tr>
<td>•</td>
</tr>
<tr>
<td>•</td>
</tr>
<tr>
<td>FS$_2$ Antecedent: [ ]</td>
</tr>
<tr>
<td>•</td>
</tr>
<tr>
<td>•</td>
</tr>
<tr>
<td>•</td>
</tr>
</tbody>
</table>

Figure 5.1

GBUG distinguishes two varieties of coreference:

1. OVERLAP-REFERENCE: The FS rooted at the source of the overlap-reference arc overlaps in reference with the FS rooted at the source of any structure sharing antecedent arc.

2. SAME-REFERENCE: The FS rooted at the source of the same-reference arc has the same reference as the FS rooted at the source of the structure sharing antecedent arc.
These two features encode a distinction noted in Chomsky and Lasnik (1991, p.65) and elsewhere between two types of coreference. In the same-reference relation, the coreferential items clearly refer to the same individual. For example in (5.2a) represented as the FS in Figure 5.2, *John* and *himself* identify the same individual. Since the same-reference relation is obligatory, (5.2a) and Figure 5.2 represent all of the possible readings of the described sentence. In the overlap-reference relation, the pro-form refers to some set of individuals which on at least one reading, includes the coindexed individuals. For example, in (5.2b) represented as the FS in Figure 5.3, *they* refers to a group which, inclusively or exclusively includes John and Mary. Since coreference with pronouns is usually optional, (5.2b) and Figure 5.3 represents a subset of the possible readings of the sentence described, i.e., readings in which *John* and/or *Mary* do not overlap *they* in reference are also possible.

(5.2) (a) John\textsubscript{i} saw himself\textsubscript{i}

(b) John\textsubscript{i} told Mary\textsubscript{j} that they\textsubscript{i,j} should leave

(Chomsky and Lasnik’s 1991 example (209i))

Our notation is more explicit than that of Chomsky and Lasnik (1991), who use the notation “coindexing” for all pro-form antecedent relations with the proviso that the coindexing notation has a different interpretation for pronouns and anaphors:

(a) Suppose NP and \( \alpha \) are coindexed. Then

(i) if \( \alpha \) is an anaphor, it is coreferential with NP;

(ii) if \( \alpha \) is a pronoun, it overlaps in reference with NP.

(b) Suppose NP and \( \alpha \) are contra-indexed. Then they are disjoint. (Chomsky and Lasnik 1991, p.65)

In contrast, our representation explicitly states the relevant details—the reader does not have to know that (s)he must interpret the same notation two (or more) different ways depending on the context. We explicitly distinguish same-reference and overlap-reference relations while recognizing that they are each a type of coreference. This permits different constraints for: (1) coreference, a relation which subsumes both same-reference and overlap-reference; (2) same-reference; and (3) overlap-reference. Same-reference and overlap-reference are used to represent the referential properties of pro-forms and empty categories. Same-reference, at times, even has the same interpretation as overlap-reference,
e.g., in sentences like (5.3a-c) they can (but need not) refer exclusively to coindexed $N$s. These two facts follow from our assumption that same-reference and overlap-reference form a natural class which we call COREFERENCE. If a pro-form X has a coreference arc which does not structure share with the antecedent arc of some $N Y$, then X and Y are disjoint in reference. An Independent constraint (Binding Theory Principle B) determines the ability/inability for reference to be possible. The role of a FS is to represent a particular reading of a sentence, e.g., a reading where overlap reference obtains or a reading where it does not.

(5.3) (a) [John and Mary]$_i$ saw a beautiful picture of them$_i$.

(b) [The boys]$_i$ said that they$_i$ were very tired.

(c) People$_i$ never seem to know what they$_i$ are doing.

**Figure 5.2**

A FS representing

*John saw himself*

```plaintext
| Category: [I] | INFL |
| Tense: [J] | Past |
| Specifier: [K] | Category: Noun |
| Antecedent: [I] |
| Head-Proj: Category: [I] |
| Tense: [J] |
| Complement: [m] Category: [n] Verb |
| Specifier: [o] Category: NP-trace |
| Same-Reference: [I] |
| Antecedent: [p] |
| Head-Proj: Category: [n] |
| Head-Proj: Category: [n] |
| Phonology: saw |
| Complement: [q] Category: Noun |
| Phonology: himself |
| Same-reference: [p] |
| Internal-Theta: [q] |
| Complement-Case: [q] |
| External-Theta: [e] |
| Internal-Theta: [m] |
| Specifier-Case: [e] |
```

255
A FS representing one reading of

John told Mary that they · · ·

Many GB theorists represent coreference between empty categories and their antecedents using indices, as above. To my knowledge NP-trace, WH-trace and pro do not ever take split antecedents. These empty categories could only take split antecedents if there were movement (NP-trace/WH-trace) or agreement (pro) relations between an empty category and two separate constituency positions. It would have to be possible for NP movement to involve one case and two theta positions; for a wh operator to bind two theta positions; or for some language to combine subject and object agreement in a single empty category. As far as we know linguistic phenomena making such analyses feasible simply do not arise. However, as discussed in Williams (1980), PRO varies in this respect: obligatorily controlled PRO (OC-PRO) is of type same-reference (does not allow split antecedents) and
nonobligatorily controlled PRO (NOC-PRO, also called PRO\textsubscript{arb}), is of type overlap-reference (does allow split antecedents) as evidenced by the possibility of the indexing showed in (5.4) and the impossibility of the indexing showed in (5.5).

(5.4) (a) John\textsubscript{i} told Mary\textsubscript{j} that it would be appropriate PRO\textsubscript{i,j} to leave together.

(Williams 1980, p.218, no. 66)

(b) Dr. X\textsubscript{i} asked Dr. Y\textsubscript{j} about PRO\textsubscript{i,j} prescribing themselves/each other\textsubscript{i,j} an antispasmodic.\textsuperscript{6}

(5.5) (a) *John\textsubscript{i} enticed Mary\textsubscript{j} into PRO\textsubscript{i,j} leaving together.

(b) *John told Mary to prescribe themselves some aspirin.

Williams ascribes the following properties to OC PRO: (1) a lexical NP cannot occur in the position of PRO; the antecedent is obligatory; (2) the antecedent precedes and c-commands PRO; and (3) the antecedent is uniquely controlled (i.e. no split antecedents) thematically selected by the verb, as in (5.6a) or grammatically determined as in (5.6b) (Williams 1980, examples 25c-d.) In contrast, Williams ascribes the following properties to NOC PRO: (1) antecedents of PRO are optional as in (5.6c) (2) they need not c-command PRO; (3) they needn’t be unique, i.e., they can take split antecedents; (4) they may follow the clause containing PRO as in (5.6d); and (5) a lexical NP may occur in place of PRO as the subject of a for infinitive or an exceptional case marked complement, e.g., (5.6e).\textsuperscript{7}

(5.6) (a) John tried [PRO to leave]

(b) They died [PRO waiting for a bus]

(c) Its a good idea [PRO to leave]

(d) PRO leaving is something Mary should do

(e) It was arranged (for Mary) to have dinner with Cecil

Koster (1984) observes that OC-PRO patterns with anaphors with respect to the properties discussed above and NOC-PRO patterns with pronouns. Thus Koster rejects the idea that PRO is simultaneously a pronoun and an anaphor; rather he proposes that OC-PROs are anaphors and NOC-PROs are pronouns. Among other things, Koster argues that OC-PROs are governed, contrary to the PRO theorem of Chomsky (1981). We do not investigate these matters further here. What is important to
our current discussion is that only NOC-PROs may take split antecedents. Thus NOC-PROs, like pronouns participate in the overlap-reference relation, but OC-PROs, like anaphors participate in the same-reference relation. (See Postal 1970, pp. 476-478, for a more detailed account of the binding properties of the subjects of nonfinite clauses in examples like the ones above.)

5.3 Coreference in Pollard and Sag (1992)

Pollard and Sag (1992, pp. 12-13) (PS92) uses arcs labeled INDEX to indicate reference (as this term is used in Discourse Representation Theory). The value of an index arc in PS92 is a FS representing agreement properties (person, gender, number.) Coreference (or binding) is represented via structure sharing between index arcs of pronouns, anaphors, etc. and their antecedents, as shown in Figure 5.4. In order for structure sharing to take place, the values of the index arcs must be unifiable and thus agree in person, number and gender.

PS92’s representation of coreference differs with the one outlined above, in that PS92 do not make the overlap-reference/same-reference distinction. A GB representation of John told Mary that they should leave using HPSG’s index feature would be same as Figure 5.3 except all antecedent and overlap-reference arcs would be replaced by index arcs—this would incorrectly mark John, Mary and
they as all sharing the same reference. One alternative is to allow plural pro-forms more than one index arc (as suggested by the multiple indexes i and j used in the text) and then follow Chomsky and Lasnik’s suggestion that indexes have different interpretations for anaphors and pronouns. The resulting FS representation would not be functional (and therefore ill-formed) because two arcs with the same source would bear the same label. Unlike the complement arc and theta arc examples, we cannot simply number the indexes $\text{Index}_1$ to $\text{Index}_n$, because plural pronouns can represent groups with a potentially unlimited number of members (suppose they refers to the group of people who have and ever will have odd social security numbers.)

PS92 require coreferential phrases to share agreement features. This runs into some of the same problems as using FS unification for agreement in general, as shown in (5.7a,b,c) (See Section 3.4 for details). Additionally, the coordinate conjunction or does not seem to work very smoothly with reflexives, as shown in (5.7d,e,f). Putting these factors aside, Pollard and Sag’s index feature works fairly well for our same-reference relations. Clearly, it is worthwhile to constrain the coreference relation so that morphological agreement is maintained in most cases.

\[(5.7) \text{(a) The group}_i\text{ destroyed itself/themselves}_i.\]
\[(b) 40 \text{ acres}_i\text{ practically plows itself}_i.\]
\[(c) [John and Mary]_i\text{ consider themselves}_i\text{ lucky.}\]
\[(d) ??[Either John or the children]_i\text{ buy themselves}_i\text{ flowers every day.}\]
\[(e) *[Either the children or John]_i\text{ buys himself}_i\text{ flowers every day.}\]
\[(f) ?[Either John or Bob]_i\text{ considers himself}_i\text{ lucky.}\]

The Overlap-Reference relation cannot be modeled by HPSG’s index relation because antecedents which overlap in reference with a pro-form do not necessarily agree with that pro-form. For example, in (5.2b) (repeated below), neither John, nor Mary is plural and John and Mary do not form a plural $\overline{N}$ constituent. In examples like (5.8), readings are possible in which the pronouns and the indexed antecedents refer to precisely the same individuals. We assume that the pronoun/antecedent agreement relations that appear to hold in (5.8) are semantic in nature, not syntactic. Our account of
split antecedent cases like (5.2b) follow from this assumption.

(5.2b) John \_i \text{ told Mary}_j \text{ that they}_i \_j \text{ should leave}

(5.8) (a) Mary \_i \text{ said that she}_i \_i \text{ would deliver the ice cream.}

(b) [John and Mary]_i \text{ couldn’t figure out how they}_i \_i \text{ were going to afford}

enough pickles for the convention.

5.4 Conclusion

We have presented GBUG’s representation for pro-form/antecedent relations and empty category/antecedent relations. GBUG coreference relations encode the distinction between same-reference and overlap-reference relations. In Chapter 6, we motivate the elimination of all empty category analyses which can be represented in terms of the same-reference relation. However, we are not able to use our alternative analysis for NOC-PROs, those empty categories which currently require an overlap-reference analysis.
Notes

1Chomsky (1981) specifically identifies the agreement relation between an English finite verb and its subject, a morphological agreement relation, as being the relevant relation for binding theory. He distinguishes this agreement relation from other instances of abstract case in which morphological agreement is not realized.

2To our knowledge, the use of coindexing for predication has never resulted in any misleading statements. In Williams (1989), it is proposed that theta roles are linguistic objects which can be coreferential with theta role recipients (theta licensees in our terms) and that this coreference relation is subject to binding theory. Ultimately Williams adopts a "linking" notation instead of coindexing for representing theta role assignment, which he claims is an asymmetric relation, in contrast with pro-form coreference, which he claims is not.

Williams is careful to specify the way in which he claims that theta relations and the pro-form coreference are of a kind, i.e., they are both subject to binding theory. However, he never attempts to interpret a heterogeneous set of indices $S$ as being homogeneous, where some indices in $S$ represent theta relations and others represent pro-form coreference relations. In contrast, some accounts of Binding Theory (e.g., Chomsky 1981, 1986a) consider agreement indices and coreference indices to be of a kind, e.g., AGR, an abstract feature of verbal morphology, is viewed as a potential antecedent of reflexives and reciprocals.

3In previous chapters, we have used same-reference and antecedent arcs to represent trace antecedent relations. This chapter provides a more complete definition of our coreference mechanism and places the mechanism represented by these arcs in context.

4One interesting detail missing from the Chomsky and Lasnik (1991) approach to intersecting reference (overlap-reference) is that only plural pronouns have this option. Note also that PRO (as discussed below) may also take split antecedents under certain circumstances—in this case it is possible because PRO is unspecified for number. However, singular pronouns, when coindexed with their antecedents refer exclusively to those antecedents.
5After I completed this dissertation, Paul Postal (personal communication) pointed out that some
wh-traces have properties which are hard to characterize with same-reference relations. For example,
consider the following sentence (Postal’s example).

(i) [Which two girls], did respectively Frank kiss t and Mike hug t?

(i) means that Frank kissed one girl and Mike hugged the other. Thus the antecedent of each
trace can be viewed as a different portion of the wh phrase. A more complex representation would be
required to represent this sentence and (relevant readings of) similar sentences. Below, I assume that
wh-traces can be represented by structure sharing. However, the existence of sentences like (i) miti-
gates against the viability of that assumption. In future work, I will examine wh extraction more
closely than I have done in this thesis and attempt to address this concern.

6I would like to thank Paul Postal for assistance in identifying examples of split antecedents for
PRO.

7These criteria lead Williams to argue that verbs like want, hate, prefer, arrange, among others
which alternate between subject control and For-to or exceptional case marking complements are actu-
ally verbs of non-obligatory control. Williams claims that these verbs select for human subjects which
are available controllers, but passives of these verbs (some, like want and the sense of hate taking an
infinitival complement don’t passivize) clearly exhibit properties of nonobligatory control. Williams
gives the following examples (his 51a-b):

(i) It was arranged to have dinner at 6

(ii) It was preferred to have dinner at 6
CHAPTER 6

What do Empty Categories Represent?

6.1 Introduction

This chapter motivates replacing each chain consisting of one antecedent and one or more ecs bound by a same-reference relation with a set of constituent licensor/licensee arcs which have the same value, i.e., we revise our previous definitions of chains (see Chapter 3) as sets of structure sharing arcs. For example Figure 6.1 is our analysis of the passive sentence *John was tickled by the gorilla*, the significant feature being that *John* is the value of each arc in an A-chain consisting of the paths: Specifier, Head-Proj Complement Specifier, and Head-Proj Complement Head-Proj Complement, where Specifier is the C-head and Head-Proj Complement Head-Proj Complement is the C-tail. We call Figure 6.1 and other similar analyses CONSTITUENT SHARING (CS) analyses because their crucial feature is the structure sharing of constituent licensor/licensee arcs. We adopt CS analyses for passive, control, long distance dependencies, and other phenomena. This type of analysis for these constructions has its origin in APG and similar analyses are currently assumed in APG/MGG, LFG, HPSG, SFG and other linguistic frameworks.
In our analysis, most of GB’s empty categories represent that a single constituent is the value of more than one constituent licensor/licensee features. Since structure sharing is not permitted in a constituent structure tree model, GB has previously represented multiple licensor/licensee roles for the same constituent as multiple positions in a phrase structure tree containing one antecedent and one or more coreferential empty categories. However, GBUG makes it possible to eliminate these empty categories. The CS analysis is preferred to the empty category (EC) analysis because: (1) The CS analysis represents the same licensing relations as the EC analysis, but uses fewer arcs; and (2) Constraints relating to binding theory, agreement properties and selection restrictions which must be stipulated in EC analysis are a natural consequence of the CS analysis.
The same-reference relation which binds NP-traces, WH-traces, OC-PRO, among others is the crucial factor which permits CS analyses in GBUG. NOC-PRO, which is bound by an overlap-reference relation cannot be represented in a CS analysis. The following similarities between NOC-PRO and pronouns motivate maintaining an EC analysis for NOC-PRO: (1) like pronouns NOC-PRO can take split antecedents; (2) the antecedent of NOC-PRO is optional; and (3) coreference with NOC-PRO is constrained by Binding Theory. In contrast, the antecedents of ecs other than NOC-PRO are determined in the lexicon or by special mechanisms for covering long distance dependencies. For ecs other than NOC-PRO, Binding Theory incorrectly predicts the existence of more than one possible antecedent in a given sentence.

We examine the following hypothesis assumed in the GB literature:

The ECS-ARE-REAL HYPOTHESIS: Ecs exist in a string of words independently of the properties they represent, i.e., ecs exist for purposes of some phonological or psychological mechanism in many of the same respects that words and phrases exist.

If the Ecs-are-real hypothesis can be shown to be correct, this would argue against adopting a CS approach, an approach which assumes that empty categories do not exist. We reject the Ecs-are-real hypothesis based on the following:

(1) We show that, assuming the Chomsky and Lasnik (1977) generate and filter model of building phrase structure (i.e., parsing) and the ecs-are-real hypothesis, an infinite number of ecs will be generated between any two positions in a string of words. Although each ill-formed ec will theoretically be filtered out by GB principles, the derivation will never terminate. This supports either the CS approach or the LEXICAL EC HYPOTHESIS, the view that ecs are features of lexical items, e.g., under the Lexical ec hypothesis, NP-traces are features of passive verbs.

(2) We examine an argument for the existence of NP-trace made in Miyagawa (1989) based on the fact that nominal quantifiers in Japanese obey constraints as if they modify the initial NP-trace (the C-tail of an A-chain) rather than the phonological \( \tilde{N} \) (the C-head). We show that Miyagawa provides evidence for a complement relation to exist for passive and unaccusative
verbs under the CS analysis. However, she does not provide evidence that favors an EC approach over a CS approach.

(3) We contrast competing analyses of the morphological process known in the literature as WANNA CONTRACTION, the contraction of various verbs or semi-modals with to, e.g., want + to → wanna. We show that following Postal and Pullum (PP) (1978, 1979, 1982, 1986) and contrary to Chomsky (1975b, 1980a), Lightfoot (1976, 1977), Chomsky and Lasnik (1977, 1978), Fiengo (1980), Jaeggli (1980), and Aoun and Lightfoot (1984), among others, a constraint on wanna contraction based on a CS account surpasses an account based on the ecs-are-real hypothesis in both descriptive and explanatory adequacy.

This chapter motivates the elimination of most empty categories from GB theory. We show that the resulting analyses are preferred because: the selectional and binding properties of ec need not be stipulated, the representations are simplified in that no coreference relations need to be stated and the problem of nonterminating derivations is solved.

6.2 The Role of Empty Categories in GB Theory

Chomsky (1973) and subsequent work in the Extended Standard Theory (EST) and GB frameworks assume that ecs encode the relationship between declarative clauses containing canonical (active finite external-theta licensor) verbs with other types of clauses. More recent accounts, e.g., Jaeggli (1986), among others also use ecs to encode the relationship between the lexical entries of non-canonical (e.g. passive) lexical entries and canonical entries. In some "atypical" type of phrase P (e.g., a passive sentence, a sentence containing an unaccusative verb, etc.), an ec occupies a syntactic position which is filled by some phrase X in the corresponding canonical phrase P'. Given some licensing relation R which holds in both P and P', both the antecedent of e and X are the L of R, where the L of R is either the licensor of R or the licensee of R. For example, the passive sentence (6.1a) means about the same as the active sentence (6.1b), excluding focus considerations, etc. In particular, the relation Internal-Theta(tickle,John) holds in both sentences. The trace t in (6.1c,d) occurs in complement position, the position in which John occurs in (6.1a). GB theory assumes that the internal theta role is assigned to the complement position in both (6.1a) and (6.1b). In (6.1b), as represented by
(6.1c,d), the trace occurring in complement position transmits the theta role to John.

(6.1) (a) John was tickled (by the gorilla)

(b) The gorilla tickled John.

(c) John$_i$ was tickled t$_i$ by the gorilla

(d) John$_i$ was [t'$_i$ tickled t$_i$ by the gorilla]

Ecs in GB theory either represent (i) a position e in some atypical phrase/lexical entry corresponding to a filled position in some canonical phrase/lexical entry or (ii) some intermediate position between the antecedent and e. Intermediate ecs typically are used by grammatical constraints to impose limitations on where initial ecs or other intermediate ecs can occur. For example, Chomsky (1986b, pp. 29-31) argues that the trace t' in (6.2a) (his 55) is needed to license the trace t (the position canonically occupied by a manner adverb or $\bar{P}$). He argues that how can not move directly from t to its current position, but rather that how moves initially to t' and then to its final position. Chomsky supports this point of view by citing other instances like (6.2b) (his 22d) in which intermediate traces are ill-formed (as evidenced by the fact that how cannot refer to the manner in which the car would hypothetically be fixed.)

(6.2) (a) How did [IP you [VP t' [VP fix the car t]]]

(b) *How did Bill [t$^3$ [wonder [CP who [t$^2$ [wanted [t$^1$ [PRO to fix the car t]]]]]]]]

In early theories of transformational grammar (TG), e.g., Chomsky (1965), movement and deletion transformations capture the relationship between various constructions and active declarative main clauses with phonetically present subjects and unergative finite main verbs. In these theories, a sentence like (6.1a) is a surface structure (S-structure) derived from a deep structure (D-structure) approximating the S-structure of a sentence like (6.1b). Ecs of the type used in Chomsky (1973) and subsequent work, project D-structure properties onto S-structure, yielding a representation like (6.1c), where t (an ec of type NP trace) is a place marker of sorts showing where John would occur at D-structure. More recent accounts of passive, among other syntactic phenomena, derive S-structure from an interaction of principles of case theory, theta theory and properties of lexical entries. For example, in
Jaeggli’s (1986) account, the passive rule transforms a canonical (finite active) entry of a verb into a passive entry by suppressing the external theta role and absorbing one Complement-Case (case absorption is not universal throughout languages, it is however mandatory in English under Jaeggli’s analysis.) The following analysis is given to (6.1c): (i) the trace in (6.1c) marks the position where the theta role is assigned on the analogy of the canonical verb entry; (ii) the position of the $\bar{N} \text{John}$ is the position in which Specifier-Case is assigned by the finite verb inflection; and (iii) John moves from the position of the trace at D-structure to its final position at S-structure in order to receive case. The case filter is satisfied by the chain of NP-movement [$\text{John, t}$] as per Chapter 3. We assume (6.1d) to be the correct analysis as noted above. The passive rule transforms the canonical lexical entry of tickle to yield tickled so that the specifier is the antecedent of an NP-trace in complement position and the Complement-Case and External-Theta relations are absorbed and suppressed as in the Jaeggli approach (see Chapter 2 for details.)

The GB analyses we are aware of all assume that at least some ecs are BASE GENERATED, meaning that the ec is generated at D-structure, is included in the lexical entry, or is otherwise not inserted by a transformation. In particular PRO and pro, ecs which predecessors to GB theory derive by deletion transformations (Equi NP deletion and pro-drop respectively), are base generated in GB accounts. However, GB analyzes differ regarding whether or not NP traces ($\bar{N}$ traces) and WH traces are base generated. Koster (1978, 1984), Correa (1987, 1988, 1991), Leacock (1991), among others argue versions of GB theory without D-structure in which all (syntactic) ecs are generated at S-structure (We are unaware of any claims that Logical Form (LF) traces are generated at LF. However, this dissertation limits its discussion to S-structure ecs.) Chomsky (1973, p.157) discusses the possibility of a No-D-Structure approach to EST and Chomsky (1981, p.90) suggests that it would not be easy to find empirical evidence distinguishing a No-D-Structure approach from one in which ecs are derived from movement transformations.

As shown in Figure 6.2, Correa (1988, pp. 96-98, pp. 167-168) shows that for any ec e, the type of e is completely predictable from the following factors: (i) whether or not e occurs in an argument position, where an ARGUMENT POSITION is a syntactic position where external and internal theta
roles are canonically assigned, i.e., specifier positions of verbs, specifier positions of small clauses, and all complement positions; (ii) whether or not e is governed; and (iii) whether or not e is case-marked. In Correa’s system, the possibility that an ec occurs at all is determined by phrase structure rules based on $\bar{X}$ theory. Any constraints on coreference between a base-generated ec and its antecedent must be viewed as constraints on the S-structure representation. In contrast, under the view that an ec e is the residue of some movement or deletion transformation, constraints on coreference between e and its antecedent may be (and in fact usually are) viewed as constraints on the transformational operation. Koster, Correa and Leacock argue that the "constraints on representation" approach is preferred on grounds of simplicity because transformations and a separate level of D-structure are unnecessary in this approach. Below, we assume a variant of the "constraints on representation" approach.

<table>
<thead>
<tr>
<th>A-position</th>
<th>Government</th>
<th>Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh-trace (variable)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Wh-trace (COMP)</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>NP-trace</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PRO</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 6.2 (Correa 1988, p. 98, no. 113)

6.3 The Constituent Sharing Analysis

This section describes and motivates the constituent-sharing (CS) analysis for passive, obligatory control and $wh$ extraction constructions, as well as CS versions of the V-chain and Determiner-Chain analyses from in Chapter 3. Since FSs representing CS analyses do not need to represent coreference, they represent the same licensing relations using fewer arcs and are thus preferred by the Feature Structure Economy Principle, defined in Chapter 2 and repeated below:

**Definition 2.3 The Feature Structure Economy Principle (FSEP):** Given two Feature Structures $FS_1$ and $FS_2$ representing the same licensing relations, $FS_1$ is preferred iff $FS_1$ represents these licensing relations using fewer features (arcs) than $FS_2$. 

269
In order to accommodate a CS analysis, a few modifications must be made to GB theory. In particular, a number of theoretical concepts which have previously been formulated in terms of nodes of constituent structure trees must be reformulated in terms of arcs of a DAG model of a FS. We begun this process in Chapter 3 when we defined command relations as relations between arcs (see Chapter 3, note 85).

The core idea of the CS analysis is that the same constituent is the value of some set of constituent licensor/licensee features. Figure 6.1 (repeated below) is an example of a CS analysis of the A-chain found in a passive sentence. Figure 6.3 is a CS analysis our V-chain analysis from Section 3.6. Each of these analyses have the following in common: (1) There exists a chain C consisting of constituent licensor/licensee arcs which share the same value; and (2) the C-Head and C-Tail of the chain each have the same source and target as some other type of licensor or licensee arc, or as is the case for Head-Proj arcs, are simultaneously a constituent licensor arc and some other type licensor arc. Figure 6.1 contains the A-chain

\[ \text{Specifier, Head-Proj Complement Specifier, Head-Proj Complement Head-Proj Complement} \]

with the C-head Specifier having the same source and target as Specifier-Case; and the C-tail Head-Proj Complement Head-Proj Complement, having the same source and target as Head-Proj Complement Head-Proj Internal-Theta. Figure 6.3 contains the V-chain

\[ \text{Head-Proj, Head-Proj Complement Head-Proj} \]

where Head-Proj arcs represent multiple licensor/licensee features including: Constituent Licensor, Theta Licensor, Case-Licensor, Modification-Licensee, N-Agreement-Licensee, Quantification-Licensee and Degree-Licensee, among others. We thus assume that: (1) Head-Proj, the C-Head of the V-Chain is simultaneously a Constituent-Licensor arc and a Case-Licensor arc; and (2) Head-Proj Complement Head-Proj, the C-Tail of the V-Chain is simultaneously a Constituent-Licensor arc and a Theta-licensor arc. Figure 6.4 contains the Determiner-Chain

\[ \text{Specifier, Head-Proj Specifier, Head-Proj Head-Proj Head-Proj Specifier} \]
where the C-head Specifier and the C-tail Head-Proj Head-Proj Head-Proj Specifier each have the same source and target as one N-Agreement-Licensor arc, i.e., local agreement holds in the sense defined in Chapter 3.
A comparison of the CS analyses in Figures 6.1, 6.3 and 6.4, on the one hand, with their counterpart EC analyses in Figures 6.1’, 6.3’ and 6.4’ reveals that exactly the same licensing relations are represented on both accounts. However, the CS accounts do not require the arcs representing coreference and and thus are preferred by the FSEP. It follows from the CS approach that licensee features which share a value impose morphological agreement and selection restrictions on that constituent. However, the corresponding fact that ecs and their antecedents have the same agreement properties and satisfy the same selection restrictions must be stipulated in the EC approach. Therefore, the CS approach is preferred by Occam’s razor.
FS representing a CS analysis of
dve novyx studentki

Category: Noun
Morph-Agr: Fem-Plural
Morph-Case: Nominative
Specifier: Category: Determiner
Phonology: dve
Morph-Agr: Fem-Sing
Morph-Case: Genitive
Head-Proj: Category: Adjective
Adjunct: Category: Adjective
Phonology: novyx
Morph-Agr: Fem-Sing
Morph-Case: Genitive
Head-Proj: Category: Adjective
Phonology: studentki
Morph-Agr: Fem-Sing
Morph-Case: Genitive


Figure 6.4
A FS representing an EC analysis of 

*John was tickled by the gorilla*

Figure 6.1
An EC Based lexical entry for

*eats*

Category: \(\text{INFL}\)
Tense: \(\text{Present}\)
Specifier: \(\text{Noun}\)
  Morph-Agr: \(\text{Third-Person-Singular}\)
  Morph-Case: \(\text{Nominative}\)
Antecedent: \(\text{[l]}\)
Head-Proj: \(\text{Category: [i]}\)
  Tense: \(\text{[j]}\)
  Head-Proj: \(\text{Category: [i]}\)
    Phonology: \(\text{eats}\)
    Antecedent: \(\text{[m]}\)
    Tense: \(\text{[j]}\)
Complement: \(\text{[n]}\)
  Category: \(\text{NP-trace}\)
    Specifier: \(\text{[p]}\)
      Category: \(\text{NP-trace}\)
      Same-Reference: \(\text{[l]}\)
      Head-Proj: \(\text{Category: [o]}\)
        Head-Proj: \(\text{Category: Verb-trace}\)
          Same-Reference: \(\text{[m]}\)
          Complement: \(\text{[q]}\)
            Category: \(\text{Noun}\)
            Morph-Case: \(\text{Accusative}\)
            Internal-Theta: \(\text{[q]}\)
            Complement-Case: \(\text{[q]}\)
          External-Theta: \(\text{[p]}\)
        Internal-Theta: \(\text{[n]}\)
      Specifier-Case: \(\text{[k]}\)

*Figure 6.3’*
FS representing an EC analysis of
dve novyx studentki

Figure 6.4'

All relations which EC analyses assume between nodes may be formulated as relations between
arcs in CS analyses since there is a one to one correspondence between each empty category in an EC

analysis and some constituent licensor/licensee arc in corresponding CS analysis. As noted by Ray Dougherty (personal communication), relations like c-command, etc. have undesirable results if interpreted as relations on nodes in FSs, e.g., a node N (an \( \overline{N} \)) can c-command a phrase X (a \( \overline{V} \)) on the basis of a path P (Specifier of \( \overline{I} \)) with N as the target, even though X dominates N following another path P' (Complement of \( V^0 \)) with N as the target. In chapter 3, we define command relations as relations between arcs and above we define chains as sets of arcs instead of sets of nodes. For example, GB’s condition that an antecedent c-commands its trace may be converted into a condition that the C-head of the chain c-commands the C-tail of the chain. Thus in Figure 6.1 satisfies this constraint since the arc Specifier c-commands the arc (i.e., the last arc in the path) Head-Proj Complement Head-Proj Complement. Additionally, control and constraints on long distance dependencies must be viewed in terms of arcs rather than nodes (or FSs rooted at nodes) for CS analyses. For example stating that the arc Specifier controls the arc Head-Proj Complement Specifier in Figure 6.5 is equivalent to saying that the values of these same arcs are in that same control relation in Figure 6.5'. Similarly any constraints which can be stated between the wh-trace and its antecedents in Figure 6.6' may be stated as a condition on the relative positions between the position of the C-head and C-tail of the \( \overline{A} \)-chain [Head-Proj, Complement Complement Complement Complement]. A description of all changes in GB concepts necessary for this change from relations on nodes or FSs to relations on arcs is beyond the scope of this dissertation (wh extraction in particular). Based on our preliminary results, we assume that the relevant linguistic generalizations captured in GB’s EC analyses may be captured in CS analyses as well.
A FS Representing a CS Analysis of a Finite Subject Control Verb

Category: | INFL
Tense:  |
Specifier: | N
Head-Proj: Category: | Specifier: | Verb
Head-Proj: Category: | Specifier: | Verb
Complement: | Head-Proj: Category: | Specifier: |
Head-Proj: Category: | Specifier: |
Complement: | Head-Proj: Category: | Specifier: |
Head-Proj: Category: | Specifier: |
Head-Proj: Category: | Specifier: |
Internal-Theta: | External-Theta: | Specifier-Case: |

Figure 6.5
A FS Representing an EC Analysis of a Finite Subject Control Verb

Category: [i] INFL
Tense: [j]
Specifier: [k] Category: Noun
Antecedent: [l]
Head-Proj: Category: [m]
   Tense: [n]
      Head-Proj: Category: [o] Verb
         Specifier: [p] Category: NP-trace
            Same-Reference: [q]
               Antecedent: [r]
                  Head-Proj: Category: [s]
                     Head-Proj: Category: [t] INFL
                        Head-Proj: Category: [u]
                           Head-Proj: Category: [v]
                              Specifier: Category: PRO
                                 Same-reference: [w]
                                    Phonology: to
                                    Specifier-Case: [x]
                                     Internal-Theta: [y]
                                        External-Theta: [z]
                                          Specifier-Case: [k]

Figure 6.5'
A FS Representing a CS Analysis of the sentence

Who did John see?

Head-Proj: i Category: WH-Complementizer-Noun
Phonology: who
Complement: j Category: k INFL
Specifier: l Category: Noun
Phonology: John
Head-Proj: Category: k
Head-Proj: Category: k
Phonology: did
Complement: m Category: n Verb
Specifier: k
Head-Proj: Category: n
Head-Proj: Category: n
Phonology: see
Complement: i
Internal-Theta: i
Complement-Case: i

Internal-Theta: m
Specifier-Case: l
Internal-Theta: i

Figure 6.6
6.4 Selection Restrictions

It follows from the CS approach that an item which is the predicate licensee of more than one predicate relation must meet the selection restrictions associated with each of those predicate relations (Ray Dougherty, personal communication). An equivalent requirement, if correct, would have to be explicitly stipulated in an EC approach as follows:

(6.3) The antecedent of an ec must obey all selection restrictions imposed on that ec and vice versa.

Assuming this restriction is valid, the CS approach is preferred over the EC approach because the former need not state (6.3) explicitly. On the other hand, if an ec is limited by selection restrictions which its antecedent violates, a CS analysis of that construction would not be possible\(^5\).
In each of the examples in (6.4), there are two theta relations: \( \text{Theta1}(X,Y) \) and \( \text{Theta2}(Z,Y) \), both of which have an \( N \ Y \) as their theta licensee. The examples in (6.4) are each ill-formed because \( Y \) fails one of the selection restrictions associated with one of the theta relations. For example: John can try something or be convinced of something, but he cannot elapse; the time can pass or elapse, but cannot try anything or be persuaded of anything; and a book can open, but it cannot be convinced of anything. The semantic anomaly of these examples follow from CS analyses of these constructions.

(6.4) (a) %John tried to elapse.
(b) %John was convinced to try to elapse.
(c) %John persuaded the time to pass.
(d) %The time tried to elapse.
(e) %Which book did John convince to open.

(6.5) contains some apparent counter examples. An idea cannot be eaten, but (6.5b,c) are well-formed nevertheless. Selectional restrictions may be violated in complements of certain verbs due to the selection restrictions of those verbs, e.g., it is possible to want or believe ridiculous things like eating an idea (cf. Jackendoff 1983, pp. 212-239, among others). (6.5b,c,d) are grammatical in spite of the fact that the idea violates a potential selection restriction. However, want, try, and say among other verbs of desire, belief and reporting allow selection restriction violations in their complements because it is possible to desire, believe or report something impossible. Therefore (6.5b,c,d) are not counterexamples. There are limits to this approach however: some selection restrictions are difficult to break even in these contexts, e.g., (6.6) is ill-formed since few people can conceive of Mary as a measured segment of time.

(6.5) (a) %John ate the idea.
(b) John tried to eat the idea.
(c) John wanted the idea to be eaten.
(d) John said that Mary ate the idea.

(6.6) %John wanted Mary to elapse.
In summary, these facts about selection restrictions follow from a CS approach, but must be explicitly stipulated in an EC approach. On this basis, a CS approach is preferred.

6.5 ECs and Binding Theory

Previous GB approaches (e.g. Chomsky 1981) assume that coreference between ecs and antecedents is determined using the same principles used to bind pronouns and anaphors. The CS approach does not allow such an interpretation for OC-PROs, NP-traces or WH-traces constructions because the targets of the arcs corresponding to these ecs are built into the lexical entries. We show that perceived parallels between NP-traces and anaphors are incorrect, as are any similarities between OC-PROs and anaphors or pronouns. However we do find significant similarities between NOC-PRO and pronouns. As noted above, only ecs bound by the same reference relation may be replaced by CS analyses. Therefore it is not surprising that NOC-PRO is pronoun-like. We conclude that except for NOC-PRO, ecs are not in any way governed by Binding Theory. Rather, lexical entries determine all properties represented by ecs in the EC approach, and the most efficient way of representing these features is by constituent sharing (except for NOC-PRO.)

Antecedents of NP-traces are always determined by FS based lexical entries containing those traces. In contrast, anaphors can vary with respect to their antecedents. For example, (6.7a) has at least two readings: one in which each other is bound by the $\bar{N}\ \text{John and Mary} \text{ and another in which each other is bound by Bob and Sally}$—either (i) John gave Bob and Sally Mary’s dog and Mary gave Bob and Sally John’s dog; or (ii) John and Mary gave Sally’s dog to Bob and Bob’s dog to Mary. Similarly (6.7b) has two interpretations: one where himself is bound by John and one where himself is bound by Bill. No such ambiguities ever exist with NP traces.

(6.7) (a) John and Mary gave Bob and Sally each other’s dog.

(b) John sold Bill a picture of himself.

Binding Theory provides a set of constraints for ruling out coreference relations between pro-forms and potential antecedents. Binding Theory cannot correctly account for the distribution of NP-traces because NP-traces always have exactly one possible antecedent. This antecedent can be pre-
precisely determined if we assume that it represents a feature of a lexical entry. For example, the EC analysis based lexical entries for *was* and *ticked*, provided in Figures 6.7 and 6.8 combine along with other FSs into Figure 6.1’, unambiguously determining that *John* is the antecedent of the NP-trace. Binding Theory wrongly predicts that NP-traces can have more than one possible antecedent. For example, in (6.8), the NP-traces can only be coreferential with the subject of the sentence, not with the first object, although both are possible antecedents under Binding Theory Principle A defined in Chomsky (1981) and elsewhere.

(6.8) (a) *A puppy* was given the little girl

(b) *The expensive gift* was handed *John*

---

**A FS representing an EC Lexical entry for *was***

- **Category:** [I] INFL
- **Tense:** [I] Past
- **Specifier:** [K] Category: Noun
  - Antecedent: [I]
- **Head-Proj:** Category: [I]
  - **Tense:** [I]
  - **Head-Proj:** Category: [I]
  - Phonology: *was*
- **Complement:** [M] Specifier: Category: NP-trace
  - Same-Reference: [I]
- **Internal-Theta:** [M]
- **Specifier-Case:** [K]

**Figure 6.7**
A FS representing a EC Lexical entry for *tickled*

Category: Verb
Specifier: Category: Noun
Antecedent: K
Head-Proj: Category: Head-Proj: Category: Phono:

Phonology: tickled
Complement: Category: NP-trace
Same-Reference: K
Complement2: Category: Preposition
Head-Proj: Category: Phonology: by

Complement: n
Internal-Theta: l
Suppressed-Theta: n

**Figure 6.8**
GB assumes that NP-traces are [+Anaphoric, -Pronominal], pros are [-Anaphoric, +Pronominal], PROs are [+Anaphoric, +Pronominal] and WH-traces are [-Anaphoric, -Pronominal]. As we have already seen, NP-traces behave like anaphors only in the sense that they must be bound, but not how they are bound. Similarly pro is like a lexical pronoun in that it need not be bound, but differs from pronouns in that pro can be bound by an $N$ in its same clause, as evidenced by the following Spanish sentence:

\[(6.9) \text{Me pro$_1$ gustan [los libros].} \]

(To) Me pro are-pleasing the books

"I like the books"
According to Chomsky (1981), WH-trace is an r-expression (normal $\overline{N}$ like John or the book) with respect to binding theory, i.e. it is neither pronoun-like or anaphor-like. This does not seem right since other types of r-expressions do not require antecedents. It is therefore unclear why Binding Theory (Principle C) should be applied to wh expressions. The principles GB needs to account for long distance dependencies are needed regardless of the pronominal and anaphoric properties of wh-trace and stating that it has such properties does not seem to buy anything at all.

Following Williams (1980), OC-PRO acts like an anaphor in that it must be bound in the superordinate clause and NOC-PRO acts like a pronoun in that it can be optionally coreferential with $\overline{N}$s in the superordinate clause. However unlike anaphors, OC-PROs are bound according to whether they occur in the complement of an object control, subject control or variable control verb. For example as shown in (6.10) tell and persuade select for object control PRO; try, promise and threaten select for subject control PRO; and badger and ask allows for variable control (either subject or object control—cf. Grimshaw and Jackendoff 1981, Wolff, Macleod and Meyers 1993, among others).

NOC-PRO, unlike other ecs, does seem to be (optionally) bound in basically the same situations as pronouns, as exemplified in (6.11.)

(6.10) (a) Mary told/persuaded Betty, PRO$_i$ to leave.
     (b) Cecil, PRO$_i$ to leave.
     (c) Beatrice, promised/threatened Hildegard PRO$_i$ to leave her in Jersey City
     (d) Wanda’s daughter, asked/badgered her, PRO$_{i,j}$ to go to the movies.

(6.11) (a) PRO$_i$ to leave would be their, pleasure.
     (b) For them, PRO$_i$ to leave would be their, pleasure.
     (c) It is important to John, PRO$_i$ to do good in the world.
     (d) It is important to John, for him, to do good in the world.
     (e) PRO$_i$ doing nice things for people is important for John.
     (f) His, doing nice things for people is important for John.
The fact that NOC-PRO is bound by an overlap relation prevents us from replacing NOC-PRO analyses with CS analyses. Assuming NOC-PRO obeys Binding Theory as proposed by Williams (1980), Koster (1984) and others, it is warranted to posit a relationship between NOC-PRO and normal pronouns. NOC-PRO could be viewed as an invisible pronoun which occurred in a number of environments including specifier arc of to complements, of gerunds and other lexical items. Further investigation is required. In particular, it is important to see whether there are any grammatical constraints applying to pronouns which also apply to NOC-PRO\(^7\) or if NOC-PRO patterns with pronouns with respect to discourse constraints on coreference.

6.6 Null Phrase Structure Rules and Nonterminating Derivations

This section shows that ecs in previous GB analyses are best interpreted as properties of lexical entries and phrases—we call this hypothesis the LEXICAL EC HYPOTHESIS. The lexical ec hypothesis is in direct conflict with the following hypothesis assumed in the GB literature:

The ECS-ARE-REAL Hypothesis: Ecs exist in "the mind’s eye", i.e., ecs exist for purposes of some phonological or psychological mechanism in many of the same respects that words and phrases exist.

If it could be shown that the ecs-are-real hypothesis was correct, that would constitute evidence against the CS approach. This section shows that any model for generating syntactic analyses of phrases (e.g., a parsing algorithm) along the lines of Chomsky (1977, 1991) would never terminate assuming the ecs-are-real hypothesis. This fact shows that the ecs-are-real hypothesis is undesirable, but does not distinguish between the lexical ec hypothesis and the CS analysis. Although this chapter provides evidence for abandoning most ecs, it is important that it is possible to represent ecs in GBUG for at least two reasons: (1) we assume some ecs, e.g. NOC-PRO; and (2) GBUG provides the means for comparing EC and CS analyses. This section assumes the lexical ec hypothesis in order to give ecs a coherent interpretation in the context of GBUG-based analyses.

If, as assumed in Chomsky (1981) among others, ecs exist in the same sense that words exist, it follows that the phonetically empty string \(\varepsilon\), may contain any number of ecs. In many explicit compu-
tational models of GB theory, e.g., Fong (1991a), each ec e is introduced by one phrase structure rules of the form

$$e \rightarrow \varepsilon.$$  

We will call these and all other phrase structure rules with left hand side (RHS) symbols of $\varepsilon$ NULL RULES. Chomsky and Lasnik (1977, 1991), among others assume some version of the GENERATE AND FILTER approach (also called the "generate and test" approach) in which all possible syntactic analyses are generated and then the ungrammatical ones are ruled out by various syntactic constraints. In our view, the "generation" procedure functions in much the same manner as a parsing algorithm in a computational model. In a bottom-up parsing approach, words project their own phrase structure—they select which phrases they can combine with to form larger phrases. In a top-down approach the set of phrase structure rules is selected first and then words are found that fill positions in these rules. Based on the discussions above, it should be clear that we are assuming a bottom-up approach. In general, both top-down and bottom-up parsing algorithms will take a string of words (and a set of phrase structure rules) and produce a chart which includes the grammatical parses of the word string. However, ecs necessarily limit the search strategy to a top-down approach.

If we assume a bottom-up strategy, the following problem occurs. Given that the generation precedes the filtering stage of derivation, the rule

$$\text{NP-trace} \rightarrow \varepsilon$$

and all other null rules may apply once for each occurrence of $\varepsilon$ in any string of words. Since $\varepsilon$ occurs an infinite number of times in any string of words, each null rule applies an infinite number of times producing an infinite number of analyses most of which would be ruled out once the filter stage of the derivational process is reached. However, the generation procedure never terminates, because it has to complete an infinite number of operations, and the derivational process never reaches the filter stage. As is well-known, this same problem shows up for all bottom-up context-free parsing algorithms as well as the Earley (1970) algorithm (which has bottom-up as well as top-down properties.)
We propose the following alternatives:

(i) assume that ecs are properties of lexical entries (the lexical ec hypothesis); or

(ii) replace ecs with sets of structure sharing arcs (the CS analysis).

By adopting a linguistic theory that cares about economy of derivation in the sense of Chomsky (1991), it becomes necessary, at least in a general way to characterize how derivations are generated. The fact that GB assumes that phrase structure is "projected" from the lexicon suggests a bottom-up, not a top-down approach. It is therefore significant that ecs pose a significant problem for all bottom-up approaches when ecs are interpreted as items in a string of words.

Correa (1987, 1988, 1991) circumvents the problem of parsing with null rules by introducing ecs with Attribute Grammar (AG) rules rather than ordinary phrase structure rules. These rules insert ecs into phrase markers if any empty syntactic position meets any of the sets of criteria listed in Figure 6.2—the identity of the ec is also determined according to Figure 6.2 (repeated below). In Correa’s implementation of GB Theory, ecs may be viewed as "existing" in strings of words. However, Correa’s ecs may just as easily be viewed as instantiations of properties of "surrounding" lexical items (like our (i)). Correa’s parsing algorithms do not "look for" ecs in the string, rather they check all unfilled constituent positions created on the basis of lexical items found in the input string and see if these positions meet the necessary criteria for ecs to be inserted. The distinction between "position in the input string" and "constituent position" is crucial here: the former corresponds to phonetics and the latter corresponds to values of constituent licensor/licensee features in GBUG. The main difference between our approach and Correa’s is that we adopt a CS approach and he adopts an approach in which ecs are inserted on the basis of lexical entries.
Given the undesirability of null rules, clearly two grammatical systems $G$ an $G'$ may be evaluated on the basis of whether or not they contain null rules. If $G$ conveys all the same information as $G'$, but differs from $G'$ solely in that $G$ lacks null rules and $G'$ contains them, $G$ is clearly superior to $G'$ on the grounds of economy of derivations: fewer derivations in $G$ will crash (or fail) in the sense of Chomsky (1992) and fewer spurious derivations will even be attempted—in fact, an infinitely fewer number of spurious derivations will be attempted. By acknowledging that ecs are do not actually "exist" as members of strings of words, we simplify our representation of a string of words and greatly reduce the number of possible derivational processes applicable to that string.

6.7 Quantifiers in Japanese

Miyagawa (1989) takes the examples in (6.12), repeated below, as evidence that NP traces "exist" in Japanese sentences containing passive or unaccusative main verbs, i.e., he adopts a version of the ecs-are-real hypothesis. This section provides a brief overview of Miyagawa’s analysis and shows that a constituent sharing analysis is compatible with Miyagawa’s findings. We show that Miyagawa has motivated the existence of a complement relation for these verbs and an NP-trace is a way to represent this relation in a phrase structure tree model. However, in a FS based model a CS analysis is an alternative. Therefore Miyagawa does not provide any evidence which is incompatible with a CS analysis of passive and unaccusative constructions.

(6.12) (a) Boku wa yuumei na gakusya ni 3-nin atta
  I TOP famous scholars DAT 3-CL met
  ‘I met three famous scholars.’
(Miyagawa 1989, p.35, no. 56)
(b) *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'
(Miyagawa 1989, p.39, no. 67)

(c) Yuube, kuruma ga doroboo ni 2-dai nusum-are-ta
last night cars NOM thief by 2-CL steal-PASS-past
'Last night, two cars were stolen by a thief.'
(Miyagawa 1989, p.38, no. 66)

(d) Gakusei ga ofisu ni 2-ri kita
students NOM office to 2-CL came
'Two students came to the office.'
(Miyagawa 1989, p.42, no. 81)

Miyagawa argues that the quantifiers in (6.12a,c,d) modify the object position, even though (6.12a) is an active sentence, (6.12c) is a passive and (6.12d) is an unaccusative sentence. In (6.12c,d), it appears that the NP trace is modified, rather than the antecedent directly. As evidenced by the ungrammaticality of (6.12b), this is not simply a case of free word order: the quantifier in object position can only modify $N$ in object position or NP traces in object position. According to Miyagawa, these facts show that verbs always assign theta roles to particular syntactic positions and that NP traces appear in these positions in examples like (6.12c,d).

Miyagawa assumes the following constraint based on Williams (1980):

Mutual C-Command Requirement: For a predicate to predicate of a NP $[\bar{N}]$, the NP or its trace and the predicate or its trace must c-command each other. (Miyagawa 1989, p.30)

He attempts to account for the distribution of nominal quantifiers (NQs) in Japanese with respect to the nominals they quantify over by maintaining that an NQ predicates of the $\bar{N}$ it quantifies. He contrasts examples like (6.12a) and (6.12b) claiming that in sentences lacking NP movement, NQs occur either adjacent to the $\bar{N}$s they modify or next to an intervening adverb. Miyagawa posits that $\bar{N}$ objects and NQs occur within the maximal projection of the verb such that the Mutual C-command Condition is met. (6.12b) is ungrammatical because the NQ 2-ri does not c-command the subject of the sentence. In contrast, (6.12c,d), sentences for which GB analyses posit NP movement, are grammatical with the
interpretation that the NQ predicates of the subject of these sentences. Miyagawa explains that:

There can only be one reason why (6.12c) [our number] is grammatical. The NQ somehow manages to c-command the subject NP. The NQ obviously cannot directly c-command the NP in the subject position because of the intervening maximal projection VP. The only choice open to us is that the subject NP starts out in the object position and is moved to the subject position, leaving behind a trace... Before the movement, the NP and the NQ mutually c-command each other directly because both occur in the VP. After the movement, the mutual c-command is maintained by the trace left by the NP. (Miyagawa 1989, p. 39)

Therefore, Miyagawa argues that NP-traces are needed to account for the distribution.

Compare Figures 6.9 and 6.10, the former a FS representing a trace analysis of (6.12c) and the latter showing a CS analysis. The Mutual C-command Requirement holds under both analyses assuming GBUG’s version of c-command which is defined on arcs. In each case the arcs \textit{Head-Proj Head-Proj Complement Head-Proj Complement1} and \textit{Head-Proj Head-Proj Complement Head-Proj Complement2} c-command each other, i.e., for purposes of her argument the CS and EC analyses are equivalent.\textsuperscript{11}
A FS representing an EC Analysis of
Yuube, Kuruma ga doroboo ni 2-dai nusum-are-ta

Category: [INFL]
Tense: [Past]
Adjunct: [Adverb]
Phonology: Yuube
Head-Proj: [Category: [Noun]
  Specifier: [Category: [Case-Marker
    Head-Proj: Category: [NQ
    Phonology: kuruma
  Antecedent: [N]
Head-Proj: Category: [Verb]
  Tense: [Past]
  Head-Proj: Category: [Verb]
  Complement: [Category: [NP-trace
    Specifier: Category: [NP-trace
    Same-Reference: [N]
    Antecedent: [N]
    Head-Proj: Category: [Noun]
    Phonology: nusum-are-ta
    Complement1: Category: [NQ
      Head-Proj: Category: [NQ
      Phonology: 2-CL
      Predication-Licensee: [N]
    Complement2: Category: [Postposition
      Head-Proj: Category: [Postposition
      Phonology: ni
      Complement: [Category: [Noun
        Phonology: doroboo
        Complement-Case: [N]
        Internal-Theta: [N]
        Suppressed-Theta: [N]
    Internal-Theta: [N]
    Specifier-Case: [N]
    Modifier: [N]

Figure 6.9
We conclude that the distribution of Japanese NQs and the Ns they quantify does not chose between EC and CS analyses. Miyagawa shows that a complement relation exists in passive and unaccusatives and that NP-traces may be used in a representation of this relation. Miyagawa does not, however, motivate a version of the ecs-are-real hypothesis. As shown above, CS analyses are preferred on independent grounds.
6.8 Wanna Contraction and the Status of Empty Categories

6.8.1 Introduction

This section compares competing analyses of the morphological process known in the literature as WANNA CONTRACTION, a morphological rule which combines infinitival to, with various verbs/semi-modals. For example wanna contraction combines to with have, got, going, ought, use and want, to form hafta, gotta, gonna, oughta, usta and wanna. Most of the EST and GB oriented accounts explain why wanna contraction is possible in sentences like (6.13a,b), but not in sentences like (6.13c,d) based on a version of the ecs-are-real hypothesis.

(6.13) (a) They/I want to leave
(b) They/I wanna leave
(c) Who do you want to leave?
(d) *Who do you wanna leave?

For example, Chomsky (1980a) proposes the following explanation:

The rule of contraction that assigns to the expression want + to the phonetic form wanna does not apply unless want and to are contiguous, and to the mind’s eye they are not in examples (5b) [who do you want t to meet Bill] and (6) [who do you want t to visit], since they are separated by the Noun Phrase trace of wh-movement, which though the mind does not know it at this point in mental computation, will ultimately have no phonetic realization. (Chomsky 1980a, pp. 159-160)

Thus according to Chomsky, an intervening trace which exists in the mind’s eye blocks wanna contraction, just as the \( \overline{\text{N John}} \) blocks wanna contraction in (6.14). The various versions of this claim collectively represent the most explicit and most commonly cited argument for the ecs-are-real hypothesis. This section shows that this argument ultimately fails on grounds of descriptive and explanatory adequacy. We show that wanna contraction actually supports a CS analysis instead of arguing against it.

(6.14) (a) I want John to drink his milk.
(b) *I wan—John—na to drink his milk.

We contrast three accounts of wanna contraction:
(1) The TO-SUBSUMPTION (TS) analysis — Putting aside the value of the category and phonology features, only verbs/semi-modals which take to complements and are subsumed by the lexical entry for to, may undergo wanna contraction. We show that the TS analysis is equivalent to an analysis, which we will call the SUBJECT SHARING (SS) analysis, made in Postal and Pullum (PP) (1978, 1979, 1982, 1986). According to the SS analysis, only verbs whose final subjects are also the final subjects of their complements may undergo wanna contraction, where FINAL SUBJECT is an RG relation which closely corresponds to GB specifier relations. The SS and TS analyses are equivalent because our lexical entry for infinitival to is basically the smallest possible FS in which GB’s equivalent to subject sharing would hold. Therefore all FSs subsumed by this entry would satisfy the SS analysis as well as the TS analysis. The effect of this constraint is that only subject raising and subject control verbs/semi-modals may undergo wanna contraction.

(2) The CASE MARKED TRACE (CMT) analysis — This is a version of the ecs-are-real account noted above which was developed in the course of a decade in Chomsky (1975b, 1980a), Lightfoot (1976, 1977), Chomsky and Lasnik (1977, 1978), Fiengo (1980), Jaeggli (1980), and Aoun and Lightfoot (1984), among others. Many of the changes proposed to this analysis over time were reactions to the PP articles listed above. Under the CMT account, only wh-traces which are case licensees block wanna contraction.

(3) The CASE ASSIGNMENT (CA) analysis is based on the CMT analysis, but does not assume the ecs-are-real hypothesis and is compatible with both CS and EC analyses. According to the CA hypothesis case assigning lexical entries for verbs are incompatible with wanna contraction.

This section demonstrates that the TS and CA analyses are preferred over the CMT analysis on the following grounds: (1) both analyses surpass the CMT analysis in descriptive adequacy; and (2) neither analysis assumes the ecs-are-real, and thus are preferred for all the reasons discussed above, i.e., coreference and selection conditions need not be stipulated and derivations will terminate. We could find very few examples distinguishing the TS and CA analyses on descriptive grounds.
The TS analysis is preferred over the CMT and CA analysis on grounds of explanatory elegance, deriving from the fact that the TS constraint on wanna contraction is (excepting category and phonology features) the lexical entry for to, the lexical item which combines with the verb/semi-modals during wanna contraction. We are unaware of any explanatory principle which motivates the CA analysis. We show that the principles previously postulated as the basis of the CMT approach are inelegant and descriptively inadequate. Furthermore the CMT approach assumes a version of the ecs-are-real hypothesis and therefore leads to nonterminating derivations.

6.8.2 The Contraction Debate

The descriptive and explanatory adequacy of two alternative accounts of wanna contraction is the topic of numerous publications between 1970 and 1986, collectively referred to by some as "the contraction debate" (the title of Postal and Pullum 1982.) This section provides a summary of the positions taken at the close of this debate in order to put our own analysis in historical context. In our view, the debate had two major results: (1) the CMT analysis of wanna contraction failed on grounds of descriptive and explanatory adequacy; and (2) the merit of the SS analysis, the Relational Grammar based approach was not recognized by most GB linguists because it could not easily be expressed in a constituent structure tree model. Aoun and Lightfoot (1984, p. 472, note 8), for example, make a cryptic remark to the effect that the SS approach is far too complex to meet general requirements.

Chomsky (1980a) proposes the following rationale for why wh traces block wanna contraction:

The rule of contraction that assigns to the expression want + to the phonetic form wanna does not apply unless want and to are contiguous, and to the mind’s eye they are not in examples (5b) [who do you want t to meet Bill] and (6) [who do you want t to visit], since they are separated by the Noun Phrase trace of wh-movement, which though the mind does not know it at this point in mental computation, will ultimately have no phonetic realization. (Chomsky, 1980a, pp. 159-160)

Chomsky (1980a) then offers the following possible explanation for why only case-marked traces block wanna contraction.

It might be, for example, that trace without Case is optionally deleted by convention prior to contraction. (Chomsky 1980a, p. 280, note 24)

Jaeggli (1980, pp. 240-242) takes basically this same positions: case marked traces, but not other ecs block wanna contraction because case-less traces (and implicitly other case-less ecs) delete prior to
contraction. However both Chomsky (1980a) and Jaeggli (1980) were prior to the GB framework: these authors assumed that there was a general device for deleting intermediate traces which is unav-
ailable in GB. Thus Chomsky (1981) provides a different analysis:

... trace with Case is "visible" to the contraction rule (37) [want+to → wanna] while trace without Case is not...Then we conclude that to be visible to a PF-rule, an empty category must have the feature Case. (Chomsky 1981, p. 181)

Based on some unpublished work by Aoun, Chomsky (1981, pp. 117, 224 note 24) proposes that θ-
roles are Logical Form (LF) properties and Case is a Phonological Form (PF) property. He claims that LF rules ignore items lacking θ-roles and PF rules ignore elements lacking case. This implies that by being assigned case, a phonological property, an ec becomes in some sense more like a phonological entity, thus blocking wanna contraction, a phonological process. Chomsky attempts to justify the CMT approach based on perceived parallels between ecs which are marked with a phonological property and phonetically realized entities.13

Postal and Pullum (PP) (1978, 1979, 1982, 1986) propose an alternative which attempts to account for the distribution of these contracted forms with a generalization about when contraction does occur, instead of when it doesn’t. PP propose that:

A contraction trigger V can have a contracted form with infinitival to only if:

a. to is the main verb of the initial direct object complement of the matrix clause whose main verb is V [Note that PP follow Pullum (1982) in maintaining that to is a subject to subject raising verb.]

b. the final subject of the complement is identical to the final subject of the matrix. (Postal and Pullum 1982, p. 130)

PP’s generalization, i.e. the SS analysis, characterizes subject to subject raising verbs and subject control verbs as the only candidates with contraction with to.

A strong point of the SS analysis is that it descriptively accounts for nearly all wanna contraction data.14 In contrast, PP cite numerous examples in which the CMT analysis fails to account for the data, including examples such as (6.15c), (6.16c), (6.17b), the purpose (or rationale) clause readings of (6.18b) and (6.19b), (6.20b) and (6.21b), sentences which the CMT account fails to mark ungrammatical.
(6.15) (Postal and Pullum 1982, p. 124, 3a,b,c)

(a) I don’t want it to become standard practice in this monastery to flagellate oneself in public

(b) *I don’t want to flagellate oneself in public to become standard practice in this monastery

(c) *I don’t wanna flagellate oneself in public to become standard practice in this monastery

(6.16) (Postal and Pullum 1982, p. 125, 6a,b,c)

(a) To regret that one does not have seems like to want

(b) *It seems like to want to regret that one does not have

(c) *It seems like to wanna regret that one does not have

(6.17) (Postal and Pullum 1982, p. 125, 7a,b)

(a) I don’t want anyone [who continues to want] to stop wanting

(b) *I don’t want anyone [who continues to wanna] stop wanting

(6.18) (Postal and Pullum 1978, p. 17, 36a,b; 1982, p. 126, 8a,b)

(a) He says he’s going to annoy me

(b) He says he’s not gonna annoy me

(6.19) (Postal and Pullum 1982, p. 126, 9a,b)

(a) One must want (in order) to become an effective overconsumer

(b) One must wanna become an effective overconsumer

(6.20) (Postal and Pullum 1982, p. 126, 10a,b)

(a) I want to dance and to sing

(b) *I wanna dance and to sing

(6.21) (Postal and Pullum 1982, p. 126, 11a,b)
(a) I don’t need or want to hear about it

(b) *I don’t need or wanna hear about it

CMT Advocates argue for alternative explanations of this data. In response to (6.18), Chomsky and Lasnik (1978, 273) claim that wanna contraction is sensitive to a distinction between the different senses of go (e.g., they have different lexical entries) and that only the progressive aspect marker sense of go is of type ”contractable.” This leads Postal and Pullum (1982) to point out that (6.19), which contains the intransitive sense of want, is also a counter example, since one cannot plausibly claim that the following two assumptions hold simultaneously: (1) intransitive want is distinct from the sense of want which allows wanna contraction; and (2) the sense of want in Who do you want to meet Bill? is the same sense as the sense of want which allows contraction—if (2) doesn’t hold, then the CMT account makes no predictions at all. Aoun and Lightfoot (1984) attempt to account for (6.15,16,17,18,19,20), by adding to the CMT approach the requirement that only verbs which govern to can undergo wanna contraction. Towards this end, they propose analyses of (6.15b), (6.16b), (6.17a), (6.19a), (6.20a) and (6.21a) (their examples 1-6) in which want does not govern to. Aoun and Lightfoot (1984) propose analyses (6.15’), (6.16’), (6.17’) and (6.19’) (their 7, 12, 13 and 14, adapted to current GB notation) for (6.15), (6.16), (6.17) and (6.19). Under the definitions of government provided above, want does not govern the following instance of to in these sentences. For example, in (6.15’), the C and N nodes dominating the to preceding flagellate block government and thus wanna contraction by Aoun and Lightfoot’s proposal.

(6.15’) want [I\text{C} \text{N} [I\text{C} \text{I} [I\text{PRO to flagellate ...]]]] to become standard practice]]

(6.16’) [I\text{N} it, \text{I} \text{V} seems like [I\text{C} \text{I} [I\text{PRO to want][I\text{C} \text{I} [I\text{PRO to regret ...]]], ]]

(6.17’) I don’t want [I\text{C} \text{N} anyone [I\text{C} who continues to want]] to stop wanting]

(6.19’) [I\text{N} one, \text{I} \text{V} [I\text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{I} \text{
Postal and Pullum (1986, pp. 105-6) argue that assuming (6.15'), (6.16'), (6.17') and (6.19') are correct, Aoun and Lightfoot’s proposal will rule out not only the intended examples, but all bona fide cases of wanna contraction, as well. For example, \( \tilde{C} \) in (6.22) blocks government of to.

(6.22) (a) Do you want \( \lfloor C \rfloor \) to leave]]

(b) Do you wanna leave

While the definitions of government given above would rule out all cases of wanna contraction, Aoun and Lightfoot make the following assumptions not supported by recent GB analyses: (i) S‘ (our \( \tilde{C} \)) is the maximal projection of the INFL element to; and (ii) the definition of government is:

\[
\alpha (X') \text{ governs } \beta \text{ iff } \forall \phi, \phi \text{ a maximal projection, if } \phi \text{ dominates } \beta \text{ or the maximal projection of } \beta, \phi \text{ also dominates } \alpha, \text{ AND if } \phi \text{ dominates } \alpha, \phi \text{ also dominates } \beta \text{ or the maximal projection of } \beta. \quad \text{(Aoun and Lightfoot 1984, p. 467)}
\]

The italicized portions of the above definition are adapted from Belletti and Rizzi (1981). We interpret this definition as follows:

\[\alpha \text{ governs } \beta \text{ iff: (a) } \alpha \text{ m-commands either } \beta \text{ or the maximal projection of } \beta \text{ and (b) } \beta \text{ m-commands } \alpha \text{ or the maximal projection of } \alpha.\]

We find that (i) contrary to Postal and Pullum (1986), Aoun and Lightfoot’s assumptions do not rule out well-formed instances of wanna contraction; (ii) all other conceptions of government we are aware of pose exactly the problem Postal and Pullum point out; and (iii) Adopting the Aoun and Lightfoot (1984) definition of government would have unfortunate consequences for a plethora of other GB concepts, e.g., the PRO theorem. We therefore reject Aoun and Lightfoot’s definition of government and accept Postal and Pullum’s conclusion that government cannot be used as a condition on wanna contraction.

Postal and Pullum (1982) point out that (6.20) and (6.21) pose the following problems for the CMT approach:

(i) The fact that want is not adjacent to the second occurrence of to in (6.20a) does not explain why I wanna dance and sing is grammatical, but (6.20b) is not; and
(ii) The fact that want is adjacent to to, in (6.21a) does not explain why (6.21b) is ungrammatical.

Aoun and Lightfoot attempt to explain (6.20) and (6.21) using the Williams' (1978) version of Ross' (1967) principle of "across the board" rule application. In this approach, syntactic rules treat coordinates as special constituents which (i) occur in the same syntactic position on "alternate planes" and (ii) are unordered with respect to each other. This predicts that both instances of to in (6.20a) are considered adjacent to want and the verbs need and want in (6.21a) are considered adjacent. Thus if wanna contraction is applied across-the-board to both instances of to in (6.20a), I wanna dance and sing results. In contrast, the ungrammatical (6.20b) results from a non-across-the-board application of wanna contraction. Similarly, in examples like (6.21a), both verbs can contract with to yielding the grammatical I don't needa or wanna hear about it, but application of wanna contraction to one verb and not the other is prohibited. Regardless of which approach to explaining wanna contraction is correct, this general method of rule application appears useful for applying wanna contraction to examples like (6.20) and (6.21).

However, as Postal and Pullum point out, the relationship between the contractable verb and to in these sentences can hardly be viewed as adjacent in the phonological sense. Rather one has to take this abstract "across the board" view of adjacency. Unfortunately, this "across the board" view of adjacency does not hold for other types of contraction, as shown in Carden (1983, pp. 42-43.) For example contraction of an auxiliary (e.g., am, has, is, etc.) with the preceding word may occur with a subset of coordinates, as in examples (6.23a,b) (Carden's no. 18a and 19a.) Although it may be useful to apply contraction "across the board," the adjacency condition on contraction cannot be stated in terms of an "across the board" constraint, unless each type of contraction uses a different notion of "adjacency," an implausible proposition. Therefore the CMT analysis fails to account for examples (6.20) and (6.21.)

(6.23) (a) I'm ready and will leave at once

(b) Either Tom or Dick's got it;

According to Aoun and Lightfoot (1984, pp. 472-473), descriptive adequacy has never been the goal of the CMT approach. They claim that work on CMT
used facts about contraction as a probe into some properties of Universal Grammar. (Aoun and Lightfoot 1984, p. 472)

They cite properties of trace and PRO as principles of universal grammar. Regarding the nature of ecs, at best, the CMT approach brings out properties of case-marked wh traces lacking in other ecs. This result has no theoretical ramifications we are aware of since neither case-marked wh traces nor all ecs other than case-marked wh traces form a natural class under any GB analysis we are aware of. Aoun and Lightfoot (1984) also fail to show that the relation government effects wanna contraction—thus no light is shed on the nature of government under the CMT analysis.

The CMT approach is not based on universal principles because it does not hold universally. Postal and Pullum (1979, 704-705) note that there exist "liberal" dialects (spoken by e.g., Andrew Radford) of English in which sentences like *who do you wanna kiss you* are perfectly grammatical. This suggests that a particular analysis which blocks some instances of possible wanna contraction cannot be universal. Therefore if case-marked traces block wanna contraction in some subset of the existing dialects, this does not constitute evidence that case "universally" distinguishes ecs which behave like phonological properties from ecs which do not.

Carden (1983) demonstrates that CMT accounts shed no light on other instances of contraction and thus it is questionable that they are truly based on universal principles, as claimed. For example, case-marked traces occur between *is* and the preceding word in (6.24a,c), yet neither of these traces block contraction of *is* with the preceding words, as evidenced by the grammaticality of (6.24b,d) (Carden’s examples 25 and 26—he attributes 6.16d to Radford). According to Carden, if the CMT analysis was truly general, it would effect all morphological operations of type "contraction." However, case-marked traces only appear to block one kind of contraction. Furthermore, as evidenced by the examples in (6.25) (Carden’s 28), some non-case-marked traces appear to block auxiliary contraction. This obscures Chomsky’s (1981) claim that case is the property which makes ecs visible to the phonological component. In order to maintain this view: (i) one would need to find some other property visible to the "mind’s eye," e.g., the fact that t in (6.25a) is an adjunct trace; and (ii) one would need to determine in a general way which property of ecs effects which type of contraction.
(6.24) (a) Who t is going?
   (b) Who’s going?
   (c) Who do you think t is gonna to win?
   (d) Who do you think’s gonna to win?

(6.25) (a) We should tell John that the meeting is/’s Friday
   (b) We should tell John where the meeting is t Friday
   (c) *We should tell John where the meeting’s Friday

Aoun and Lightfoot (1984, p. 472, note 8) claim that the SS analysis is more of a descriptive generalization than an analysis because it is *far too complex to meet general requirements*. As far as we know, this is the only direct critique of the SS analysis during the contraction debate. As noted in Postal and Pullum (1986), it is very unclear which "general requirements" Aoun and Lightfoot refer to. We believe that Aoun and Lightfoot’s complexity assessment reflects the difficulty of representing a SS analysis in a constituent tree model. Below, we provide a GBUG based version of the SS analysis. GBUG makes it possible to clearly examine the bearing of the SS analysis in terms of GB theory.

We have shown that the CMT analysis fails on several grounds: (1) descriptive adequacy; (2) there is no sense in which it serves as a probe into universal grammar; and (3) it has not been shown how case marked traces exist in the mind’s eye, are adjacent to other constituents in any sense. This evidence combined with the evidence in previous sections shows that the ecs-are-real hypothesis is false.

6.8.3 The To Subsumption Analysis

This section motivates the To subsumption analysis, our account of wanna contraction based on the SS analysis of Postal and Pullum (1978, 1979, 1982, 1986). We propose the following:
The TO SUBSUMPTION (TS) analysis:

1. Ignoring the values of all phonology and category features, only verbs/semi-modals which take to complements and are subsumed by the lexical entry for to, provided in Figure 6.11, may undergo wanna contraction, a morphological process.\textsuperscript{17}

2. The lexical entry for the contracted form is derived as follows: (a) change the complement of the verb from a to infinitive complement to a bare infinitive complement; and (2) change the phonology of the entry according to a morphological rule, e.g., \textit{want} + to $\rightarrow$ wanna.

For example, Figure 6.12, the lexical entry for \textit{want} is subsumed by Figure 6.11 if we ignore the values of phonology and category features. Figure 6.13 the lexical entry for \textit{wanna} differs from Figure 6.12 in exactly the respects proposed above.

\begin{center}
\textbf{A CS Analysis Version of the Lexical Entry For Infinitival to}
\end{center}

\begin{verbatim}
Category: \textbf{INFL}  
Specifier: \textbf{N} 
Head-Proj: Category \textbf{N} 
  Head-Proj: Category \textbf{N} 
  Phonology: to 
  Complement: \textbf{K} Category: Verb 
  Specifier: \textbf{K} 
  Internal-Theta: \textbf{K} 
\end{verbatim}

\begin{center}
\textbf{Figure 6.11}
\end{center}
According to the TS analysis, wanna contraction combines both the phonological and syntactic features of two lexical items. It follows from this account that only those lexical entries with syntactic features that are compatible with the syntactic features of to allow this combination. Therefore the TS account, not only describes the wanna contraction data, but also explains the data.
We model the SS analysis of Postal and Pullum (PP) (1978, 1979, 1982, 1986) in GBUG as follows: Only verbs which are subsumed by Figure 6.14 may undergo wanna contraction. Clearly, this is the same class of verbs as satisfy the TS analysis. We based the TS analysis on the SS analysis, adding an explanation for why wanna contraction can occur.

Further support for the TS analysis is derived from the examination of the following English contraction rule which we call HAVE CONTRACTION:\(^{18}\)

\[(6.26) \text{Modal} + \text{Have} \rightarrow \text{Contracted form}\]

We propose the following analysis:

<table>
<thead>
<tr>
<th>A CS GB Version of the SS Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category: Verb</td>
</tr>
<tr>
<td>Specifier: (N)</td>
</tr>
<tr>
<td>Head-Proj: Category: INFL</td>
</tr>
<tr>
<td>Specifier: (k)</td>
</tr>
<tr>
<td>Head-Proj: Category: (k)</td>
</tr>
<tr>
<td>Head-Proj: Category: (k)</td>
</tr>
<tr>
<td>Phonology: to</td>
</tr>
</tbody>
</table>

**Figure 6.14**

The HAVE SUBSUMPTION (HS) analysis:

1. Ignoring the values of all phonology and category features, only lexical items which allow perfective have complements and are subsumed by the lexical entry for have provided in Figure 6.15, may undergo have contraction, a morphological process.

2. The lexical entry for the contracted form is derived as follows: (a) change the complement of the verb to a past participle verbal complement; and (2) change the phonology of the entry according to a morphological rule, e.g., *should +have \(\rightarrow\) should've*

The contracted forms *should've, could've, would've, might've* and *may've* are derived by (6.26) and phonological properties of *can, shall, will* prevent (6.26) from applying to derive *can've, shall've, may've*. 
*will’ve. Following R. Lakoff (1970, pp. 230-231) and Pullum and Wilson (1977: pp. 759, 772, 784), among others, we assume that modals can have either subject control (equi) or subject to subject raising properties—most modals are subject to subject raising verbs, but *may and *can are ambiguous between the two readings as exemplified by the examples in (6.27). Pullum and Wilson’s no. 67. (6.27) can mean "elephants are able to kill crocodiles" (root) on the subject control (root) reading. On the subject raising or epistemic reading, (6.27) can mean "elephants are likely to kill crocodiles". Similarly the root reading of *may is something like "is allowed to", and the epistemic reading is very similar to that of epistemic *can. See Pullum and Wilson (1977) for details.

(6.27) Elephants can kill crocodiles

Similarly, as shown below, the auxiliary verb have is also assumed to have a subject to subject raising interpretation. Therefore, we propose that one of the reasons that the contraction rule (6.26) is possible is that the lexical entries of modals are subsumed by the lexical entry for have, if we put aside category and phonology features. For example, Figure 6.16 is subsumed by Figure 6.15 if we ignore the values of phonology and category features. Figure 6.17, the lexical entry for would’ve derived in the manner proposed above.

A CS Analysis Version of the Lexical Entry For Perfective have

| Category: | Verb |
| Specifier: | N |
| Head-Proj: | Category: N |
| Head-Proj: | Category: N |
| Head-Proj: | Category: N |
| Head-Proj: | Category: N |
| Phonology: have |
| Complement: k Category: Past-Participle-Verb |
| Specifier: k |
| Internal-Theta: L |

Figure 6.15
Therefore both wanna contraction and have contraction are morphological processes which combine:

(1) one subject to subject raising lexical item (*have* or *to*);

(2) one verbal/INFL/modal lexical entry that has either a subject to subject raising or subject control property.
Both types of contraction follow from our TS analysis because, ignoring morphological and category features (i.e., only considering licensor/licensee features relevant), the lexical entries for have and to subsume the lexical entries of the verbal/INFL/modal lexical entries with which they combine. The output of both contraction operations have the same control or raising properties as the verb/INFL/modal which combines with have or to. For example both ought and oughta have subject to subject raising properties; both want and wanna have subject control properties; both should and should’ve have subject to subject raising properties; and both may and may’ve can be either (e.g., The alligator may’ve eaten the possum; The rhinoceros may’ve been in the zoo before the earthquake, but it isn’t there now).

This section motivated our version of PP’s analysis of wanna contraction. The relation between the lexical entries of the input and output of the contraction operation provides a linguistic explanation for how the contraction operation occurs and which lexical items contract. The fact that a similar constraint holds for have contraction further supports the TS/SS analysis because it shows that the TS/SS/HS analysis has some generality. GBUG provided us with the means for showing that Aoun and Lightfoot (1984) were incorrect in their assumption that the SS analysis is too complex to meet general requirements.

6.8.4 Wanna Contraction and Case Marking

In previous sections we provided grounds for rejecting the CMT analysis based on literature from the contraction debate as well as our own evidence against the ecs-are-real hypothesis. This section develops the following new hypothesis based on CMT:

The Case Assignment (CA) Analysis: Entries for verbs which assign case cannot undergo wanna contraction.

According to the CA analysis, only lexical entries for verbs of type "contractable" which satisfy the constraint represented by the FS in Figure 6.18 may undergo wanna contraction. The CA approach preserves the claim by CMT advocates that the case relation blocks wanna contraction, but does not assume any version of the ecs-are-real hypothesis. The CA analysis is also comparable to the TS analysis with respect to descriptive adequacy. Therefore the CA analysis is a strong candidate for
The main source of counterexamples to the CMT analysis provided by PP were examples in which a contractable verb V could not combine with an instance of to heading an infinitival phrase which was not the complement of V.\(^9\) Figure 6.18, however, limits wanna contraction to V and complements headed by to. The effect is that the CA approach, unlike the CMT approach, correctly accounts for (6.15) to (6.19), repeated below. Thus, in our view, the main source of the failure of the CMT approach is its reliance on the ecs-are-real hypothesis and the claim that the empty category was in any sense adjacent to V. In our view, the relation Complement(V,infinitival-phrase) makes contraction possible rather than adjacency. Wanna contraction is a morphological operation on two adjacent lexical items. However, this morphological operation should not be interpreted as an operation which converts one sentence in which V and to are adjacent, with one sentence in which they are not. The fact that want and to only contract when to is the head of the complement of want is evidence to this effect. This suggests that perhaps Aoun and Lightfoot (1984) were on the right track when they claimed that wanna contraction only occurs in sentences where want governs to.

(6.15) (Postal and Pullum 1982, p. 124, 3a,b,c)

(a) I don’t want it to become standard practice in this monastery to

flagellate oneself in public

(b) ?I don’t want to flagellate oneself in public to become standard practice in this
monastery
(c) *I don’t wanna flagellate oneself in public to become standard practice in this monastery

(6.16) (Postal and Pullum 1982, p. 125, 6a,b,c)
(a) To regret that one does not have seems like to want
(b) ?*It seems like to want to regret that one does not have
(c) *It seems like to wanna regret that one does not have

(6.17) (Postal and Pullum 1982, p. 125, 7a,b)
(a) I don’t want anyone [who continues to want] to stop wanting
(b) *I don’t want anyone [who continues to wanna] stop wanting

(6.18) (Postal and Pullum 1978, p. 17, 36a,b; 1982, p. 126, 8a,b)
(a) He says he’s going to annoy me
(b) He says he’s not gonna annoy me

(6.19) (Postal and Pullum 1982, p. 126, 9a,b)
(a) One must want (in order) to become an effective overconsumer
(b) One must wanna become an effective overconsumer

6.8.5 Descriptive Adequacy

There is a great deal of convergence between the CA and TS analyses. Both approaches only allow verbs taking to infinitival complements to contract. Under both approaches, we have proposed FS representations of constraints which must be satisfied by the lexical entry of any verb that undergoes wanna contraction. The CA analysis rules out case-assigning. The TS analysis selects the subset of TO-verbs in which the subjects of the verb and its complement are the same (or coreferential) as the only candidates for wanna contraction.

Six of the most common types of verbs taking infinitival complements in English are:
(1) Subject Control Verbs, e.g., \( I \), want \([\text{PRO}_i\, to\, leave]\)

(2) Object Control Verbs, e.g., \( I \) persuaded John \([\text{PRO}_i\, to\, leave]\)

(3) Prepositional Object Control Verbs, e.g., \( I \) beckoned to Mary \([\text{PRO}_i\, to\, leave]\)

(4) Arbitrary Control, e.g., \( I \), said \([\text{PRO}_j\, to\, leave]\)

(5) Subject to Subject Raising, e.g., \( I \), seem \([t_i\, to\, know\, you]\)

(6) Exceptional Case Marking/Raising to Object, e.g., \( I \) want \([John\, to\, leave]\)/\( I \) want John \([t_i\, to\, leave]\)

The TS analysis correctly selects verbs of type (1) and (5) as the only possible candidates for wanna contraction. The CA analysis, rules out (2) and (6) because the verbs are case assigners, predicting that (1), (3), (4), and (5) are "contractable." If we claim that verbs of type (3), and the case marking preposition are composite lexical items, e.g., beckon-to, count-on, prevail-upon, then these verb-preposition combinations would be case assigners and thus also ineligible for wanna contraction under the CA analysis. In previous accounts (Postal 1971: 213-214, Van Riemsdijk 1978: 218-223, Hornstein and Weinberg 1981), verb preposition combinations have been reanalyzed as complex verbs if they can undergo pseudo passive and these verb preposition combinations do, as shown in the following examples (the CMT approach accounts for type (3) verbs without reanalysis):

(6.28) (a) John was counted on/prevailed upon to deliver the secret documents
   
   (b) John was beckoned to to enter the house with no windows.

Another type of evidence may be found in right node raising examples like (6.29), in which a verb can conjoin with a verb-preposition pair.

(6.29) (a) John trusted—and counted on—his brother to deliver the love letter.
   
   (b) John pleaded with—and then ordered—his brother to deliver the documents.
The CA and CMT analyses incorrectly predicts type (4) verbs will undergo wanna contraction, but the TS analysis accurately predicts that they do not. Type (4) verbs are infrequent in English, including say, motion and gesture, as used in (6.30), the significant feature being that PRO must refer to some element outside of the sentence. (Verbs taking arbitrary controlled -ing phrase are more common, e.g., advise, permit, recommend.)

(6.30) John \(_i\) said/motioned/gestured PRO \(_j\) to leave

There is a subtype of type (1) verbs which take \(\overline{N}\) and to infinitive complements, e.g., promise and threaten in sentences like I promised/threatened Mary to take her to New Jersey. Promise and threaten are the only English verbs fitting this description we are currently aware of. The CMT/CA analyses correctly predict that no verb of this type undergo wanna contraction. However TS analysis only accounts for these examples if we assume a binary branching analysis of these verbs, e.g., I \([\text{promised Mary}] \text{ to feed the dog}\].

In and of themselves these distributional differences may not be sufficient for choosing between the TS and CA analyses because type (4) and promise-type verbs are so rare. However, as shown above, other factors favor the TS approach.

6.8.6 Summary

This section showed that the CMT analysis fails on grounds of descriptive and explanatory adequacy. This is significant since the CMT analysis is the main argument cited in support of the ecs-are-real hypothesis. Therefore we conclude that wanna contraction does not provide any evidence that the ecs-are-real hypothesis holds.

During the contraction debate, CMT advocates failed to appreciate or even discuss the SS analysis of wanna contraction, in part we surmise, due to the differences between the constituent structure tree model assumed in GB theory and the graph based model of RG, the theoretical framework assumed by PP. In particular, it is hard to represent the SS analysis in a constituent structure tree model which assumes ecs and no structure sharing. However, we show that in a GBUG based version of GB theory, the SS analysis can be represented quite easily. We show that the SS generalization represented
as a constraint is also a characterization of the lexical entry for infinitival to. Based on this observation, we motivate a rule of wanna contraction which combines both the syntactic and morphological properties of the verb/semi-modal and to—we show that only verbs which are subsumed by the set of syntactic properties in the lexical entry for to can contract with to. Furthermore, we show that the same type of model of contraction explains contraction of modals with the auxiliary verb have. Therefore the SS/TS analysis is preferred over the CMT analysis on grounds of descriptive and explanatory adequacy.

Further investigation showed that the CA analysis, a variant of the CMT analysis which does not maintain the ecs-are-real hypothesis, obtains descriptive adequacy comparable to the SS/TS analysis. However, we were unable to motivate the CA analysis on any explanatory grounds. The significance of the CA analysis is that it shows why the CMT analysis was descriptively inadequate. Ironically, the key concept that needed to be changed to correct the descriptive inadequacy was precisely the concept which CMT advocates claimed was the source of linguistic explanation—the ecs-are-real hypothesis.

6.9 Conclusion

This chapter motivated CS analyses of most constructions and syntactic categories for which previous GB accounts assumed an EC analysis. Furthermore, we demonstrated that ecs must be interpreted as properties of lexical items and cannot be viewed as comparable phonologically or psychologically to words or phrases in "the mind’s eye."

A comparison of EC and CS approaches revealed that: (1) binding and selection factors which needed to be accounted for explicitly in EC analyses followed from a CS analysis; and (2) EC analyses required more arcs to represent the same licensing relations as the corresponding CS analyses. Therefore CS analyses are preferred to EC analyses on grounds of economy.

A comparison of the ecs-are-real hypothesis and the lexical ec analysis revealed that no derivation can ever terminate that assumes that: (1) ecs exist in the mind’s eye; and (2) syntactic derivations are generated according to the generate and test approach of Chomsky and Lasnik (1977, 1991). Furthermore, we showed that phenomena purported to be evidence for the ecs-are-real hypothesis actually either: (1) motivated constituent relations represented by ecs rather than the ecs themselves
(nominal quantifiers in Japanese); or (2) argued against the ecs-are-real hypothesis (wanna contraction in English).

In conclusion, we have motivated the abandonment of most EC analyses of English in favor of CS analyses and shown that NOC-PRO, the one ec which we assume, is a feature of lexical items and phrases.
Notes

1Dougherty (1975) suggests that trace theory is motivated by a desire to do all semantic interpretation off of S-structure.

2Figures 6.3’ and 6.4’ are the same as Figures 3.59 and 3.53 respectively.

3In Figures 6.1 and 6.1’, we represent that the theta role assigned to the object of *by* has been suppressed according to Jaeggli’s (1986) account of passive. We have omitted an explicit representation of how this theta role gets re-assigned by means of the preposition *by*.

We also assume that *by* phrases are complements, even though according to Grimshaw (1990, pp. 109-110), they have similarities with both complements and adjuncts. (See Chapter 1, note 27.)

4As indicated in our discussion of f-command in Chapter 3, it is possible that control should be stated in terms of predicate licensor/licensee arcs, e.g., the control relation in Figure 6.5 could be described as External-Theta controls Head-Proj Internal-Theta External-Theta.

5Paul Postal (personal communication) observes that some types of extraction are impossible out of what he calls ‘antipronominal contexts’. For example, neither a pronoun, the gap left by topicalization, or parasitic gaps (the trace *t* is distinguished from the parasitic gaps *pg* in examples (iii), (vi) and (viii) can follow the copula in the environment:

Existential—*there* be _______

as shown below (compare (i), (ii) and (iii) with (iv), (v) and (vi). While this antipronominal restriction applies to these gaps, as shown in (v) and (vi), it does not appear to apply to the antecedent of the gap, as shown in (vii) and (viii) (assuming WH pronouns are a type of pronoun). This is evidence that gaps of topicalization and parasitic gaps should be represented using ecs, rather than structure sharing. (We assume that many of the ecs which GB characterizes collectively as Wh-traces, actually have very different properties.)

(i) There are gorillas in the zoo.

(ii) Gorillas, I (really) like to watch *t*.
(iii) Which gorillas did everyone who expected to find pg in the zoo look for t there?

(iv) *There are them in the zoo.

(v) *Gorillas, there are t in the zoo.

(vi) *Which gorillas did everyone who thought there were pg in the zoo look for t there?

(vii) Her, I (really) like to watch t.

(viii) Whom did everyone who expected to find pg in the zoo look for t there?

6 In my idiolect the object control readings of ask and badger are preferred, but the subject control (asking permission) readings are also available.

7 Postal (personal communication) observes a few instances in which subjects of gerunds are antipronominal contexts as exemplified by (i) and (ii) (see note 5). We assume NOC-PRO to be the subject of the gerunds in the (b) examples.

(i) (a) *PRO being you frightened the amnesiac.

            (b) *Her being you frightened the amnesiac.

(ii) (a) *PRO going down worried IBM

              (b) *Its going down worried IBM

We have not found many other examples showing that NOC-PRO subjects of infinitives or gerunds pattern with pronouns. However, we find that NOC-PRO subjects of gerunds only appear in types of gerundal complements that allow possessive subjects (including possessive pronouns).

For example, (iii) to (x) are instances of the available classes of gerundal complements listed in Wolff, Macleod and Meyers (1993)—we refer to these classes as Comlex classes. The Comlex classes, like Brandeis verb classes (cf. Grimshaw and Jackendoff 1981), can be broken down into substrings: oc stands for object control, sc for subject control, ac for arbitrary control (our NOC-PRO), possing for a gerund with a possessive subject, np for an N, p for a preposition, ing for a gerundal phrase with no subject, e.g., np-p-ing-oc stands for a complement structure consisting of an N followed by a preposition and a gerundal phrase without a subject, where the gerund is controlled by the N.
object.

(iii) I caught the dog eating the chocolate pudding (np-ing-oc)
(iv) I spent the time eating chocolate pudding (np-ing-sc)
(v) no np-ing-ac or np-possing
(vi) She stopped thinking about grapes (ing-sc)
(vii) She opposed (John’s/their/her) moving to Cleveland (possing, ing-ac)
(viii) I accused the dentist of using the wrong drill (np-p-ing-oc)
(ix) I spent all my money on purchasing silverware (np-p-ing-sc)
(x) They asked me about (John’s/their/his) moving to Cleveland. (np-p-possing, np-p-ing-ac)

These examples suggest that gerunds with NOC-PRO and possessive subjects pattern together and obligatorily controlled complements have a wider distribution over types of complement structures. Of course, the fact that we have not observed np-ing-ac or np-possing complements does not mean that they don’t exist. It is suggestive nevertheless. Assuming NOC-PRO patterns with lexical \( N \)s more strictly than other reported types of PRO argues for maintaining NOC-PRO as an empty category, i.e., as a property of lexical items as argued above.

8 Readers familiar with natural language processing (NLP) are cautioned that we use the term generation in a way that is standard in GB theory, but not in NLP. We use generation with the meaning \textit{algorithm for deriving potential analyses from a string of words}—NLP uses the term parsing with this meaning. In contrast the term generation in NLP usually refers to a \textit{procedure for deriving a string of words in a target language from a semantic or logical representation}.

9 For example, see Sudkamp (1988, p.89).

10 Williams (1980, p.206) the following constraint:

The C-command Condition on Predication: If NP \( \bar{N} \) and X are coindexed, NP must c-command X or a variable bound to X.

In Williams notation, coindexing marks predicates and their arguments, pro-forms and their antecedents, and ecs and their antecedents. Miyagawa’s Mutual C-command requirement is
essentially the same as a stricter version of the C-command Condition on Predication that Williams proposes (p. 204, note 1), but does not ultimately choose due to examples like:

(i) John became rich

In (i) rich does not c-command John. However, become is a subject to subject raising verb and (i) has the following structure:

(i) John [t', became [t rich]]

Therefore Miyagawa’s proposal accounts for (i) since rich c-commands t, a trace of John.

11 We believe the following aspects of Figures 6.9 and 6.10 are in need of refinement:

(1) We make essentially the same assumptions regarding the ni-phrase of Figures 6.9 and 6.10 as we did with the by-phrases of Figures 6.1 and 6.1’. At the same time as analyzing ni as a postposition, we analyze other case markers as specifiers of ηs, assuming that case is assigned by the same morphemes as in English. This conformed to Miyagawa’s set of assumptions that the case markers were part of ηs, although he did not describe these constituents in terms of constituent licensing relations. We believe that further study is required for a more uniform representation.

(2) In order to preserve Miyagawa’s Mutual C-command Requirement, we assume that NQ is a type of complement of the verb. However, this does not seem like a reasonable assumption, given that quantifiers in English are usually considered to be specifiers of nouns and non-theta-assigning predicates tend to be adjuncts. However, either of these latter choices would have violated the Mutual C-command Requirement.

Presenting an alternative analysis of these phenomena is beyond the scope of this dissertation.

12 Larry Horn, cited in G. Lakoff (1970), is the first person to our knowledge to suggest that some noun phrase present at an earlier stage of derivation blocks wanna contraction in the relevant examples. Selkirk (1972, 121-130) is credited with being the first to note that only traces of wh movement block wanna contraction.
The definition of adjacency assumed by CMT advocates is flawed as noted in Chomsky and Lasnik (1977, p. 478.) The requirement that case assignment occurs under adjacency behaves differently with respect to traces and phonetically realized Ns as evidenced by the ungrammaticality of (i) (their 150a) and the grammaticality of (ii) (their 150b). Therefore it is unclear why adjacency should treat traces like phonetically realized Ns with respect to wanna contraction.

(i) *John believes sincerely Bill to be the best man

(ii) Who does John believe sincerely t to be the best man?

Brame (1980: 88n, 1981: 286 note 13) notes that subjunctive and quasi-imperative clauses are incompatible with wanna contraction, as shown in the following examples from Brame (1981):

(i) a. The director requires that all the actors want to give their most.
   b. *The director requires that all the actors wanna give their most.

(ii) a. Want to do that and you’ll be rewarded.
   b. *Wanna do that and you’ll be rewarded.

As pointed out in Postal and Pullum (1982, p.132), neither the SS or CMT (CA) approaches account for these data.

We do not discuss the appropriateness of Aoun and Lightfoot’s constituent analyses here. See Postal and Pullum (1986, pp. 106) for arguments that purpose or rationale clauses as in (6.18) and (6.19) are V internal.

We believe that a more formal approach to coordination is desirable. Perhaps a GBUG analog to the GPSG approach discussed in Gazdar, Klein, Pullum and Sag (1985, 169-181) can be developed.

Our lexical entry for to is based on Pullum (1982) as well as current VP internal subject analysis assumptions. Figure 6.11 is based on the assumption that to acts much like a subject to subject raising verb, but is of category INF as following GB assumptions.
Accepting Pullum’s (1982) idea that *to* is actually a verb would simplify our analysis because the values of all the category feature values in Figure 6.11 would subsume the category features of candidates of wanna contraction. Therefore we would not have to stipulate category features as exceptions to the relevant notion of subsumption.

18 We assume that the contraction operation which combines pronouns and *have*, e.g. 
\[ I + have \rightarrow I've, \] is distinct from the contraction operation discussed here.

19 We assume that all *to* infinitival complements lacking complements are \( \overline{I} \)s headed by *to*. Following Koster (1984), this allows OC-PRO to be governed. This follows from a purely head-driven approach which projects maximal projections from heads. Under this view, \( \overline{C} \)s are not projected without evidence of a complementizer. (Note that this proposal may save Aoun and Lightfoot’s (1984) proposal that a contracting verb must govern *to*.)

Most GB approaches, \( \overline{C} \)s are projected above all these \( \overline{I} \)s headed by *to*, except for ECM complements (which vary between different accounts.) For a Barriers (Chomsky 1986b) model, the difference between having a \( \overline{C} \) above the \( \overline{I} \) is minimal, as shown by comparing (i) and (ii). In (i) an intermediate trace is required for antecedent government because the \( \overline{C} \) would be a barrier for government otherwise. Since \( \overline{I} \)s can only be barriers by inheritance, no intermediate trace is required in (ii).

(i) \( \text{Who}_i \text{ did Irving persuade you} [\overline{C}_t \text{'t' } \overline{I}_t \text{ to kidnap t}_] \]

(ii) \( \text{Who}_i \text{ did Irving persuade you} [\overline{I}_t \text{ to kidnap t}_] \]

As pointed out by Mark Baltin (personal communication), a crucial type of example for determining whether non-ECM infinitives are \( \overline{C} \)s or \( \overline{I} \)s would be examples like (iii) and (iv) in which an adjunct infinitive is embedded under a complement infinitive. Assuming that all infinitives are \( \overline{I} \)s, the complement infinitive would be a barrier by inheritance under our approach. Under the Chomsky approach, there would be no barrier, as both of these phrases would be \( \overline{C} \)s with an empty position for an intermediate trace.
(iii) Why did Bill persuade Mary \( \text{PRO} \) to drive to Boston \( \text{PRO} \) singing old ABBA songs? Is this an
intermediate trace? Why did Bill persuade Mary \( \text{PRO} \) to drive to Boston \( \text{PRO} \) singing old ABBA songs? Is this an
intermediate trace?

(iv) Why did Bill persuade Mary \( \text{PRO} \) to drive to Boston \( \text{PRO} \) singing old ABBA songs? Is this an
intermediate trace? Why did Bill persuade Mary \( \text{PRO} \) to drive to Boston \( \text{PRO} \) singing old ABBA songs? Is this an
intermediate trace?

There is no possible reading of the sentence Why did Bill persuade Mary to drive to Boston singing old ABBA songs? in which why can refer to why Mary is singing. Under the Chomsky (1986b) approach shown in (iii), both the infinitive phrases would be \( \text{C} \)s. Under our approach shown in (iv), they would both be \( \text{C} \)s because they are both obligatorily controlled and both have no overt complementizer—the fact that one is an adjunct phrase and one is a complement is irrelevant. Chomsky’s approach provides intermediate traces, so there are no barriers to government, thus making the wrong prediction that such a reading is possible. In contrast, our approach makes the correct prediction that this reading of the sentence is ungrammatical because the first \( \text{C} \) is a barrier by inheritance.

Our approach is preferred by economy of representation because we project fewer levels of phrase structure. In addition, our approach makes the correct predictions for examples like (iii) and (iv).
CHAPTER 7

What is Constituency?

7.1 Introduction

This chapter demonstrates that most constituent licensor arcs may be eliminated from FS representations because they are completely predictable from the predicate licensor/licensee and agreement licensor/licensee arcs which share the same source and target. 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’. Similarly, if one arc represents more than one licensor/licensee feature, e.g., like Head-Proj in previous chapters, we say that those features are parallel features. Any constituent licensor/licensee arc which is not parallel to any predicate licensor/licensee or agreement licensor/licensee arc, is a theoretical device for satisfying various constraints, i.e. the equivalent of intermediate traces in ec analyses.

We assume that in the grammar of a particular language, if all arcs with a label L is always parallel to arcs with label L’, then L’ subsumes L. In our GB grammar of English, each predicate licensor/licensee arc and agreement licensor/licensee arc is parallel to some constituent licensor/licensee arc. Therefore, we assume the following:

The PRIORITY OF CONSTITUENCY HYPOTHESIS (PCH): Each predicate licensor feature, predicate licensee feature, agreement licensor feature and agreement licensee feature is subsumed by some constituent licensor or licensee feature.

Each Head-Proj arc above represents some set of parallel licensor/licensee features, one of which is a constituent licensor feature. For example, the Head-Proj arc Head-Proj Head-Proj in Figure 3.23 (repeated below) simultaneously represents an adjunct licensor feature and a modification licensee feature. Although other predicate licensor/licensee features and agreement licensor/licensee features are parallel to constituent licensor/licensee features, these sets of parallel features have been
represented by separate arcs, e.g., the Specifier (Specifier licensee) feature and Quantifier (Quantification licensor) features are parallel in Figure 3.23, but are instantiated as two arcs rather than one.

The subsumption hierarchy in Figure 7.1 follows from our analysis. We assume that each arc labeled with an item in the hierarchy may represent more than one licensor/licensee feature, as specified in Figure 7.2.¹
A Subsumption Hierarchy for Licensor and Licensee Features in English

Figure 7.1
The following principle of economy of representation follows from the FSEP and the semantics of disjunctive feature labels.

The LICENSING REDUNDANCY PRINCIPLE (LRP): Given two arcs with the same source and target $A_F$ and $A_{F'}$ representing licensor/licensee features $F$ and $F'$, if $F \subseteq F'$, then $A_F$ is superfluous.

We have shown above that defined subsumption relations are equivalent to disjunctions. Therefore it follows that the feature Complement is equivalent to Complement-Case $\mid$ Internal-Theta $\mid$ Other-Complements, since it subsumes the features Complement-Case, Internal-Theta, as well as some finite set of other types of complement features. Several parallel arcs with labels $A$, $B$, $C$, $\cdots$ have the same interpretation as a single arc with a conjunction of labels $A$, $B$, $C$, $\cdots$. Therefore, since $A \mid B \mid C \cdots$ $\Lambda$ $A$ is equivalent to $A$, a constituent licensor/licensee arc $A$ with label $L$ may be eliminated if $A$ is parallel to one or more licensor/licensee arcs $A_1$, $\cdots$, $A_n$ with labels $L_1$, $\cdots$, $L_n$ and $L$ subsumes any $L_i$, where $1 \leq i \leq n$. For example, each Complement arc which is parallel to an Internal-Theta or Complement-Case arc may be eliminated.
Based on the LRP and the subsumption hierarchy in Figure 7.1, this chapter motivates replacing FSs like Figure 6.1 (repeated below) with FSs like Figure 7.3. Figure 6.1 contains several pairs of arcs A1 and A2 with the same source and target which are replaced by a single arc A1' in Figure 7.1. For example, Figure 6.1 contains the following types of arc pairs: (a) Complement, Internal-Theta; (b) Complement, Complement-Case; (c) Specifier, External-Theta; (d) Specifier, Specifier-Case; and (e) Specifier, Quantifier. These arc pairs correspond to the following types of arcs in Figure 7.1: (a) Internal-Theta; (b) Complement-Case; (c) External-Theta; (d) Specifier-Case; and (e) Quantifier. Our analysis leads to the conclusion that Figures 6.1 and 7.3 represent all the same licensing relations, but that Figure 7.3 represents these relations using fewer arcs and is therefore preferred by the FSEP. By using fewer arcs and a subsumption hierarchy Figure 7.3 is explicit about which features imply each other. For example, it is explicitly assuming that Internal-Theta features always are parallel to Complement features. In contrast, this assumption is independent of the Figure 6.1 representation.
A FS representing a constituent sharing analysis of

John was tickled by the gorilla

Figure 6.1
7.2 Surface and Non-Surface Constituent Relations

We assume that all Head-Proj features and all constituent licensor/licensee features which are parallel to Degree, Quantifier, Modifier, Complement-Case and Specifier-Case features are surface constituent features in English, where SURFACE CONSTITUENT FEATURES are licensor or licensee features of SURFACE CONSTITUENT RELATIONS, constituent relations which have the following properties, among others: (1) they obey constituency tests (see below); and (2) they are governed by the word order constraints provided in Chapter 4. We assume that the constituent licensee features which are parallel to Degree Licensor, Quantifier, Modifier, Complement-Case and Specifier-Case arcs are completely predictable. Therefore, by the LRP and Figures 7.1 above, we assume that these features are surface features which include the parallel constituent licensee feature. Thus Degree Licensor, Quantifier and Specifier-Case are each surface Specifier features; Modifier is a surface adjunct feature and Complement-Case is a surface Complement feature.
Previous literature in generative grammar, structuralist linguistics, and other frameworks have employed various tests for surface constituency. Thus constituents consisting of licensors and licensees of surface relations obey constituency tests, but constituents consisting of licensors and licensees of other types of licensing relations, i.e. theta relations on our account, do not always. (7.1) to (7.5) below demonstrate the effects of constituency tests. Adjacent items form a surface constituent if the same set of adjacent items can occupy different positions in pairs of sentences which paraphrase each other. For example, the sentences in (7.1) show that the bracketed groups of words form constituents. A group of words form a constituent if this unit functions as if it fills a gap in so-called VP deletion constructions as exemplified in (7.2), where the gap is indicated by ____.

A group of words forming a constituent can be the answer to a question, be coreferential with some pronominal, or be a coordinate with another constituent, as exemplified in (7.3), (7.4) and (7.5). In GB theory, each of the above tests determine that some set of elements forms a surface constituent. Thus the following data show that the bracketed phrases are each surface constituents in (7.1) through (7.5).

(7.1) (a) Sam ate [green eggs and ham]

(b) [Green eggs and ham] were eaten by Sam

(7.2) (a) Sam [ate green eggs and ham] and so did I ____

(b) Sam [ate green eggs and ham] and I did too ____

(c) [Eat green eggs and ham] is what Sam did ____

(7.3) (a) What did Sam do? [Eat green eggs and ham]

(b) What did Sam eat? [Green eggs and ham]

(7.4) (a) Sam [ate green eggs and ham] and I did it too.

(b) [The big green monster] ate itself.
(7.5) (a) John [ate green eggs and ham] and [gave blood] on the same day.

(b) Do you want to [sleep in a hotel] or [drive all night long].

(c) Underdog will either [swallow his secret energy pill] or [be eaten by monsters.]

Internal-Theta is a surface (complement) feature when the value of Internal-Theta is not an $\overline{N}$ and GB assumptions do not required a case relation in order for the $\overline{N}$ to be in a surface position. Also Internal-Theta subsumes the Minimal A-chain (MAC) features which is a surface feature, considering the fact that is is subsumed by Complement-Case in Figure 7.1. Thus, Internal-Theta is unlike the other features in Figure 7.1 in that it is not consistently a surface or nonsurface feature.

External-Theta and Internal-Theta (when it has an $\overline{N}$ value) are not surface constituent features, but rather are viewed as constituent features based on theoretical considerations (see below) and analogy. In finite sentences with active transitive matrix verbs, Internal-Theta licensees are always complements (of the verb) and External-Theta licensees are always specifiers (of INFL). Therefore, it is unsurprising that these licensee features are assumed to be parallel to complement and specifier features respectively. In other sentences, where these regularities do not hold, further analogy is assumed. For example, there are no constituent structure tests which show that eaten and green eggs and ham form a constituent in (7.6) Nevertheless, according to GB theory, the relations Complement(eaten, green eggs and ham) and Specifier([eaten t], green eggs and ham) hold, as represented by t and t' in (7.6). The complement licensee feature represented by t is justified either by parallels between passive and active sentences or by parallels between passive and active lexical entries for verbs, i.e., t represents the "initial" or D-structure position. GB assumes that both passive and active verbs "assign" theta roles to the same syntactic position. Thus in an ec analysis cast in GBUG, the relations Internal-Theta(V,$\overline{N}$) and Complement(V,$\overline{N}$) hold for an active verb V and the corresponding relations Internal-Theta(V',NP-trace) and Complement(V',NP-trace) hold hold for the corresponding passive verb V'. The specifier licensee feature represented in (7.6) by t' is justified by the combinatorial properties of the passive verb, e.g., the subject of be is coreferential with the subject of its complement whether that complement is a passive, an adjective, etc. (cf Stowell 1978).
(7.6) [Green eggs and ham] were [t’, t] by Sam and I.

GB theory assumes that each relation Internal-Theta(X,Y) is accompanied by a relation Complement(X,Y), as stated in the following citation from Chomsky (1981).

Consider structural configurations of the form (5), where \(\alpha\) is an immediate constituent of \(\gamma\):

(5)  
   (i) \([\ldots \alpha \ldots \beta \ldots]\)  
   (ii) \([\ldots \beta \ldots \alpha \ldots]\)

If \(\alpha\) is a 0-level category of X-bar theory, then the category \(\beta\) must satisfy the subcategorization frame of \(\alpha\). In this case, we will say that \(\alpha\) subcategorizes the position occupied by \(\beta\).

Clearly, \(\theta\)-marking is closely related to subcategorization. The two notions are not identical however. Thus if \(\alpha\) is VP in (5) with \(\gamma = S\), then \(\alpha\) may \(\theta\)-mark the NP subject \(\beta\) but does not subcategorize it. But we must require that if \(\alpha\) subcategorizes \(\beta\), then \(\alpha\) \(\theta\)-marks \(\beta\). Were this condition not to hold, then the definition of "subcategorize" just given would fail to capture the intended sense of this notion. (Chomsky 1981, 36-7)

Chomsky later incorporates this requirement into the Projection Principle. The projection principle is usually defined as something like the following:

**Definition 7.1: The Projection Principle**

Representations at each syntactic level (i.e., LF, and D- and S-structure) are projected from the lexicon, in that they observe the subcategorization properties of lexical items. (Chomsky 1981, 29)

Adding the requirement that heads \(\theta\)-mark their complements, Chomsky provides the following alternative definition of the projection principle—currently this requirement is assumed to be a defining property of GB theory. Note that in the definition below, \(L_i\) and \(L_j\) are variables over levels of representation, like D-structure, S-structure and LF.

**Definition 7.2: The Projection Principle (revised)**

(i) if \(\beta\) is an immediate constituent of \(\gamma\) in (5) at \(L_i\), and \(\gamma = \alpha\), then \(\alpha\) \(\theta\)-marks \(B\) in \(\gamma\)

(ii) if \(\alpha\) selects \(B\) in \(\gamma\) as a lexical property, then \(\alpha\) selects \(B\) at \(L_i\)
On the basis of Chomsky’s projection principle, each Internal-Theta arc in each GBUG is parallel to some Complement arc, i.e., the Complement arc is completely predictable from the presence of the Internal-Theta arc. This follows from our assumption that Internal-Theta is a type of Complement relation. Given the LRP, this motivates eliminating all Complement arcs with the same source and target as some Internal-Theta arc. However Chomsky’s projection principle makes a stronger claim. Interpreted in GBUG, the projection principle states that each Complement arc has the same source and target as each Internal-Theta arc and vice versa. If this were true, then Complement and Internal-Theta features would be equivalent. In later sections we show that although Complement $\subseteq$ Internal-Theta, the reverse does not hold.

Chapter 3 provided evidence based on Marantz (1981, 1984) that an External-Theta arc always has the same source and target as some Specifier arc. On this basis, we assume that the feature External-Theta is a type of Specifier feature, as shown in Figures 7.1 and 7.2. Chomsky (1981) states that

If an argument appears in subject position at D-structure, then the position is $\theta$-marked at every level. If no subject appears at D-structure, then the position is not $\theta$-marked at any syntactic level. (Chomsky 1981, p.41)

This requirement, together with the projection principle are called the EXTENDED PROJECTION PRINCIPLE. In GBUG, the extended projection principle is the assumption that Specifier $\subseteq$ External-theta. Note, that unlike the projection principle, the stronger claim that all specifier positions are external-theta positions is not made by either Chomsky or ourselves.\(^{5,6}\)

To summarize, surface constituent relations are testable. Each set consisting of a constituent licensor and all its licensees forms a constituent which satisfies various constituency tests and is subject to word order constraints. In contrast Internal-Theta and External-Theta relations are nonsurface (or D-structure) relations. They are modeled as constituents in order to maintain theory-internal parallels with surface constituent relations. One benefit of these theory internal parallels is that they enabled theta relations to be modeled in a constituent structure tree. Furthermore, it enables us to
assume that all Internal-Theta licensees are complements, both surface Internal-Theta licensees (non-Ns) and nonsurface Internal-Theta licensees (Ns)

We left N-Agreement licensor/licensee features out of Figure 7.1 because these features are not part of the grammar of English. Based on the data discussed in Chapter 3, these features would have different positions in our subsumption hierarchy in Spanish and Russian. In Spanish, the N-Agreement licensee feature subsumes Head-Proj and the N-Agreement Licensor Feature comes in two varieties: N-Agreement-Specifier and N-Agreement-Adjunct, the former subsumed by Quantifier, and the latter subsumed by Modifier. In Russian, the quantifier is the N-Agreement licensor, and the N-Agreement Licensee is always some projection of the head. Unfortunately, we cannot assume that Head-Proj is always the N-Agreement licensee feature because, this would result in our analysis of quantified Ns having two paths with the same source and target that consist entirely of Head-Proj arcs, thus violating functionality. This finding shows that some licensor/licensee features, perhaps should not be viewed as constituent licensor/licensee features (or as parallel to constituent licensor/licensee features).7

7.3 Believe Type Verbs

7.3.1 Introduction

Exceptional Case Marking analyses of sentences like (7.7) are in conflict with our assumption that the feature Complement-Case is subsumed by the feature Complement and thus, if correct, may be viewed as evidence against the PCH. In this section, we show that the Raising to Object (RO) analysis is preferred over the Exceptional Case Marking (ECM) analysis on grounds of economy, as well as on the basis of some recent evidence presented in Postal and Pullum (1988) and Dougherty and Leacock (1993). This evidence provides the basis for maintaining the PCH and rejecting the ECM analysis.

(7.7) (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]
Given a matrix *believe*-type verb V, followed by \( \bar{N} \) and infinitival *to*, the relations Complement-Case(V,N) and Specifier(*to*,N) hold under an Exceptional Case Marking ECM analysis (Chomsky 1980b, 1981, among others). GB precludes the possibility that the relation Complement(V,N) holds because clause (i) of Chomsky’s projection principle (see Definition 7.2 above) requires that all heads theta-mark their complements.8 Thus if the ECM analysis is correct, the following subsumption relation does not hold: \( \text{Complement} \subseteq \text{Complement-Case} \). On this basis, parallel Complement and Complement-Case arcs cannot be reduced to Complement-Case arcs; and parallel Complement-Case and Internal-Theta arcs cannot be reduced to MAC arcs as assumed above. Thus an ECM analysis would force us to represent one additional arc in representations of all case assigning verbs other than *believe*-type verbs.

(7.7’) lists EC versions of raising to object analyses (based on Rosenbaum 1967, Postal 1974, among others) of sentences (7.7). In a CS version of RO analyses, the following relations hold: Complement-Case(V,N) and Specifier(*to*,N), where the relation Complement(V,N) is implied since Complement \( \subseteq \) Complement-Case.9 Under the RO analysis, the matrix verb licenses a complement relation, without licensing an internal-theta relation in violation of clause (i) of the projection principle.

(7.7’) (a) John believes ketchup \(_i\) [ti to be a vegetable]

(b) Cecil considered ketchup \(_i\) [ti a vegetable]

(c) Mildred made Herman \(_i\) [ti angry]

(d) John saw Irving \(_i\) [ti stealing the documents]

Figure 7.4 is a CS representation of either (7.7a) and (7.7a’) reduced to the fewest number of arcs by the LRP. Under the RO interpretation, the Complement-Case arc represents both a Complement feature and a Complement-Case feature. Under the ECM interpretation, the Complement feature is not represented. Although it appears that the RO and ECM analysis are equivalent by the FSEP, this equivalence is only apparent. A comparison of Figure 7.5 and 7.5’, alternative analyses of the sentence *Mary ate the sandwich*, reveal that the RO analysis is preferred by the FSEP. Figure 7.5 assumes an RO analysis of *believe*-type verbs. Therefore the features Complement, Internal-Theta and...
Complement-Case are represented as one MAC arc (see note 1). In Figure 7.5', which assumes an ECM analysis of believe-type verbs, an Internal-Theta arc represents the Complement and Internal-Theta features, and the Complement-Case feature is represented by a separate arc. The RO analysis is preferred over the ECM analysis by the FSEP because it allows FSs to reflect the generalization that Complement-Case always represents both case licensee and constituent licensee features.

A CS analysis of
John believes ketchup to be a vegetable

Category: [I] INFL
Tense: [i] Present
Specifier-Case: [k] Category: Noun
Phonology: John
Head-Proj: Category: [h]
  Tense: [h]
  Head-Proj: Category: [h]
    Tense: [h]
      Internal-Theta: Category: [h] Verb
      External-Theta: [k]
      Head-Proj: Category: [h]
        Head-Proj: Category: [h]
          Phonology: believes
          Internal-Theta: Category: [m] INFL-Verb
          Specifier: [n] Category: Noun
          Phonology: ketchup
          Head-Proj: Category: [m]
            Head-Proj: Category: [m]
              Phonology: to be
              Internal-Theta: Category: [o] Noun
              External-Theta: [n]
              Head-Proj: Category: [o]
                Phonology: A vegetable

Complement-Case: [n]

Figure 7.4
CS Analysis Number 1 for
Mary ate the sandwich

Category: INFL
Tense: Past
Specifier-Case: Category: Noun
Phonology: Mary
Head-Proj: Category:
Tense: [ ]
Head-Proj: Category: [ ]
Tense: [ ]
Internal-Theta: Category: Verb
External-Theta: [ ]
Head-Proj: Category [ ]
Head-Proj: Category [ ]
Phonology: ate
MAC: Category: Noun
Quantifier: Category: determiner
Phonology: the
Head-Proj: Category: [ ]
Phonology: sandwich

Figure 7.5
The remaining subsections provide evidence which motivates an RO analysis. We examine sentences containing expletive N complements discussed in Postal and Pullum (1988). These constructions provide evidence that the projection principle is incorrect and thus that the RO analysis is permitted. We examine sentences, from Dougherty and Leacock (1993), which have a well-formed S-structure under an RO analysis, but not under an ECM analysis. These data argue for rejecting clause (i) of the projection principle, rejecting the ECM analysis, and adopting the RO analysis. Thus the RO analysis is motivated by both descriptive adequacy and economy.

### 7.3.2 Expletives in Complement Positions

Postal and Pullum (1988) discuss RO analyses in light of current GB assumptions. They show that the only GB principle ruling out RO analyses is clause (i) of the projection principle, which they (p. 638, no. 8) state as follows:
(7.8) If $\alpha$ subcategorizes the position $\beta$, then $\alpha \theta$-marks $\beta$

In addition to prohibiting RO analyses, (7.8) 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 1986a, 131-144). Postal and Pullum show that contrary to this prediction, expletives do in fact occur in complement positions. They show that many of their examples with expletives require RO analyses, ECM analyses being unavailable. If Postal and Pullum’s arguments are correct, the RO analysis is available. As shown above, the LRP and the FSEP prefer RO analyses over ECM analyses. Therefore Postal and Pullum’s evidence is crucial to our analysis.

Postal and Pullum discuss the major theoretical issues which previous literature used to choose between ECM and RO analyses. They show that, other than (7.8), all of these issues have disappeared.

GB theory no longer bans *vacuous movement rules*—movement between Constituent positions representing adjacent string positions. Otherwise many standard GB analyses, including (7.9) (Postal and Pullum’s example 17) would violate this condition.

(7.9) [Who $i$ [t made this coffee]]

In order for a verb $V$ to case mark an $\bar{N} N$, $V$ must govern $N$. Various proposals regarding the infinitival complements of ECM verbs seek to explain how it is possible for $V$ to govern the subject of the embedded clause. Chomsky (1981) claims that there are usually two phrasal nodes between a verb and the subject of its clausal complement, as shown in (7.10)

(7.10) (a) I believe $[\text{CP e } [\text{IP John is here}]]$

(b) I want $[\text{CP e } [\text{IP PRO to be here}]]$

However, case marking cannot occur across two phrasal nodes because case marking must occur under government and two phrasal nodes block government. Thus case marking by *believe*-type verbs is *exceptional*. Chomsky argues that this exceptional case marking (ECM) is only possible because English has a marked rule of S’-deletion [i.e., $\bar{C}$-deletion] for complements of the *believe* category, permitting the verb to govern the subject of the embedded complement...' (Chomsky 1981, p.66)

341
The particular version (based on Chomsky 1986a) of S’-deletion we assume is that the infinitival complements selected by believe-type verbs are I’s, rather than C’s with the C deleted, as shown in (7.11a).

(7.11) (a) John believes I ketchup to be a vegetable]
(b) John is likely I t to be here]
(c) John seems I t to like you]
(d) John believes ketchup I t to be a vegetable]

Chomsky (1981, 66-67) and subsequent work also require S’-deletion for complements of subject to subject raising predicates like seem and likely in sentences like (7.11b,c), in order that the NP-trace be governed (by seem or likely.) Postal and Pullum point out that this same concern would arise for RO analyses and thus believe-type verbs would have to be of type S’-deletion under RO analyses. Postal and Pullum (p. 65) cite Chomsky (1981) regarding this fact:

raising to object in essentially the sense of Postal (1974), with base-generated [NP e] as object of the verb...is ruled out by the requirement that trace be governed...unless we require S’-deletion in addition, to raising, in which case raising is superfluous. (Chomsky 1981, p.146, note 91)

However assuming that I’s are selected and are not the result of some S’-deletion process, the raising operation is not superfluous, particularly if it allows us to combine the Complement-Case feature and the Compliment feature under the LRP.

Postal and Pullum show that the significant difference between the RO and ECM approaches is that the former is based on a theory without assumption (7.8) and the latter is based on a theory with (7.8). Postal and Pullum show that expletive it purported to only occur in non-θ-marked positions does occupy complement positions in a wide range of sentences including those listed in (7.12-14) (These correspond to Postal and Pullum’s examples 20a,b, 21j, 22a, 23b, 30a, 32d, 38b, 39e, 43, 44 and 46. (7.12a,b) originate with Jespersen (1937, 63) and (7.12c) originates with (Chomsky 1981, 190).) On the basis of these examples, Postal and Pullum show that the prediction that expletives cannot occur in complement positions is false. From these data it follows that assumption (7.8) must be abandoned and RO analyses are legitimate. As shown above, LRP and FSEP considerations also prefer RO analyses.
(7.12) (a) I take it that you will pay
(b) He never gave it a thought that Bolshies 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.

(7.13) (a) I figured/made/reasoned it out.
(b) You two are going to battle it out when things go wrong between you.

(7.14) 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.

According to Postal and Pullum, the examples in (7.12) are extraposition constructions in which the expletive\textsuperscript{12} \textit{it} occurs in complement position and is "linked" (coreferential) with some other complement. Postal and Pullum reject a possible exceptional case marking (ECM) alternative on the following grounds.\textsuperscript{13} Under the apparent ECM interpretations of these sentences, expletive \textit{it} is a specifier of a "small clause" complement and is exceptionally case-marked by the matrix verb. For example, in (7.12c) the relations Specifier(\textit{for him to win}, \textit{it}) and Complement-Case(\textit{hate}, \textit{it}) would hold, but the relation Complement(\textit{hate}, \textit{it}) would not hold. The constituent structure analysis would be:

\begin{equation}
(7.12c') \text{John would hate [it [for him to win]]}
\end{equation}

Postal and Pullum show that examples like (7.12d) are problematic for a small clause analysis because \textit{to the candidate}, the \textsuperscript{P} complement of the verb, is in between the \textit{it} and the \textit{that} complement. Barring discontinuous constituency, the \textsuperscript{P} would have to be a part of the small clause, as shown:

\begin{equation}
(7.12d') \text{They never mentioned [it [to the candidate]}
[that the job was poorly paid]].
\end{equation}
Postal and Pullum show that this Constituent analysis cannot be correct because, semantically, *to the candidate* is a separate complement of the verb and must be assigned a separate theta role. Furthermore there is no relationship, semantic or otherwise, between the phrases *to the candidate* and [*that the job ...*] which would justify their being co-constituents. We find this argument convincing.

Postal and Pullum also claim that the following fact is incompatible with a small clause analysis: an adverb modifying the verb may intercede between *it* and the other complements, but not between the verb and *it*, as shown (Postal and Pullum’s no. 26):

(7.15) (a) They mentioned it immediately to the candidate that the job was poorly paid.
(b) *They mentioned immediately it to the candidate that the job was poorly paid.
(c) They mentioned (*immediately) that.

In our view, this may however be explained by the adjacency condition on case assignment since the adverb in these examples intervenes between the case assigner and the expletive. However, since the case filter only requires that *Ns receive case to be "visible" for theta marking, it is unclear why an expletive requires case (unless it is linked to a non-case-marked *N, e.g., There was a man in the room.*

Stroik (1990) argues against Postal and Pullum’s analysis of these expletives, limiting his comments to the type of examples found in (7.12). He makes the following claims:

(1) The S-structure of sentences like (7.12a) is:

(7.12a’) I take [\(\text{C} \] [\*\(\text{e}_i\) \[\text{C} \]] \[\text{C} \text{that you will pay}\]]

(2) The matrix verb, e.g., *take*, exceptionally case marks the expletive *it*.\(^{14}\)

(3) The \(\overline{C}\) in (7.12a’) moves from the position occupied by *e* at D-structure to the S-structure position shown above.

(4) \(\overline{C}\) movement occurs over adverbs as shown in (7.16) (his 13, 14, 16). (7.16b) is the S-structure of (7.16a) and (7.16c) is the D-structure.

(7.16) (a) We regret it very much that we could not hire Mosconi.
(b) \[\overline{f}_i \text{We} \overline{f}_g \[\overline{v}_i \text{regret} \overline{c}_j \text{it e}_j \text{very much}] \overline{c}_k \text{that we could not hire Mosconi].}\]
(c) \[ T^I \in V \Downarrow \text{regret} \in V \Downarrow \text{it} \in V \Downarrow \text{that we could not hire Mosconi} \Downarrow \text{very much} \]].

Stroik’s analyses of these constructions requires at least two highly unusual proposals: (1) \( C \)s headed by the complementizer that somehow license expletive specifiers; (2) an \( X \) may undergo movement. Most accounts, e.g., Chomsky (1986a, p.4), assume that only \( X^0 \)s and \( \bar{X} \)s may undergo movement. We show that Stroik’s account cannot be correct.

Stroik does not provide any other examples of it—\( that \) clause constituents in English, because none exist. There are no small clauses headed by \( that \) clauses, and if they were, the specifier would be considered an argument position according to the extended projection principle. For example, there is no verb consider’ like consider except its complement could be a \( that \) clause predicate with an \( \bar{N} \) subject. For example, (7.17) cannot mean that John thinks that the idea is characterized by the predicate that Mary went to the movies—this is true regardless of which English verb is substituted for considered’

(7.17) John considered’ \( \in C [\text{the idea} \in C \text{that Mary went to the movies}] \]

Stroik’s support for \( X \) movement comes from a number of sources, each of which we find problematic.

Williams (1986, pp. 286-287) claims that \( N \), not \( \bar{N} \)s undergo movement in relative clauses due to the following "reconstruction" effects. Williams claims that (7.18a) is ungrammatical and (7.18b) is grammatical because the extracted constituent does not contain the specifier, i.e., it is the \( \bar{N} \) (7.18) = Williams 59b,c; Stroik’s note 7, examples (a) and (b.) He claims that in (7.18a), the \( \bar{N} \) moves into the position of the trace by reconstruction,\(^{15}\) so that the anaphor can be bound by John and Mary, the subject of the embedded clause. According to Williams, this option is unavailable for (7.18b) because the anaphor is not in the \( \bar{N} \).

(7.18) (a) I saw the pictures of each other that John and Mary took.

(b) *I saw each other’s pictures that John and Mary took.

(c) I saw the gorillas’ pictures of each other that John and Mary stole.
This cannot be the correct analysis. Otherwise counterfactually, the gorillas would not be coreferential with each other in (7.18c). We conclude that the ungrammaticality of (7.18a) must be due to other factors. Therefore, Stroik can not use (7.18) as evidence of other instances of $\bar{X}$ movement. Even if Williams' analysis is correct, this would be an example of $\bar{N}$ movement at LF (or NP-structure). It would be insufficient evidence for S-structure movement of a $\bar{C}$.

Stroik argues that the following sentence is an example of right node raising (RNR) in which the $\bar{C}$ did John leave undergoes movement.

(7.19) When—and why—did John leave?

However, a more natural analysis for (7.19) is that the conjoined complementizers when and why are bound to a wh trace adjunct as shown:

(7.19') [When and why]$_{t}$ did John leave $t_{i}$?

Postal (1974) defines RNR as follows:

(86) (a) Jack may be—and Tom certainly is—a werewolf.
(b) Tom said he would—and Bill actually did—eat a raw eggplant.
(c) Tony should have—and Pete probably would have—called Grace.
(d) Terry used to be—and George still is—very suspicious.

Roughly, given certain paired sentences of identical constituents in disjoint clauses, RNR places a double of the sequence on the right by Chomsky adjunction and deletes all original occurrences. Note the characteristically sharp intonation breaks at the points of deletion indicated by dashes in (86). (Postal 1974, p.126)

Stroik's claim that (7.19) is an example of RNR is similar to the claim that the conjunction of subject $\bar{N}$s is actually RNR, e.g.,

(7.20) John—and Mary—went to the store.

In our view, it is more likely that (7.19) and (7.20) contain the conjunction of two words than the conjunction of two clauses with the bulk of the clauses duplicated and deleted, including the main verb. Stroik's analysis of (7.19) is ad hoc—if (7.19) is RNR, then so are most sentences containing conjunctions. RNR sentences have one sharp intonation break after each conjunct at the site of deletion—(7.19) and (7.20) lack these intonation breaks. Furthermore, the deleted (or moved) item in RNR is
typically the complement of some verb or modal, not a constituent including the main verb. Thus contrary to Stroik, (7.19) is not an example of \( \overline{C} \) movement (or deletion.)

Stroik also claims that the \( \overline{C} \) you think that Mary left raises over the \( \overline{P} \) by tomorrow in (7.21a), but the \( \overline{I} \) you love me will not raise over the \( \overline{P} \) in (7.21b) (Stroik’s nos. 18a,b.)

(7.21) (a) I want you to tell me why, by tomorrow, you think that Mary left.

(b) *I want you to tell me that, by tomorrow, you love me.

For Stroik, the assumption that why is the specifier of a \( \overline{C} \) and that is the head of the \( \overline{C} \) are crucial for determining that these two seemingly comparable strings of words actually form two different types of constituents a \( \overline{C} \) and an \( \overline{I} \) respectively. Note however that only adjunctive wh words, e.g., why, where, how, tolerate an intervening phrase. In contrast, argument wh words like what and who are ungrammatical in this same context, e.g.,

(7.22) (a) *I want you to tell me what, by tomorrow, you told Mary.

(b) *I want you to tell me who, by tomorrow, murdered the monkey.

Therefore Stroik’s analysis cannot be correct. He has only shown that there is a difference between adjunct wh clauses and other \( \overline{C} \)s, but has failed to demonstrate that a \( \overline{C} \) movement analysis is desirable in any way. If the strings of words that Stroik claims do undergo movement, there is no particular reason to claim that these strings form \( \overline{C} \)s or \( \overline{I} \)s. (7.22) shows that Stroik’s proposal for \( \overline{C} \) movement does not distinguish the grammatical examples from the ungrammatical ones and therefore does not buy Stroik any theoretical ground. \( \overline{X} \) movement has not been shown to occur in any other structure and positing \( \overline{X} \) movement would not contribute anything to the analysis of these data.

We conclude that Stroik’s data do not motivate a \( \overline{C} \) movement analysis for any of his examples.

Even assuming that Stroik’s analysis is correct for the examples he provides, it is unclear how Stroik would account for (7.12g,h), repeated below:

(7.12g) John will see to it that you have a reception.

(7.12h) He made a point of honor of it, I recall, to respect agreements he had negotiated.
Stroik would have to make proposals such as the following to account for this data: (1) In comple-
ments of some small set of nouns including honor, of is a complementizer which case-marks the
specifier of the following infinitive; (2) To takes $\bar{C}$ complements in this small set of environments;
and/or (3) see to and make a point of honor of are both complex (ECM) verbs. (1) requires that a small
set of nouns select $\bar{C}$s headed by a unique complementizer which is of the category preposition in all
other environments. (2) requires that the preposition to takes a $\bar{C}$ small clause complement in exactly
one syntactic environment in English. We know of no other proposals requiring a preposition to take a
$\bar{C}$ small clause complement headed by the complementizer that.\textsuperscript{16} In our ideolec, proposal (3) is not
supported by the standard tests for complex-verb-hood, as shown in (7.23)\textsuperscript{17}.

(7.23) (a) *It will be seen to that you have a reception.

(b) *It will be made a point of honor of to respect the agreements.

Therefore it seems unlikely that Stroik can extend his analysis to these additional examples.

Stroik (p. 13) explicitly states that he will not attempt to deal with "unlinked" expletives,
expletives without coreferential antecedents, as exemplified in (7.13-14). Presumably to maintain his
position, Stroik would have to come up with further clearly exceptional proposals.

We found several other faults in Stroik’s proposal including: his proposal requires verbs like
give to be exceptionally case marking verbs only for this set of constructions, but no others; and he
erroneously claimed that all $\bar{N}$s require case, when actually only $\bar{N}$s in theta-marked positions require
case, e.g., the following grammatical sentences contain non-case marked $\bar{N}$s.\textsuperscript{18}

(7.24) (a) John graduated from college [a genius]

(b) Mary left the room [a rich woman]

It appears that expletives require case, otherwise none of the sentences in (7.12-14) would passivize.
However, some of these examples do passivize, as shown in (7.25) ((7.25e) = Postal and Pullum’s
(58b).) In each of these examples, expletive it must be a C-head of an A-chain.
(7.25) (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.....

Furthermore, as both Postal and Pullum, and Stroik note, expletive it must be adjacent to the matrix verb, just like (other) complements, as shown:

(7.26) (a) I take (*sincerely) it that you pay.
(b) He never gave (*whole-heartedly) it a thought that ....
(c) *They hold, without a doubt, it against me that I ....

These data suggests that these instances of expletive it must be case licensees. Expletives are not theta licensees and therefore it is not clear why they require case. We propose that the fact that these expletives appear in complement position makes them require case. As we have proposed above, the Internal-Theta relation is actually a type of Complement relation. We propose that the claim that only theta-marked Ns require case is too strong, and that expletive complements are part of a larger set of case-seeking Ns. In contrast, there seems to be no argument available to Stroik for explaining why these expletives require case even though other Ns in non-argument positions do not.

In summary, since Stroik makes a series of ad hoc assumptions to explain a small fraction of the crucial data, his arguments against Postal and Pullum’s conclusion fail.

Preceding Postal and Pullum (1988), some linguists attempted to eliminate sentences like (7.12-14) from consideration. Postal and Pullum argue specifically against such claims. For example, Williams (1980, 222) argues for ignoring sentences like (7.27a) on the grounds that they occur too sporadically. Similarly, Chomsky (1981, 147-148) argues against the need for consideration of (7.27b) because it is idiosyncratic.

(7.27) (a) John saw to/regretted it that Bill had a good time.
(b) They forced it to rain (by seeding the clouds.)
Postal and Pullum argue that it is irrelevant whether or not a particular construction is sporadic or idiosyncratic—why should informal measures of statistical frequency have anything 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 reject these considerations as valid concerns regarding the constructions in (7.12-14). They propose that Chomsky’s and Williams’ claims in this regard are attempts to dismiss data without argument because these data are problematic for their analyses. We agree with Postal and Pullum.

Chomsky (1981, 147-148) also claims that (7.27b) should be dismissed from consideration because, in addition to being idiosyncratic, it is also *derivatively generated* in the sense of Chomsky (1965: 227, note 2; 1972: 27) Chomsky applies this term to expressions that are ungrammatical due to violations of subcategorization and/or selection restrictions, e.g., *John found sad, John elapsed that Bill would come*, and *Colorless green ideas sleep furiously*. Chomsky claims that the grammar does not generate these expressions directly, but "derivatively generates" these and any other strings that deviate from the grammar, but to which the grammar must still provide some description (presumably because people often assign even ungrammatical expressions some representation.) It is unclear to us as well as to Postal and Pullum in exactly what sense (7.27b) (or any example in (7.12-14)) is *derivatively generated*.

Chomsky (1981) claims that the fact that (7.27b) does not passivize, i.e., *it was forced to rain* (as per Postal 1974, 1977) is evidence that (7.27b) is derivatively generated. But it is unclear why passivization is a criterion for determining whether a construction is "derivatively generated," given the other examples of "derivative generation" above. Postal and Pullum show that the lack of passivization in (7.27b) and many of the sentences in (7.12-14) does not mean that such sentences should be removed from consideration. Similarly, the fact that the following sentences, taken from Postal and Pullum’s example (57), are ungrammatical does not mean that other sentences with main verbs *cost, have, resemble, promise* and *sell* are "derivatively generated" or are for any other reason less linguistically significant than sentences with other main verbs.
(7.28) (a) *A lot of money was cost by this.
   (b) *An accident was had by John.
   (c) *John is resembled by Bill.
   (d) *Bill was promised to leave.
   (e) *The books were sold without having finished them.

Furthermore, as Postal and Pullum (p. 656, note 9) point out, many of the examples in (7.12-14) do passivize, e.g., the examples in (7.25), repeated below:

(7.25) (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 conclude that:

The failure of some active structures to have passive alternates does not cast doubt on their grammatical status or support the notion of "derivative generation," and would not even if we were unable to find closely similar structures that do passivize. (Postal and Pullum 1988, p.657)

Many of the examples in (7.13) and (7.14) may involve idioms. According to Postal and Pullum:

Idiomaticity in some example of a direct object expletive could only be significant if it challenged the crucial claim that the structure has the form $[\text{VP} \text{V NP}]$. One possible proposal would be that the correct analysis of such cases involves a complex verb $[\text{V} \text{it...}]$, of which it is just a phonetic or morphological subpart. But, at least under the assumptions of Chomsky (1981), there is a syntactic argument against this. Consider the passives mentioned in footnote 9 [similar to our (7.25)]. Such examples raise extreme problems for any attempt to claim that direct object or object of preposition extraposition expletives are part of verbs. (Postal and Pullum 1988, p. 658)

They show that if it was really a subpart of the verb, then passives of the extraposed examples would contain two instances of it, as exemplified in (7.25’a,b,c,e). The ungrammaticality of these examples show that it is not part of the verb in these examples. We found that this argument can also be extended to the nonextraposition examples that passivize as shown in (7.25’c,d). ((7.25c) is an example of passive both with and without extraposition.)
(7.25) (a) *It was never mentioned it to the candidate that the job was poorly paid.
   (b) *It is (always) held it against me that I am an extraterrestrial.
   (c) *It was figured it out (that everyone in the office was an extraterrestrial.)
   (d) *It has to be fought it out.
   (e) *If it can be prevented it from raining,.....

Postal and Pullum provide one additional argument that * it in these constructions behave syntactically as Ns distinct from the verbs. Presentational * there requires intransitive verbs, but cannot occur with any of the expletive * it object verbs above, as exemplified in (7.29) (Postal and Pullum’s (65) - (67).) If * it was part of these verbs the compound verbs [V it] would be intransitive.

(7.29) (a) Near the fountain (there) was resting a lion.
   (b) *Near the fountain (there) was having a rest a lion.
   (c) In the back room (there) were pigging out two pale-faced and heavily-built teenagers.
   (d) *In the back room (there) were pigging it up two pale-faced and heavily-built teenagers.
   (e) Through the narrow street (there) rushed some morose-looking villagers.
   (f) *Through the narrow street (there) rushed it some morose-looking villagers.

According to Postal and Pullum, * it in the above examples contributes absolutely nothing to the meaning of these sentences. Any suggestion that these instances of * it have a reference and are not true expletives would have to give * it a different meaning for each verb, e.g., * it in * pig it up would refer to some sort of food; * it in (7.13b) would refer to some solvable problem or something that is hard to see; etc. Also, we observe that many of these constructions have synonymous alternatives without * it, e.g., (7.12c,d). Regarding the meaning of expletive * it in examples like (7.13-14), Postal and Pullum (p.650) cite Allerton:19

The impersonal kind of semantically empty elaborator can be illustrated with:

(173) (a) Oliver bought it. [buy it = "be deceived"]
   (b) Oliver gave it to me straight. [give it = "warn, reprimand"]

The * it in these and similar examples (get it into X’s head, have it out/off with X, etc.) is of course part of a verbal idiom, semantically speaking. But from the syntactic point of view, it seems to fill one particu-
lar elaboration function, that of the object, thus permitting the use of a divalent verb in place of a monovalent one, or a trivalent verb in place of a monovalent [Postal/Pullum: apparently an error for divalent] one. The it used in such patterns is unlike the “meteorological” it of zero-valent verbs: whereas the it of it rained (etc.) does not correspond to any valency function but merely fills an obligatory surface structure slot, the it of (173) is used rather to fit a particular class of verb into a structure that is alien to it but normal for other verbs; thus buy and give (also: get and have) are object-requiring verbs, but use of empty it allows them to occur without a substantive object, giving them the freedom of intransitive verbs. (Allerton 1982, pp. 131-132)

Lasnik and Saito (1991) generally support Postal and Pullum’s analysis of the expletive complement examples. However, they claim the following empirical problem exists, citing examples (7.30) and (7.31) (their nos. 78-86).

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)

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

(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

(7.31) Examples with verbs that take infinitival raising to object complements, but not expletive complements

(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, we find that some verbs with RO infinitival complements do allow expletive complements when negated, or with modals as shown in (7.32). Therefore, the empirical problem noted by
Lasnik and Saito doesn’t hold.

(7.32) (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 examples like (7.12-14) is one in which expletive *it* occupies a complement position and that this fact raises problems for (7.8). Once (7.8) is abandoned all obstacles to an RO analysis are removed. Obviously, exceptional case marking involves a "unique" process—instances in which an element case marks the specifier of its complement, rather than the complement itself. Since other "raising" type analyses exists, we believe that there is more independent support for an RO analysis than an ECM analysis, once it becomes clear that (7.8) is incorrect.

Assuming an RO analysis, Case is uniformly a type of S-structure constituent licensing relation according to the LRP, as discussed above. By showing that an RO analysis is more viable than an ECM analysis, this section has argued for the subsumption hierarchy of licensor/licensing relations diagramed in Figure 7.1 above.

### 7.3.3 Sentences with No S-structure

Dougherty and Leacock (1993) argue that sentences like the following (Dougherty and Leacock’s example 15a, 16, 18) can have no S-structure.

(7.33) (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.

We show that a CS analysis provides the means for assigning these sentences S-structures and we explain Dougherty and Leacock’s findings by showing that an EC analysis is incompatible with an S-structure representation. Therefore, these constructions provide further evidence for preferring a CS analysis over an EC analysis. We also show that CS analysis provides the means for comparing ECM
and RO analyses of these and other coordinated sentences. Under an RO analysis, the following constraint (much like Ross’ 1967 "Across the Board" Constraint) is obeyed by all coordinated verbs:

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

Under the ECM analysis, (7.34) holds for all sets of coordinated verbs, except those which include at least one ECM verb and one non-ECM verb. Thus (7.34) and the examples in (7.33) support an RO analysis and mitigate against an ECM analysis.

We follow Williams (1978) in modeling conjuncts as parallel structures which function as if they occupy the same syntactic position on two different dimensions—these parallel structures need to be "linearized" to derive the correct word order. In GBUG, this parallel-ness of structure is modeled as structure sharing, i.e., we assume that conjoined verbs share complements, as shown in Figure 7.6, an EC analysis of John persuaded and reminded the doctor to visit Bill (Dougherty and Leacock’s example 10a). Williams’ goal was to model Ross’ (1967) across the board constraint. Our goal is to analyze the Dougherty and Leacock data in such a way as to maximize the extent which coordinated verbs have the same complement structure.
Figure 7.6 poses no significant problems\textsuperscript{20} for our analysis of coordination because *persuade* and *remind* each have the same control/raising properties. However, the above type of representation for conjunction fails for EC based approaches for ECM and RO analyses of the verb *expect*, when *expect* is conjoined with *persuade*, as shown in Figures 7.7 and 7.8. The inconsistent parts of these figures are annotated with question marks and the initial of the applicable verb (*P* for *persuaded*; *E* for *expected*.) Under the ECM account (Figure 7.7), it is unclear what the subject of the infinitival complements of the conjoined verb [*persuade and expect*] would be. The specifier position of the infinitival complement would have to be filled simultaneously by: (1) the \( \bar{N} \) the doctor, as required by *expect*; and (2) PRO, as required by *persuade*. Also under the ECM approach, the \( \bar{N} \) the doctor would have to simultaneously be part of the matrix clause and independent of it. Under the RO analysis represented in
Figure 7.8, the subject of the infinitive would have to be simultaneously PRO and NP-trace. Dougherty and Leacock propose that these examples do have logical forms (LFs), even though they do not have S-structures (or D-structures for that matter.) They argue for a revision of linguistic theory in which sentences need not have a well-formed or complete syntactic representation at every level, contrary to the projection principle.\textsuperscript{21} However, we show that a CS analysis provides a means for representing the S-structures of these sentences.

| An (inconsistent) LRP-Reduced EC/ECM Analysis of |
| John persuaded and expected the doctor to examine Bill |
| Category: INFL |
| Specifier-Case: Category: Noun |
| Phonology: John |
| Antecedent: |
| Head-Proj: Category: i |
| Head-Proj: Category: l |
| Internal-Theta: Category: k, Verb |
| Conjunction: Phonology: and |
| Conjunct1: External-Theta: Category: NP-trace |
| Same-Reference: |
| Head-Proj: Category: x |
| Head-Proj: Category: y |
| Internal-Theta1: Category: m |
| Phonology: persuaded |
| Antecedent: n (P?) |
| Internal-Theta2: Category: p, INFL-Verb |
| External-Theta: (?E) m |
| (P?) Category: PRO |
| Same-Reference: |
| Head-Proj: Category: e |
| Head-Proj: Category: f |
| Phonology: to examine |
| Internal-Theta: Category: g |
| Phonology: Bill |
| Complement-Case: m |
| Conjunct2: External-Theta: l |
| Head-Proj: Category: k |
| Head-Proj: Category: l |
| Phonology: expected |
| Internal-Theta: Category: o |
| Complement-Case: |

Figure 7.7
Figure 7.9 is a CS version of an RO analysis of (7.33a). In a CS analysis, both RO and object control (OC) verbs license two complements: one N N and one infinitival phrase I. A RO verb $V_{\text{RO}}$, e.g., expect, licenses the complement relations Complement-Case($V_{\text{RO}}$ N) and Internal-Theta($V_{\text{RO}}$ I). An OC verb $V_{\text{OC}}$, e.g., persuade, licenses the complement relations MAC($V_{\text{OC}}$ N) and Internal-Theta($V_{\text{OC}}$ I). The difference between the two types of verbs boils down to the difference between the relations MAC and Complement-Case, i.e., only MAC is subsumed by Internal-Theta (cf. Figure 7.1). In contrast, EC analyses makes it seem as though the differences between these two verbs is the type of ec specifier found in the complement. Since each verb requires a different ec, there is no way to chose which ec is the specifier of the $I$ complement of the conjoined verbs. A CS analysis is preferred.
over an EC analysis because only a CS analysis provides a means for representing the S-structure for this type of sentence.

A CS version of an ECM analysis is also possible for (7.33a), as shown in Figure 7.10. (Apart from the structure sharing and elimination of ecs, this is very similar to the Dougherty and Leacock’s analysis. See note 21.) Figures 7.9 and 7.10 differ in that Figure 7.9 assumes that Complement \(\subseteq\) Complement-Case, but Figure 7.10 does not. Thus in Figure 7.10, Internal-Theta and Complement-Case arcs do not reduce to MAC arcs as they do in Figure 7.9. Furthermore, it is assumed that the doctor is not the complement of expect in Figure 7.10 in contrast with Figure 7.9.

| Category: | [i] INFL |
| Specifier-Case: | [i] Category: Noun |
| Phonology: | John |
| Head-Proj: Category: | [i] |
| Head-Proj: Category: | [i] Verb |
| Internal-Theta: Category: | [k] |
| Conjunction: Phonology: | and |
| Conjunct1: Category: | [k] |
| External-Theta: | [j] |
| Head-Proj: Category: | [k] |
| Head-Proj: Category: | [k] |
| Phonology: persuaded |
| MAC1: Category: | [i] Noun |
| Phonology: the doctor |
| Internal-Theta2: Category: | [m] INFL-verb |
| External-Theta: | [l] |
| Head-Proj: Category: | [m] |
| Head-Proj: Category: | [m] |
| Phonology: to examine |
| MAC: Category: | Noun |
| Phonology: Bill |

**Figure 7.9**
Constraint (7.34), repeated below, is a specific statement based on previous assumptions about the way conjuncts are parallel in some sense (see Ross 1967, Williams 1978, among others.)

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

The conjoined verbs in Figure 7.9 each have all the same complements, one to an N and one to an I. In contrast, the conjoined verbs in Figure 7.10 take a different number of complements, thus violating principle (7.34). As exemplified in (7.35), such violations are blocked in every other circumstance we are aware of. By adopting an RO analysis and rejecting an ECM analysis, we are taking the strong position that (7.34) always holds.
(7.35) (a) John will either telex or mail the documents to each of the interested parties.
(b) Mary simultaneously verbalized and signed the message to the audience.
(c) Clarence breaks and fixes the toys all the time.
(d) *John will either photocopy or mail the documents to each of the interested parties.
(e) *Mary simultaneously heard and signed the message to the audience.
(f) *Clarence breaks and laughs the toys all the time.

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

(7.36) (a) John could neither promise nor persuade himself to study harder.
(b) John promised and expected himself to study harder.

_Promise_ and _persuade_ each have one _N_ complement and one infinitival complement. Due to the different control properties of these two verbs, most sentences are ill-formed which contain the conjoined matrix verb _[promise and persuade]_ and its _N_ and infinitival complement. However, when an anaphor bound to _John_ is in object position, conjunction is allowed. As shown in Figures 7.11 and 7.12, these control properties cannot be represented properly in a DAG model, since the control properties of the two verbs conflict. Figure 7.11 captures that these two verbs have the same complement structures, but fails to represent the conflicting control properties in a straightforward manner. Figure 7.12 explicitly represents the control properties of the two verbs, but fails to clearly represent that the two verbs have the same complement structures. We assume that these difficulties reflect limitations of the DAG model for representing conjunction, rather than an argument that these sentences lack an S-structure as argued in Dougherty and Leacock (1993). In the ideal representation, the infinitival complement would be a complement of both verbs, even though the specifier of the complement would structure share with a different _N_ on each plane. Apart from this problem of representation, we assume that the complements of the conjoined verbs are the same. Similarly, under an RO analysis, both of the conjoined verbs in (7.36b) have one _N_ and one infinitival complement. The fact that _himself_ is in object position
also gives (7.36b) a grammatical reading. Assuming an RO analysis, neither of these examples violate (7.34). Principle (7.34) holds only if no ECM verbs exist. Therefore Principle (7.34) supports an RO analysis and mitigates against an ECM analysis.

An inconsistent CS Analysis of
John could neither promise nor persuade himself to study

| Category: [1] INFL |
| Specifier-Case: [1] Category: Noun |
| Phonology: John |
| Antecedent: [2] |

Head-Proj: Category [2]

Head-Proj: Category [1]

Phonology: could

Internal-Theta: Category [1] Verb

Conjunction: Phonology: neither · · · nor

Conjunct1: Category [3]


Head-Proj: Category [1]

Phonology: promise

MAC1: [2] Category: Noun

Phonology: himself

Same-Reference: [2]


Head-Proj: Category [1]

Phonology: to study

MAC: Category: Noun

Phonology: Bill

Conjunct2: Category [1]

External-Theta: [1]

Head-Proj: Category [1]

Head-Proj: Category [1]

Phonology: persuade

Complement-Case1: [3]

Internal-Theta2: [2]

Figure 7.11
S-structure representations of linguistic expressions are required for identifying the constituents which satisfy constituency tests and for constraints on word order, among other purposes. We assume that an adequate linguistic theory must be able to represent these concepts. On this basis, we have shown that only a CS analysis can explain how believe-type verbs can conjoin with other verbs taking one N and one to-infinitive complement, even if these other verbs are not believe-type verbs. Furthermore, we have proposed constraint (7.34) which accurately accounts for head coordination in English provided that an RO analysis, not an ECM analysis is adopted for believe-type verbs. (7.34) supports an RO analysis over an ECM analysis. We have not been successful at solving problems with representing all of Dougherty and Leacock’s (1993) examples, i.e., not the examples involving the
verb *promise*. However, we assume that this problem reflects a limitation of the DAG model, not a lack of an S-structure representation as argued by Dougherty and Leacock.

Thus the differences between our analyses and Dougherty and Leacock’s bear on the question: *Is a well-formed S-structure required of every grammatical sentence?* We attempt to answer this question below.

### 7.3.4 W-Verbs and Raising to Object

W-verbs (coined in Postal 1974) are a group of English verbs including *want, desire* and *wish* which are sometimes given RO or ECM analyses in generative frameworks. This section briefly discusses some of the issues raised in Bresnan (1972, 1976), Postal (1974), Bach (1977) and Lasnik and Saito (1991) which raise problems for an RO analysis of W-verbs. Bresnan (1972, 1976), Chomsky (1981) and Lasnik and Saito (1991), among others assume that the *N* following a W-verb is the subject of the infinitival clause, receiving case from a null case-assigning complementizer, usually derived by deletion of *for*, but Lasnik and Saito suggest that this null complementizer could also be base generated. Another possibility is that W-verbs are the only ECM verbs and that the differences between W-verbs and RO verbs may be derived from the differences between ECM and RO analyses. As noted above, the effect on our analysis would be that Complement-Case would not be a constituent feature. We argue that an RO analyses should be maintained for both W-verbs and *believe*-type verbs and the differences between these two sets of verbs do not bear on this issue.

Figure 7.13 charts some of the differences between B-verbs and W-verbs. We found two consistent differences: (1) Following Postal (1974), W-verbs also take $\overline{C}$ complements headed by *for*, but B-verbs do not, as shown in (7.37); and (2) Following Fitzpatrick and Sager (1974, 1981) and Wolff, Macleod and Meyers (1993), the verb following *to* in the infinitival complement of B-verbs is restricted to *be, have*, stative verbs like *love* and *know* and generic uses of other verbs (e.g., *speak* in (7.38d)), but there is no such restriction on the complements of W-verbs, as shown in (7.38). Postal (1974) and others note that W-verbs are inconsistent with respect to whether passive or complex NP shift is possible. In contrast, B-verb complements can always occur in these construc-
tions. The examples in (7.39) to (7.41) exemplify this distribution of W-verbs and B-verbs. Postal (1974, p. 190 footnote 2) proposes that expect is ambiguously a B-Verb and a W-verb. Postal (personal communication) extends this analysis to mean and intend. Assuming that W-verbs cannot occur in these environments, but B-verbs can, these data are explained. While it is tempting to explain the sentences containing the verb require in the same manner, it is not clear why passivization is grammatical with raising to object complements of require, but not complex NP shift.25,26

(7.37) (a) I want (very much) for my daughter to become a linguist
   (b) I would prefer for you not to do tie your shoes so tight
   (c) *I (would) consider (very much) for my daughter to be a linguist
   (d) *I (would) believe (very much) for them to be botanists

(7.38) (a) I expected you to leave/be a linguist/have been to China/know French (W-verb)
   (b) I would hate for you to leave/become a linguist/have too much pizza/know French (W-verb)
   (c) I consider/believe/assume you to be a linguist/have eaten the most pizza (B-verb with have/be)
   (d) I considered/believed/assumed our guide to know/speak French (B-verb with stative verb or in generic context)
   (e) *I consider/believe/assume you to leave/see London/eat the most pizza (B-verb violating these restrictions)

(7.39) B-verbs and Passive
   (a) They are considered to be botanists
   (b) They were believed to be a type of protozoan
   (c) The painting was claimed to be a forgery
   (d) The gun was assumed to be the murder weapon
(7.40) W-verbs and Passive

(a) You are expected to come (Postal 1974, p. 178)

(b) You were intended to die (Postal 1974, p. 178)

(c) *You are liked to visit us (Postal 1974, p. 178)

(d) *You are wanted to love me (Postal 1974, p. 179)

(e) You are required to get a license before you can drive.

(7.41) B-verbs and Complex NP Shift

(a) I believe to be arrogant—all those who disagree with my fountain theory of language. (Postal 1974, p. 405)

(b) I estimated to be over 175 feet long—all of the dinosaurs which we caught yesterday in Central Park (Postal 1974, p. 302)

(c) I assumed to be a forgery—the painting we found at the garage sale

(d) I reported to be annoying—the applicant for the job who was wearing two ties instead of one

(e) *I required to pass three difficult exams—the student who just entered our department

(7.42) W-verbs and Complex NP Shift

(a) I expect to be arrogant—all those talking dinosaurs who wear funny hats and cheap suits

(b) *I want to leave now—all of those men who are tracking mud on my carpet.

(c) *I wish to succeed in that venture—all of my friends who have just returned from the Venezuelan Highlands (Postal 1975, p.406)
A few of the differences between W-verbs and B-verbs, first noted in Bach’s (1977) review of Postal (1974), are discussed further in Lasnik and Saito (1991). Lasnik and Saito claim that quantifier scope and binding property distinctions made by Postal (1974) and Bach (1977) are particularly relevant because they may require the postverbal $N$ to be in the matrix clause at S-structure for *believe* and in the subordinate clause at S-structure for *want*. Thus if this reasoning is sound, W-verbs cannot have RO analyses and some other type of analysis would have to be offered, e.g., a null complementizer analysis or an ECM analysis. We find Lasnik and Saito’s arguments unconvincing because: (1)
each point involves very close grammaticality judgements among ill-formed sentences; (2) only sentences containing want and believe are discussed making it unclear to what extent their claims extend to other W-verbs and B-verbs; and (3) the sentences being compared are not sufficiently parallel to warrant drawing any conclusions.26

According to Lasnik and Saito (1991), the pronoun him can be coreferential with Bob in (7.43a) easier than it can in (7.43b) (their examples 62 and 63). This could be accounted for if (7.43a), but not (7.43b) was a violation of the constraint on coreference from Langacker (1969) which states that a pronoun cannot precede and command its antecedent (approximately the same as condition C of Chomsky’s (1981) version of Binding Theory). However, to our ear, the relative ill-formedness of these examples is due to the selection restrictions of the phrase even more fervently. It is possible to want something with passion, but believing is not something one normally does more or less passionately. In contrast, (7.44) compares two well-formed examples which presumably have the same structure as the examples in (7.44). With these well-formed examples, we detect no difference in relative acceptability of the coreference with him.

(7.43) (a) Joan wants him to be successful even more fervantly than Bob’s mother does
   (b) *Joan believes him to be successful even more fervantly than Bob’s mother does

(7.44) (a) Joan wanted him to be successful even before Bob’s mother did.
   (b) Joan believed him to be successful even before Bob’s mother did.

Lasnik and Saito (1991) imply that the relative well-formedness of examples (7.45a) and (7.45b) (their examples 64 and 65) may be based on whether the N those men is in the matrix or subordinate clause. This assumes that coreference between those men and each other is more felicitous in (7.45b) than in (7.45a) and that anaphors contained in adjuncts obey the same conditions as anaphors contained in complements, i.e., Condition A of the Binding Theory. In our view, the relative grammaticality is actually due to the fact that the subordinate clause of (7.45a) is a passive and the subordinate clause of (7.45b) is an active infinitival clause. In contrast we find (7.45c) just as good as (7.45b) and (7.45d), just as bad as (7.45a). Therefore these examples do not provide an argument against an RO
analysis of \textit{want}.

\begin{enumerate}[label=(\arabic*)]
\item \textit{*I wanted [those men to be fired] because of each other\textquotesingle s statements}
\item \textit{I believed [those men to be unreliable] because of each other\textquotesingle s statements}
\item \textit{I wanted [those men to be in my play] because of each other\textquotesingle s resumes}
\item \textit{*I believed [those men to be fired] because of each other\textquotesingle s statements}
\end{enumerate}

In (7.46), the relative positions of \textit{none} and \textit{any} are at issue. Lasnik and Saito posit that (7.46b) (their 67) is better than (7.46a) (their 66) because \textit{none} is in the matrix clause at S-structure in (7.46b), but not in (7.46a). On our analysis, (7.46a) and (7.46b) are not comparable for two reasons: (1) the subordinate clause is passive in (7.46a), and is active in (7.46b); and (2) We assume that the usage of \textit{after} in (7.46a) is infelicitous with the future orientation of W-verbs (cf. Postal 1977: p. 177). (7.45c) and (7.45d) are examples which are comparable to (7.45a) and (7.45b) without these drawbacks: both verbs in (7.45a) and (7.45c) have passive subordinate clause complements; and both verbs in (7.45b) and (7.45d) have active subordinate clause complements without any conflict in futurity. Not surprisingly, (7.46a) and (7.46c) are just as good/bad; and (7.46b) and (7.46d) are just as good/bad. Therefore these examples do not argue against an RO analysis of \textit{want}.

\begin{enumerate}[label=(\arabic*)]
\item \textit{*I wanted [none of the applicants to be hired] after reading any of the reports}
\item \textit{I believed [none of the applicants to be qualified] after reading any of the reports}
\item \textit{*I believed [none of the applicants to have been hired] after reading any of the reports}
\item \textit{I wanted [none of the applicants to be qualified] before reading any of the reports}
\end{enumerate}

Lasnik and Saito propose that the differences between (7.47a) (their 68) and (7.47b) (their 69) can be accounted for if we assume that \textit{each} must be c-commanded by its antecedent. Thus the difference between (7.47a) and (7.47b) could be accounted for if we assume that \textit{them} is in the matrix clause at S-structure in (7.47b), but in the embedded clause at S-structure in (7.47a). However, as with the examples above, (7.47a,b) are not parallel with respect to the passive/active status of the complement clause. Not surprisingly, one comparison of (7.47a) and (7.47c), and one comparison of (7.47b) and
(7.47d) shows that believe and want function the same way with respect to these constructions.

(7.47) (a) ??*I wanted [them to be fired] for three reasons each
(b) ??I believed [them to be incompetent] for three reasons each
(c) ??*I believed them to be fired for three reasons each
(d) ??I wanted them to be generous for three reasons each

The data above show that want and believe cannot be distinguished on the basis of the data which Lasnik and Saito (1991) deemed relevant to whether an RO analysis can be adopted for want. We showed that the Lasnik and Saito did not take into account the differences in meaning between want and believe. Desire is more emotional (can be done with greater fervor) than belief and desire also presupposes the futurity or irrealis of its complement, where belief does not. Another problem with Lasnik and Saito’s account is that they often compared examples with passive and active complements. Given the extremely fine grammaticality judgements on which Lasnik and Saito’s claims were based, these elements undermined the distinctions they were trying to make.

We conclude that none of the differences which we observe between B-verbs and W-verbs prevent an RO analysis of either.

7.3.5 Conclusion

In this section, we have compared ECM and RO analyses. We have found three criteria for preferring an RO analysis: (1) The LRP; (2) Evidence that expletives occur in object position; and (3) A constraint on coordinating heads of phrases. In previous GB accounts, the ECM analysis is preferred because the ECM analysis is compatible with the projection principle, but the RO analysis is not. However, we have presented evidence that the projection principle does not hold.

The projection principle and LRP are working hypotheses for constructing linguistic theories. To our knowledge, no evidence in the form of linguistic examples have ever been shown to actually prefer a theory with the projection principle over one without. Above we have found two types of linguistic examples which show that part of the projection principle is inaccurate. Simultaneously these examples lend support to the LRP.
7.4 The Minimalist Framework

7.4.1 Introduction

This section discusses the LRP, the Projection Principle and believe-type verbs in relation to some recent work in the Minimalist framework (cf. Chomsky 1992). In the Minimalist framework, LF is derived directly from the phonetic form (PF) and the lexicon, with no intermediate levels of representation: no S-structure and no D-structure. Chomsky (1992) proposes that given this framework without intermediate levels of representation, the projection principle loses its force:

The Projection Principle and the Theta Criterion have no independent significance at LF. But at D-structure, the two principles are needed to make the picture coherent; if the picture is abandoned they will lose their primary role. These principles are therefore dubious on conceptual grounds, though it remains to account for their empirical consequences, such as the constraint against substitution into a θ-position. If the empirical consequences can be explained in some other way and D-structure eliminated, then the Projection Principle and the Theta Criterion can be dispensed with. (Chomsky 1992, p.28)

GBUG, the subsumption hierarchy of licensor/licensee features, the LRP and various other ideas introduced above provide a model for Minimalist analyses. The main difference between GBUG based GB analyses and GBUB based Minimalist analyses relate to Case Theory. In the GBUG version of the Minimalist framework, Case is always a type of Specifier relation. Under this approach, Complement-Case and MAC features are no longer needed and complements of believe-type verbs are handled in a manner similar to the RO analysis, except raising is at LF, i.e., in a GBUG version of the Minimalist framework, case is an LF relation rather than a surface relation as assumed above. Given the brevity of Minimalist proposals so far, further research is required to determine which of the licensor/licensee features defined above as surface licensor/licensee features are actually LF licensor/licensee features in the Minimalist framework.

In the Minimalist approach, no levels of S-Structure or D-structure are assumed. Surface Structure positions are replaced by an operation called SPELL-OUT which interfaces LF with PF—it spells out the phonetic value of a lexical item at any point in a derivation of LF from the lexicon. Chomsky claims that a separate level of S-structure is not crucial to a syntactic description of a sentence. However, this elimination of S-structure is only apparent since items which are SPELLED-OUT are Constituent-Licensors/Licensess at the point in the derivation in which SPELL-OUT occurs. The syn-
tactic positions these spelled-out constituents occupy are not PF positions because constituent structure does not exist at PF. If these positions in Chomsky’s LF did not have special status preventing them from being deleted, then well-formed derivations would "crash" at PF. Thus the set of SPELLED-OUT positions are the set of S-structure positions. These S-structure positions are crucial to any syntactic analysis for at least two reasons: (1) the generalizations captured by constituency tests are generalizations about S-structure constituents; and (2) word order constraints order S-structure constituents. While it is possible to "annotate" PF to represent these factors, this solution does not really eliminate S-Structure. It just gives S-structure a different name.

As Chomsky notes, the elimination of D-structure is sufficient for eliminating the projection principle. Therefore previous arguments against RO analyses are no longer valid. We investigate a new argument proposed in Chomsky (1992) to rule out an RO analysis. We show that Chomsky overlooks what an RO analysis would actually look like in the Minimalist Framework and that this analysis cannot in fact be ruled out in the manner Chomsky proposes. Additionally, we discuss some observations made in Lasnik and Saito (1991) which show that "raising" must take place at S-structure rather than LF as assumed in the Minimalist approach. Assuming a well-formed S-structure is required for any well-formed analysis of an utterance, our previous arguments for RO analyses still hold.

### 7.4.2 A GBUG-based Approach to the Minimalist Framework

Based on Pollock (1989), the Minimalist theory assumes that projections of INFL are replaced by the following structure (excluding $\text{NEG}$):

\[
(7.48) \ [\overline{\text{AGR}}_{-S} \text{Tense} \overline{\text{AGR}}_{-O} \overline{\text{V}}]]
\]

where $\overline{\text{AGR}}_{-S}$ replaces $\overline{I}$ under previous approaches, $\overline{I}$ is the complement of $\text{AGR}-S^0$, $\overline{\text{AGR}}_{-O}$ is the complement of $\text{T}^0$ (Tense). $\text{AGR}-S$ represents subject agreement properties of verbal inflection. $\text{AGR}-O$ represents object agreement properties of verbal inflection. (cf. Chomsky 1991, 1992 and Chomsky and Lasnik 1991)

In the Minimalist theory, case is a type of agreement property assigned at LF, rather than S-structure as in GB theory. Chomsky (1991: 18, 1992: 10-11) and Chomsky and Lasnik (1991, 81-82)
argue that $N$ subjects are raised to the specifier position of AGR-S to receive case and $N$ objects are raised to the specifier position of AGR-O to receive case, i.e., they assume that case is always assigned by some verbal inflection node AGR-X to the specifier of the projection of AGR-X. This includes instances of exceptional case marking, which they interpret as raising the subject of a complement to the specifier of the ARG-O dominating the case assigning verb (Chomsky 1992, p.11.) If this approach is correct, case is uniformly assigned by $AGR-X$ to its specifier where $X = O \mid S$ and the LRP reduces case to a type of LF specifier features in all instances we are aware of.\textsuperscript{28} Thus a GBUG based approach to the Minimalist theory assumes the subsumption hierarchy in Figure 7.1' instead of Figure 7.1 and the previous distinction between Specifier-Case and Complement-Case translates as a distinction between the case licensors AGR-S and AGR-O.

Another major feature of the Minimalist approach is that it is assumed that the verb moves to each $X^0$ position of AGR-O, Tense and AGR-S amalgamating with abstract agreement features in these positions. We represent this in terms of our V-chain analysis from Chapter 3 (For further details on the motivation of V amalgamating with various INFL features, see Pollock 1989, Chomsky 1991, 1992, Chomsky and Lasnik 1991 and the sources cited therein.)

Figures 7.14 and 7.15 are our versions of Minimalist analyses of the simple sentence *Mary ate the sandwich*. The key features of Figure 7.14 are: (1) the chain of structure sharing arcs: Case and
A Subsumption Hierarchy for Licensor and Licensee Features in English

Figure 7.1

Head-Proj Internal-Theta Internal-Theta Head-Proj Internal-Theta External-Theta and (2) the chain of structure sharing arcs: Head-Proj Internal-Theta Internal-Theta Case and Head-Proj Internal-Theta Internal-Theta Head-Proj Internal-Theta Internal-Theta. Each chain consists of one LF feature, i.e. case, and one theta feature. Figure 7.15 incorporates our V-chain analysis from Chapter 3. Unlike the Minimalist literature, we are not assuming a derivational approach. Therefore the effects of all amalgamations of verb and inflectional features are represented in Figure 7.15. Since Figure 7.15 represents the amalgam and not the verb and inflectional features separately, the syntactic categories:
verb, AGR-O, (Past) Tense and AGR-S are not distinguished. These categories are combined together yielding a single category \textit{AGR-S-Past-AGR-O-Verb}. For purposes of exposition, we assume that representations with and without V-raising are equivalent because questions of V-raising do not bear on our analyses.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.14}
\caption{A GBUG/CS/Minimalist Analysis of \textit{Mary ate the sandwich}}
\end{figure}
7.4.3 Why S-structure Positions are Required in GB Analyses

Chomsky (1992) proposes that S-structure should be eliminated if it can be shown that there are no S-structure conditions:

The basic issue is whether there are S-structure conditions. If not, we can dispense with the concept of S-structure, allowing SPELL-OUT to apply freely in the manner indicated earlier. Plainly this would be the optimal conclusion. Chomsky (1992, 33-34)

This section presents evidence that S-structure cannot eliminated. In particular, we show that word order restrictions depend on S-structure information and that constituency tests capture generalizations which are lost if S-structure is not represented.

Chomsky argues that S-structure is not needed for the following phenomena: (1) determining whether wh words move overtly or at LF (in situ); (2) binding theory conditions; and (3) the Case filter. Chomsky (p. 34) regards (1) as the result of "parametric differences" which reduce to morphological properties of a language represented at PF. (2) reflects a variety of issues which have been discussed in the literature for some time—however, none of these issues bear on this dissertation and we
will ignore them here (see Chomsky 1992, 35-36 for details.) Regarding (3), Chomsky (p.41) notes the case filter is an INTERFACE condition, a well-formedness condition on the final level of representation assumed in ones theory. Thus whether one proposes that the final level of representation is LF or S-structure, it is this final level that the Case filter restricts.

However, the spell-out operation discussed in Chomsky (1992) is insufficient for replacing all functions of S-structure. In the Minimalist theory, spell-out defines a relation between LF and PF for each word in a string such that set of constituent positions which spell-out operates on are equivalent to S-structure constituent positions in GB theory. These positions cannot be viewed as PF positions because, constituent structure is irrelevant at PF. These S-structure positions cannot be viewed as intermediate positions in a derivation of LF that can be eliminated by a principle of economy because eliminating S-structure positions removes substantive information from the linguistic representation of a sentence. These positions have a special status.

One possibility to solve the problems raised by this section is to annotate PF with S-structure information, so that PF would be the level of representation at which some S-structure generalizations may be captured. However, this would be tantamount to keeping S-structure, annotating it with phonetic information and giving it a new name. Under this proposal S-structure wouldn’t be eliminated, just renamed.

PF consists purely of phonetic information, including intonation phrases, PF constituents of a sort. However, there is no relation between PF units and syntactic units other than the fact that they may represent different aspects of the same word string. For example, intonation phrases and syntactic phrases are distinct—thus (7.49a) and (7.49b) represent the same sentence, the former breaking it down into intonational phrases, the latter into syntactic phrases. It is unclear how the Minimalist approach really relates PF and LF. At best, spell-out correctly relates the word order at PF with the representation at LF. However, the constituent information at LF does not seems to relate to the constituent information at PF in any way. Therefore, while both LF and PF may be derived from the lexicon, there need not be any clear relation between the two, unless PF is annotated with S-structure information making PF in the Minimalist theory equivalent to S-structure in GB theory.
We assume that separate levels of S-structure and LF need not exist. There is nothing to prevent a single representation, provided it is clear which constituent licensor/licensee features are S-structure features and which ones are LF features. This model is implicit in our discussions about the Minimalist framework so far. A complete formulation of a GBUG based LF does not bear on the issues discussed here, but will be a topic of future research. See Chapter 8 for details.

Each of the constituency tests represented by (7.2) to (7.5), repeated below, shows that a verb and its complement form a separate surface constituent from the subject of the phrase. At PF, presumably no syntactic constituency is represented—instead a sentence is divided into “intonation phrases” which may be sensitive to syntactic distinctions, but are not isomorphic with syntactic constituents. For example, in (7.2a,b), the string Sam ate forms an intonation phrase at PF, but does not form a constituent at a LF or S-structure. Assuming an "across the board" analysis of the variety proposed in Williams (1978), the LF of (7.2b) is something like Figure 7.16, or alternatively Figure 7.17 assuming a Minimalist analysis. The S-structure constituent ate green eggs and ham must be represented in Figures 7.16 and 7.17 because this constituent is shared under both approaches. Therefore, in addition to spell-out some other operation relating to coordination must also recognize the special status of the surface constituent ate green eggs and ham.

(7.2) (a) Sam [ate green eggs and ham] and so did I ___
(b) Sam [ate green eggs and ham] and I did too ___
(c) [Eat green eggs and ham] is what Sam did ___

(7.3) (a) What did Sam do? [Eat green eggs and ham]

(7.4) (a) Sam [ate green eggs and ham], and I did it, too.

(7.5) (a) John [ate green eggs and ham] and [gave blood] on the same day.
(b) Do you want to [sleep in a hotel] or [drive all night long].
(c) Underdog will either [swallow his secret energy pill] or [be eaten by monsters.]

### A Williams-style Analysis of

*Sam [ate green eggs and ham] and so did I*

<table>
<thead>
<tr>
<th>Category: INF</th>
<th>Past</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conjunction:</td>
<td>Phonology: and so</td>
</tr>
<tr>
<td>Conjunct1:</td>
<td>Category: Noun</td>
</tr>
<tr>
<td></td>
<td>Phonology: Sam</td>
</tr>
<tr>
<td>Head-Proj:</td>
<td>Category: Verb</td>
</tr>
<tr>
<td></td>
<td>Tense: Past</td>
</tr>
<tr>
<td>Specifier-Case:</td>
<td>Category: Noun</td>
</tr>
<tr>
<td></td>
<td>Phonology: I</td>
</tr>
<tr>
<td>Head-Proj:</td>
<td>Category: Noun</td>
</tr>
<tr>
<td></td>
<td>Specifier-Case:</td>
</tr>
<tr>
<td></td>
<td>Phonology: did</td>
</tr>
<tr>
<td>Head-Proj:</td>
<td>Category: Noun</td>
</tr>
<tr>
<td></td>
<td>Internal-Theta:</td>
</tr>
<tr>
<td>MAC: green eggs and ham</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 7.16**
In Chapter 4, we developed a theory of word order in which lexical items and phrases are ordered based on surface constituent licensor/licensee features (and structure sharing predicate licensor/licensee features). Previous theories of word order also reflect surface constituent positions rather than LF positions. Thus in the example above our word order subsumption hierarchy would determine that *green* precedes *eggs and ham*; that *ate* precedes *green eggs and ham* and that *ate green eggs and ham* follows *Sam*. These word order rules order surface constituents, not LF constituents, and not PF intonation phrases. Surface constituent relations are crucial because neither the phonological nor LF properties of a lexical items and/or phrases determine word order. There is no theory of word order we are aware of which would predict that multisyllable words precede monosyllabic words, that verbs with initial consonants precede those with initial vowels, etc. Similarly, we are

Figure 7.17
unaware of any theory of word order based solely on quantifier scope. Unless the syntactic positions containing items which are Spelled-Out have a special status, word order cannot be predicted from syntax. Giving these positions a special status is equivalent to stating that they are S-structure positions.

In summary, Chomsky’s claim that S-structure can be eliminated is premature. S-structure positions are syntactic positions which follow the word order of PF. S-structure constraints have at least two functions which cannot be cast in terms of LF or PF terms: (1) they constrain word order (at PF); and (2) they represent constituency test type generalizations—in particular S-structure constituents are the constituents which must be copied or structure shared in LF representations of conjunction. Therefore we assume the GB position that case is a surface constituent relation. Any theoretical reason for raising subjects and objects of verbs to AGR-S and AGR-O must be motivated by some other constituent licensor/licensee feature. For example, a set of logical licensing relations which determine the scope of Ns could motivate LF raising to model the scope of the quantifiers contained in the N—see Chapter 8 for a preliminary discussion of this proposal.

7.4.4 Believe-type Verbs in the Minimalist Framework

Chomsky (1992, 32-33) proposes that the subjects of B-verb complements raise to AGR-O to receive case, as shown in (7.50).

(7.50) (a) Mary believed John to have left. (PF)

(b) \[
\begin{align*}
\text{\underline{AGR}}-S & \quad \text{Mary} \\
\text{\underline{AGR}}-S & \quad \text{believed} \\
\text{\underline{T}} & \quad \text{John} \\
\text{\underline{AGR}}-O & \quad \text{to have left} \\
\text{\underline{V}} & \quad \text{left}
\end{align*}
\] (LF)

Chomsky attempts to rule out the possibility that John raises to the complement of believe by requiring that movement occurs to a position external to the targeted phrase marker, where the targeted phrase marker is the nonterminal projection immediately dominating the trace of movement. This is clear from the following Minimalist definition of Move-α:

GT is a substitution operation. It targets K and substitutes K' for \( \emptyset \) in K. But \( \emptyset \) is not drawn from the lexicon; therefore it must have been inserted by GT itself. GT, then, targets K, adds \( \emptyset \), and substitutes K' for \( \emptyset \), forming K', which must satisfy X-bar Theory...Alongside the binary substitution operation

381
GT, which maps \((K, K')\) to \(K'\), we also have the singularly substitution operation \(\text{Move-}\alpha\), which maps \(K\) to \(K'\). Suppose this operation works just as GT does: it targets \(K\), adds \(\emptyset\), and substitutes \(\alpha\) for \(\emptyset\), where \(\alpha\) in this case is a phrase within the targeted phrase-marker \(K\) itself. We assume further that the operation leaves behind a trace \(t\) of \(\alpha\) and forms the chain \((\alpha, t)\).

Suppose we restrict substitution operations still further, requiring that \(\emptyset\) be external to the targeted phrase-marker \(K\). Thus GT and \(\text{Move-}\alpha\) extend \(K\) to \(K'\), which includes \(K\) as a proper part. For example, we can target \(K = V\), add \(\emptyset\) to form \([_p \emptyset V]\), and then either raise \(\alpha\) from within \(V\) to replace \(\emptyset\) or insert another phrase-marker \(K_1\) for \(\emptyset\). In either case, the result must satisfy X-Theory, which means that the element replacing \(\emptyset\) must be a maximal projection YP, the specifier of the new phrase-marker \(\overline{V} = \beta\). (Chomsky 1992, 31-32)

Chomsky attempts to draw two types of results from the "extension" requirement: (1) Relativized Minimality effects and (2) a prohibition against raising to object.

Relativized Minimality, (Rizzi 1990) is a portion of the empty category principle (ECP) requiring a trace \(t\) bound by some item \(X\) not to have a "closer \(\alpha\)-governor" where \(\alpha\) is a variable over the set \{head antecedent\}, depending on the type of movement. Rizzi (1990, 1-11) uses this constraint to rule out examples like the following (examples taken from Chomsky 1992, p.32):

(7.51) (a) *John \(t_1\) seems it is certain \(t_1\) to be here.

(b) *Fix \(t_1\) John can \(t_1\) the car.

(c) *How \(t_1\) did John wonder what Mary fixed \(t_1\)?

(7.51)(a) is ruled out because \(it\) is a closer potential antecedent governor than \(John\); (b) is ruled out because \(can\) is a closer potential head governor than \(fix\); (c) is ruled out because \(what\) is a closer potential antecedent governor than \(how\).

Chomsky proposes that the extension condition rules out the examples in (7.51) by constraining the instances of move-\(\alpha\) which would produce them. In each of these cases the "closer governor" occupies some specifier position which successive instances of move-\(\alpha\) would ultimately move \(John, fix\) or \(how\). These examples are ruled out because this position cannot be simultaneously occupied by the intermediate trace and the other lexical item.

The extension condition seems to define move-\(\alpha\) as movement to a specifier position, thus ruling out any movement to a complement position, i.e., specifically prohibiting RO analyses. However, the extension condition relies crucially on the assumption that the specifier position is defined configurationally as the sister to \(\overline{X}\), i.e., that in the configuration \([_\gamma \overline{Y} \overline{X}]\), where \(\gamma\) a projection of \(\overline{X}, \overline{Y}\) is
always the specifier of the phrase. As discussed in Chapter 3, this is untrue. \( \bar{Y} \) can be an adjunct, a complement or a specifier, as shown:

\[
(7.52) \begin{align*}
(a) & \quad \bar{V} \text{slowly} \bar{V} \text{roasted the eggplant]} \\
(b) & \quad \bar{V} \text{give John} \text{[some money]} \\
(c) & \quad \bar{V} \text{John} \text{[eat [the eggplant]]}
\end{align*}
\]

Chomsky (1992, p.9) specifically, makes the proviso that he is ignoring adjunction for the purposes of this claim. However, it is not clear whether he is ignoring adjunct modifiers, movement by adjunction, or both. For right now, we assume that the extension condition can account for both types of adjunct positions configurationally. However, this promissory note must be filled or the extension condition fails as a generalization. We now turn to syntactic configurations of multiple complements, a crucial topic for RO analyses. Chomsky uses a variant of (7.52b) based on Larson (1988). Chomsky (1992, 19) assumes the following initial binary branching structure for multiple complements of verbs like put:

\[
(7.53) \quad \bar{V}_1 \text{John} \bar{V}_1 \text{[} \bar{V}_1 \text{e} \bar{V}_2 \text{[} \bar{V}_2 \text{[} \bar{V}_2 \text{the book} \bar{V}_2 \text{put} \bar{V}_2 \text{[} \bar{V}_2 \text{on the shelf}]}
\]

In (7.53), the verb put assigns one internal theta role to the \( \bar{Z} \) and then raises to the position \( \bar{V}_1 \) to assign another internal theta role to \( \bar{N}_2 \). If raising to object was permitted, we would expect a similar structure, as shown in (7.54). Notice that this structure would not violate Chomsky’s extension requirement.

\[
(7.54) \quad \bar{V}_1 \text{Mary} \bar{V}_1 \text{[} \bar{V}_1 \text{e} \bar{V}_2 \text{[} \bar{V}_2 \text{[} \bar{V}_2 \text{e} \bar{V}_2 \text{believe} \bar{V}_2 \text{[} \bar{V}_2 \text{John to leave}]}
\]

This structure behaves just like (7.53), except the verb does not assign the \( \bar{N} \) (e or John) a theta role. There is nothing to prevent \( \text{John} \) from moving to the position occupied by \( e \), because it extends the domain \( \bar{V}_2 \) to a \( \bar{V} \) in the sense discussed above. Therefore Chomsky’s extension requirement does not rule out an RO analysis in a Minimalist approach.

Lasnik and Saito (1991), drawing on Postal (1974), present evidence that subject to object raising at S-structure is preferred over subject to AGR-O raising at LF with respect to accounting for
several binding and quantifier scope phenomena. Based on examples including (7.55) to (7.57) (their examples 13-14, 30-32 and 25-27), they show that the object of B-verbs is in some sense structurally higher than the subject of tensed clauses. According to Lasnik and Saito, the $\tilde{N}$ few students is higher in the phrase structure tree in (7.55b) than (7.55a) in order to explain why few can have either wide or narrow scope in (7.55a), but only wide scope in (7.55b). Similarly, the relative height of none and any are at issue in (7.56) and whether the anaphor each other is in the same clause as its antecedent is at issue in (7.57).

(7.55) (a) The FBI proved that few students were spies (wide/narrow)
   (b) The FBI proved few students to be spies (wide only)

(7.56) (a) The DA accused none of the defendants during any of the trials
       (b) ?The DA proved [none of the defendants to be guilty] during any of the trials
       (c) ?*The DA proved [that none of the defendants were guilty] during any of the trials

(7.57) (a) ?The DA proved [the defendants to be guilty] during each other’s trials
       (b) ?The DA accused the defendants during each other’s trials
       (c) ?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 B-verbs and the subjects of the tensed complements are S-structure differences rather than LF differences based on the interaction between wh extraction, quantifiers and binding theory. Following Chomsky (1981), they claim that (7.58) (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 (7.59) (their 47 and 48) shows that anaphors must be bound at S-structure and not just at LF. Lasnik and Saito also show that the contrast in (7.60) can only be due to S-structure facts since nothing rules out LF raising of the negative element Neg of unlikely which would allow (7.60b) 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.29
(7.58) (a) Which book that John\(_i\) read did he\(_i\) like (not at D-structure)

(b) *He\(_i\) liked every book that John\(_i\) read (not at LF)

(7.59) (a) John\(_i\) wonders which picture of himself\(_i\) Mary showed to Susan

(b) *John\(_i\) wonders who showed which picture of himself\(_i\) to Susan\(_i\)

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

(b) *Anyone is unlikely to address the rally

Contrary to Chomsky (1992), we have shown that an RO analysis is possible in the Minimalist framework. Lasnik and Saito (1991) provide evidence that an RO analysis is preferred to a raising to AGR-O analysis. This is unsurprising in light of the evidence presented above that GB/Minimalist approaches are descriptively and explanatorily inadequate without S-structure representations and our evidence in favor of an RO analysis of B-verbs.

7.4.5 Summary

We conclude that contrary to Chomsky (1992), the concept of S-structure cannot be dispensed with. Constituency tests, word order, and Raising to Object constructions all provide evidence that surface constituent relations are required for an explanatory linguistic theory. The Minimalist framework provides a more articulated version of LF than GB, attempting to incorporate S-structure properties into the process of deriving LF. However, the evidence above showed that surface constituent relations are inadequately represented in a Minimalist analysis.

7.5 Conclusion

We have shown that Constituent licensor/licensee features in GB theory represent a wide range of properties, many of which are parallel to other types of licensor/licensee features as evidenced by our licensor/licensee feature subsumption hierarchy. Surface constituent features are testable on the basis of constituency tests. Nonsurface features are interpreted as constituent features based on their relation to surface features, e.g., the feature Internal-Theta is a type of complement features because it is sometimes parallel to a surface complement feature and is always a surface complement feature for
non-Ns.

A large portion of this chapter was devoted to motivating a Raising to Object analysis for believe-type and arguing against an Exceptional Case Marking analysis because the latter contradicts our assumption that Complement-Case is a surface complement feature. We showed that analyses of two separate syntactic configurations, previously discussed in Postal and Pullum (1988) and Dougherty and Leacock (1993), motivate an RO analysis. Postal and Pullum (1988) show that the GB assumption that complements must be theta-marked is incorrect due to their analysis of expletive complement constructions. Thus contrary to the projection principle, an RO analysis is possible. Our interpretation of conjoined B-verb and object control complements from Dougherty and Leacock (1993) show that an RO analysis of B-verbs follows from the assumption that only verbs with the same complement structure can be conjoined—this constraint would have to be significantly weakened under an ECM analysis. On these bases, we conclude that an RO analysis is required in some cases. On the basis of economy considerations, an RO analysis is preferred over an ECM analysis.

Finally we discuss the Minimalist Framework in light of the following issue:

The basic issue is whether there are S-structure conditions. If not, we can dispense with the concept of S-structure, allowing SPELL-OUT to apply freely in the manner indicated earlier. Plainly this would be the optimal conclusion. Chomsky (1992, 33-34)

We show that the elimination of S-structure is not possible because the generalizations of constituency, theories of word order and various other linguistic phenomena require an adequate representation of surface constituent relations.
1 Minimal-A-chain (MAC) arcs are subsumed by both Internal-Theta and Complement-Case arcs. Therefore:

\[
\text{Internal-Theta } \cup \text{ Complement-Case } \rightarrow \text{MAC}
\]

Because Internal-Theta and Complement-Case arcs are unifiable, they cannot have the same source without violating functionality, as defined Chapter 2. Our unification algorithm (Appendix 2A) resolves potential conflicts so that same source Complement-Case and Internal-Theta arcs do not occur. The final version of Procedure 2.3 in Appendix 2A derives a MAC arc in all instances in which other unification algorithms might derive a pair of same source Internal-Theta and Complement-Case arcs. Below, when we assume a non-binary branching structure, we are careful to number multiple Internal-Theta licensees and Complement-Case licensees, to prevent a violation of functionality.

2 When INFL is a verbal inflectional affix (as opposed to infinitival to, an auxiliary verb or modal), it is not clear whether the INFL or the verbal Head-Proj feature is a surface feature because there are no surface constituents between these two heads. Based on our analysis of V-chains in Chapter 3, we would assume that the INFL head is a surface head. The fact that constituent structure tests show that the verb-affix amalgam form a constituent with any complements of the verb shows that either the \( I \), the \( V \) or the \( \bar{V} \) forms a constituent. Since all other instances of Head-Proj are surface relations, we put this matter aside, assuming that there is something wrong with analyses which divide a finite verb into INFL and a base verb at S-structure.

These facts are among a growing body of evidence that GB’s treatment of INFL as the head of a separate phrasal projection is incorrect. For example, most non-GB linguistic frameworks assume that: (1) inflectional morphology instantiates syntactic features on verbs; (2) modals and auxiliary verbs are either types of verbs or separate parts of speech; and (3) infinitival to is either a type of verb, a separate syntactic category or a complementizer.

3 As defined above, the quantifier relation here only refers to the relation between a quantifier and the variable it binds. We assume scope is represented separately. Therefore the fact that quantifier is a
surface feature is exemplified by the adjacent positions of every with respect to book, and which with respect to book in (i) and (ii).

(i) Some philosopher owns every book.
(ii) Which book is owned by every philosopher.

4 An alternative view of this constituency test is that did is coreferential with the finite I

5 In GBUG, a predicate-structure is a FS representing all predicate relations, including theta relations. According to Chomsky (1981, p. 335) and subsequent work, D-structure is a pure representation of theta relations. Thus the information represented in predicate-structure subsumes the lexical information previously represented in D-structure. These concepts differ in that predicate relations do not always mirror some constituent relation. For example, modifiers are predicate licensors, but constituent licensees.

6 For some constructions it has been proposed that prepositions assign case to their objects, but their objects receive theta roles from the head selecting the prepositional phrase. These examples include passive by-phrases—the external theta role of the active entry is suppressed in the passive entry (e.g., Jaeggli 1986) and then reassigned to the object of the by-phrase, transmitted through the preposition. Similarly, George (1980) proposes that some verbs assign theta roles to objects of prepositions, e.g., the object of the to-phrase in John gave the book to Mary. Also \( N \) complements of nouns and adjectives are often claimed to require the insertion of the semantically empty preposition of in order for the complement to get case. This idea originated with Rouveret and Vergnaud (1980, 130-132) with respect to the rule of a insertion in French.

In each of these cases, a head assigns a theta role to the complement of a preposition (except when verbs and prepositions are reanalyzed as complex verbs.) We assume that the head assigns the theta role to the \( P \) and that the head of the \( P \) assigns a theta role to its object. We assume that some other mechanism links these two theta roles together such that the preposition receives and then retransmits the theta role. This enables us to maintain our hypothesis that all theta features are constituent licensee features.
We assume that both V-chains and A-chains link surface arcs with nonsurface arcs (although see note 2), i.e., the C-head is always a surface arc. Determiner chains are also characterized as linking surface arcs (the C-head) and non-surface arcs (the C-tail) if we assume that the quantifier arc (a surface arc) includes the N-Agreement licensor feature.

Lasnik and Fiengo (1974) and Chomsky (1977, pp. 101-102), Leacock (1990, pp. 64-73) claim that the degree words too and enough license subordinate infinitival clauses following the adjective in (ia,b). They also claim that the infinitival clauses licensed by the degree words are complements of the adjectives. This violates Definition (7.2i) because an item other than the head of the phrase selects the complement of the head. Therefore, if these analyses are correct, it appears that RO analyses are not the only analyses which violate (7.2i).

The selection facts are demonstrated in (i). (ia,b) are grammatical sentences with degree words and infinitival clauses following the adjectives. (ic,d) are ungrammatical sentences with these same adjectives, but no degree word. Thus, the degree words license the adjectives. Similarly Leacock shows that comparatives (also specifiers of adjectives and adverbs) license both an infinitival clause and a than phrase, as exemplified in (ie).

(i) (a) John is too stubborn [(for me) to talk to (Mary/e)]
(b) John is old enough [(for me) to play with (Mary/e)]
(c) *John is stubborn [(for me) to talk to (Mary/e)]
(d) *John is old [(for me) to play with (Mary/e)]
(e) John is more stubborn [(for me) to talk to e] [than Michael (is)]

As shown in (ii), the same pattern may be observed for adverbs with the specifier too.

(ii) (a) John talked too quickly [(for me) to notice (Mary/e)]
(b) *John talked quickly [(for me) to notice (Mary/e)]

Baltin (1987) offers an alternative analysis in which the infinitival clause is a complement of the degree word, even though the clause is not adjacent. He claims that a degree word and its complement forms a constituent at D-structure, but the degree complement is adjoined to the matrix sentence at S-
structure. In a GBUG representation of this analysis the relations Degree(Degree-word, adj/adv), Theta(Degree-word, infinitival-phrase) and Adjunct(S, infinitival-phrase) hold in a FS. Baltin posits separate D-structure and S-structure representations specifically to avoid complications with c-command which would result from structure sharing (one of the possible analyses he investigates). However as noted above, the fact that we define command relations on arcs instead of nodes facilitates the proper use of command relations with non-trees.

Whatever analysis is ultimately adopted, the following facts will have to be taken into consideration: (1) the degree word selects both its degree licensee and the infinitive phrase; (2) the surface constituent structure (according to constituent structure tests) is as follows:

(iii) [[Degree adjective/adverb] infinitive-clause]

(3) the adjective or adverb is the head of the phrase; and (4) following Baltin (1987), the infinitival phrase functions as if it was adjoined to the sentence with respect to binding theory considerations.

In Chapter 3, note 36 we discuss tough adjectives in modifier position. Like degree words, these adjectives select the complement of the phrase even though they are not the heads of the phrase. In the following example, (repeated from Chapter 3) the adjective modifier difficult, selects the infinitival complement, but the adjective unfair does not. These examples (pointed out by Leslie Barrett, personal communication), thus require similar treatment to the so-called degree complements.

(iv) (a) Hildegard is a difficult person to talk to.
     (b) *Hildegard is an unfair person to talk to.

9In this section we sometimes use EC notation to informally represent CS analyses in sentence format. However, we use CS analyses explicitly in FS diagrams.

10As noted above, we assume that all complementizer-less infinitival complements are ųs. So for us, this distinction is moot.

11Postal and Pullum argue against the assumption that expletives only occur in non-θ-marked positions, although they claim that the validity of their arguments are unaffected by this issue. They argue that it (of the contraction it’s) in (ii) occupies the same position as the gerundal phrase in (i).
Therefore assuming this position is assigned a theta role in (i), it appears in a theta marked position in (ii).

(i) Complaining about it now is no use.
(ii) It’s no use complaining about it now.

12Postal and Pullum claim that the following are defining characteristics of expletives: (1) expletive pronouns cannot be antecedents of emphatic pronouns, like the italicized *itself* in (i); (2) expletives do not coordinate like other Ns, as exemplified in (ii); (3) expletives do not occur in nominalization of-phrases, as exemplified in (iii); and (4) expletives do not appear as subjects in tough movement constructions, as shown in (iv). Examples (i) to (iv) are taken from Postal and Pullum’s examples (1), (3), (6) and (7). See Postal and Pullum (1988) for details. (Note that these judgements belong to Postal and Pullum—in my idiolect (ivb) is grammatical.)

(i) (a) For him to smoke is *itself* illegal.
    (b) *It is *itself* illegal for him to smoke.
(ii) (a) He and it were respectively proved to be a person and claimed to be a robot.
    (b) *It and there were/was respectively proved to be raining and claimed to be floods in the valley.
(iii) (a) my estimate of it to be impossible (* to fly)
    (b) her resentment of it (*that he won)
(iv) (a) We prevented it from becoming obvious that things were out of control.
    (b) *It was tough to prevent t from becoming obvious that things were out of control.

13Postal and Pullum also reject two other analyses of the extraposition examples: (1) An analysis based on Rosenbaum (1967) in which the expletive is the head of a complex N and the that clause antecedent is its complement; and (2) Kayne’s (1985, 114-115) analysis of sentences like I’ve believed John for a long time now to be a liar. According to Kayne’s analysis, the phrase to be a liar moves rightward without its subject, but then creates a new subject position as shown:
(i) (Kayne’s 74) ...believed [John] [e] for a long time [[P [e] [to VP]]]

We direct the reader to Postal and Pullum (1988) for details.

14Stroik assumes that all $N$s require case. However, in fact only theta-marked $N$s which are not predicative do. For example, a scholar is a modifier in (i), not a complement—it does not receive a theta role. A genius is a predicate complement—under a small clause analysis, its specifier is (exceptionally) case marked, but not the predicate; under an RO analysis, it is also not case marked.

(i) John left school [a scholar]

(ii) John considered Mary [a genius]

The case filter only applies to non-predicative $N$s in theta positions or to expletives "linked" to $N$s requiring case.

15Williams assumes that reconstruction takes place at NP-structure (see Williams 1986 for details.) We assume that this occurs at LF (if at all.)

16Jespersen (1924, pp. 89, 248), Emonds (1976) and others propose that before, after, when and others are prepositions that take finite clause complements. Note however, that these clauses never are headed by overt complementizers, as shown in (i).

(i) (a) Bernard laughed over 14 million times before (*that) he died.

(b) When (*that) he died, Bernard stopped laughing.

(c) After (*that) the funeral ended, everyone talked about how much they missed that wonderful laugh of his.

17Paul Postal finds (7.23a) not so bad and (7.23b) grammatical. Thus assuming Postal’s dialect, Stroik’s small clause analysis would assign a structure to these sentences. However, Stroik’s analysis of these sentences would still suffer from many of the same deficiencies discussed above.

18On the analysis that copulas are raising verbs, e.g., Stowell (1978), the bracketed predicate $N$s in the following examples are also non-case-marked, even though the main verb presumably selects (theta marks) this phrase.
(i) George was/remained [a nice guy].
(ii) John considered Mary [a friend].

Allerton uses the term valency in the sense of Tesnière (1953, 1959). This basically corresponding to the number of arguments of a verb, e.g., a monovalent verb has one argument; a divalent verb has two, etc. This is not exactly the same thing as the number of theta roles however. For example, Allerton assumes that copulas (be, seem, etc in sentences like John is/seems (to be) ill) are divalent—he includes the subject as part of the valency count. In contrast, we assume that the copula assigns one theta role selecting a predicate (e.g., ill) and the predicate assigns a theta role to its subject.

By "relatively" unproblematic, we mean unproblematic relative to the examples in (7.33). Actually coordination is one of the processes which FS unification alone is insufficient for describing. For example see the section on agreement in Chapter 3.

Assuming EC and ECM analyses are correct, Dougherty and Leacock (1993) show that there can be no S-structure assigned to sentences like (7.33). They propose that not all sentences have "well-formed or complete" S-structures—sentences like (7.33) are assigned partial S-structures. For example, (7.33a) is assigned the following partial S-structure (Equivalent to Dougherty and Leacock, p. 17, note 3, example (iii) (b)):

(7.33a’) \[ [N \text{John}] [V \text{persuaded and expected}] [N \text{the doctor}] [V \text{to visit Bill}] \]

An LF is formed from this partial S-structure. The constituents are "copied" to form an LF which satisfies the complement requirements of each verb. In LF persuade takes two complements (an \( N \) and an \( I \)), made from one copy of the set of constituents: \( \text{\[N \text{the doctor}\ [V \text{to visit Bill}]\]} \), and expect would take one complement made from its copy of the same set of constituents.

While we only present CS analyses of the conjunction of promise and persuade, an EC analysis faces a similar problem. In an EC analysis, the Same-Reference feature of PRO has a different value for each verb.
Fitzpatrick and Sager (1974, 1981) distinguish between two types of verbs taking NP and infinitival complements. In the first type, the infinitive is unrestricted, but in the second type, either *have* or *be* must follow *to*. They do not explicitly distinguish RO/ECM complements from the various types of controlled complements. Wolff, Macleod and Meyers (1993) do make that distinction.

Catherine Macleod and myself have also observed the wider distribution while using the classes defined in Wolff, Macleod and Meyers (1993) in the Comlex Syntactic dictionary project. Paul Postal (personal communication) observed this wider distribution independently. In particular, Postal drew my attention to the use of generic senses of non-stative verbs in these complements.

Postal (1974, 406) cites Lasnik (personal communication) as the source for the idea that W-verbs do not undergo Complex NP shift.

It is interesting that the verbs which do not passivize with these complements can passivize with other complements for which RO analyses are assumed as shown:

(i) John was wanted on the ship
(ii) The gun was established as the murder weapon

The passive form of the verb *suppose* must be carefully distinguished from the homophonous adjective. The following examples show that the B-verb *suppose* allows both passive and complex NP shift constructions.

(i) Mary is supposed by many people to be a good linguist
(ii) Mary is often supposed to be a good linguist.
(iii) I often suppose to be a good linguist, anyone who happens to speak more than eight languages.

We repeat many of Lasnik and Saito’s examples below and assume their grammaticality judgements for those examples.

We assume that the Minimalist approach would assume the following structure for $\bar{P}$s:
because this would be consistent with Minimalist assumptions. This analysis is necessary if we are to
assume that all case features are specifier features in a GBUG based Minimalist account.

Lasnik and Saito’s raise one interesting problem with respect to adjunct modifiers. They pro-
provide evidence that objects of verbs act as if they c-command some modifiers of verbs in sentences like
(i) (their 53). One way to insure this would be to assume raising to AGR-O.

(i) There arrived two knights on each other’s horses
8.1 Conclusions

This dissertation has reached the following conclusions:

(1) Licensing relations are basic to all GB analyses of phrase structure, command relations, government, binding theory, word order and other topics.

(2) We adopt GBUG analyses containing structure sharing constituent licensee arcs in place of previous GB analyses containing empty categories including NP-trace, OC-PRO, (some instances of) WH-trace and others. The crucial factor which makes it possible to eliminate these empty categories is that each is bound to its antecedent by a same-reference relation. However, the empty category NOC-PRO, which is bound by an overlap-reference relation, cannot be eliminated by this same mechanism.

(3) Constituent Licensing Relations get their interpretation from other relations. For example, a pair of one constituent licensor (i.e., Head-Proj) feature \( \text{CON}_1 \) and one constituent licensee feature \( \text{CON}_2 \), which is parallel to a pair of case licensor/licensee features \( \text{CAS}_1 \) and \( \text{CAS}_2 \), represents a surface constituent relation due to this correspondence. On this basis, we show that many constituent licensor/licensee features should be eliminated from GBUG representations of GB analyses because they are completely predictable from other features.

The GBUG based model of GB analyses is absolutely essential to our account because none of the above conclusions are obvious in a constituent structure tree model. The latter does not represent predicate and agreement relations at all and represents constituent relations inadequately, as shown in Chapter 3. Therefore, constituent structure trees provide an unclear basis for GB analyses because the most crucial properties of GB analyses (licensing relations) can only be represented indirectly.
We have shown that GBUG based GB analyses are easy to compare with analyses in other unification-based linguistic frameworks. We have identified parts of non-GB analyses that capture the same or similar generalizations as parts of GB analyses, and we have identified generalizations captured by neither analysis. We have also successfully borrowed ideas compatible with GB theory from analyses cast in other unification based linguistic frameworks. In contrast, phrase structure tree based GB analyses are more difficult to compare with other linguistic frameworks. We examined aspects of Relational Grammar, Head Drive Phrase Structure Grammar, Categorial Grammar, Lexical Function Grammar and others in terms of GB theory, particularly in the areas of morphological agreement, coreference, wanna contraction and command relations. From the discussion of command relations (Chapter 3) came one observation which we believe is particularly significant and should be taken into account in future comparisons of linguistic theories:

(8.1) Structural Distinctions have Relational Hierarchy equivalents.

Assuming a binary branching analysis of verbs with multiple complements along the lines of Larson (1988) and Chomsky (1992), c-command usually gives the same precedence results as the relational hierarchy. For example, in (8.2) and (8.3), the subject c-commands all the objects of the verb and the direct object c-commands the indirect object and locative (oblique). Similarly, the relational/obliqueness hierarchy (repeated below):

(3.103) subject < direct object < indirect object < oblique

gives the same result. Following RG assumptions: the first object of a double object construction is the initial indirect object; the second object of a double object construction is the initial direct object; and locative phrases are considered initially oblique. For this reason, HPSG’s o-command yield’s the same result as GB’s c-command in these instances and in most of the relevant instances. Therefore defining constituent structure analyses which achieve the desired results with c-command is equivalent to defining primitive grammatical relations which achieve the desired results with o-command.

(8.2) \[ I_j \text{John} \rightarrow I_j \text{INF} \rightarrow \text{t} \rightarrow \text{V} \rightarrow \text{put} \rightarrow \text{V} \rightarrow \text{the baby} \rightarrow \text{V} \rightarrow \text{t} \rightarrow \text{V} \rightarrow \text{near a picture of himself/herself}]\]

(8.3) \[ I_j \text{Mary} \rightarrow I_j \text{INF} \rightarrow \text{t} \rightarrow \text{V} \rightarrow \text{give} \rightarrow \text{V} \rightarrow \text{t} \rightarrow \text{V} \rightarrow \text{John} \rightarrow \text{V} \rightarrow \text{t} \rightarrow \text{V} \rightarrow \text{the picture of himself/herself}]\]
In our view, many of the undesirable properties of ec analyses result directly from the use of the constituent structure tree model for analyses that are beyond its descriptive capacity. NP-traces, Wh-traces and OC-PRO and the mechanisms for binding these ecs enabled GB theory to maintain a tree-based model for describing syntactic phenomena which involve structure sharing in a DAG based model. However, the cost of maintaining a tree-based model include: (1) the loss of linguistic generalization; (2) the need for special binding mechanisms, many of which are inappropriate; and (3) a system of indices independent of the tree-based model. Given the standard GB interpretation that chains of ecs and their antecedents represent single semantic units, the indices are in many ways treated as if they represent structure sharing. Therefore, it appears that the tree-hood of GB’s constituent structure trees is only apparent.

Constituent structure tree based GB accounts model all types of licensing relations in terms of constituent relations. Since theta relations, case relations, etc. are not directly modeled in constituent structure trees, constituent structure "positions" are required for these non-constituent relations to apply, i.e., case and theta roles are assigned from constituent positions to constituent positions. In a GBUG-based model of these analyses, constituent licensing relations are parallel to all other types of licensing relations. GBUG-based accounts make it clear that parallel constituent licensor/licensee features (i.e., constituent positions) are redundant and therefore should be eliminated on grounds of economy.

In summary, this dissertation developed GBUG as a tool for linguistic analysis and explored fundamental issues of GB theory in terms of GBUG. The latter’s explicit representation of licensing relations enabled us to: (1) compare GB with other linguistic frameworks; and (2) correct inaccuracies, fill gaps and remove redundancies from GB analyses. In particular, it was shown that many constituent licensing relations are epiphenomenal of other licensing relations and that chains of same-reference ecs are epiphenomenal of constituents which are licensors and/or licensees of more than one licensing relation. The remainder of this chapter discusses avenues for future research.
8.2 A GBUG Representation of Quantifier Scope and WH Extraction

This dissertation covered Wh extraction only superficially and did not discuss quantifier scope at all. This section describes our research strategy for future investigations of these areas.

The nature of the level of Logical Form (LF) is an active area of research in GB theory. LF is the level at which the scope of quantification, negation and wh operators are determined and the level at which some binding theory constraints are satisfied. Research in the Minimalist Framework assumes that other aspects of linguistic analysis may also be represented at LF. This section briefly explains how GBUG can be expanded to represent quantifier scope. We also briefly discuss possible ways to model the addition of intermediate constituent relations (intermediate traces) by adjunction, a concern also relevant for constructing S-structure chains constraining wh extraction of adjuncts.

Our research strategy assumes the following:

1. **SCOPE** is a licensing relation represented by two features with the same source: (a) a SCOPE LICENSOR feature; and (b) a SCOPE LICENSEE feature. The value of the scope licensor feature is a quantifier, a lexical item or phrase representing negation or a wh operator. The value of the scope licensee feature is the phrase over which the scope licensor has scope. Different readings of sentences are represented with separate LFs with different values for the scope licensor feature—this represents that different readings of sentences assume different scopes.

2. The feature **Quantifier** is a Specifier Licensor feature, a Quantification Licensor feature and a Scope Licensor feature.

3. The feature **WH-Operator** is a constituent licensor (head) and a scope licensor.

4. **WH** is the value of the WH-Operator feature. This phrase can have internal structure, e.g., the phrases *which book, on which day, what* are all **WHs** which function as WH-operators.

5. **WHs** are always the value of (at least) two constituent licensor/licensee features: (a) they are the head of a $\bar{C}$ consisting of the **WH** and an $\bar{I}$ containing a gap; (b) the **WH** is the value of some constituent licensor/licensee feature within the $\bar{I}$, i.e., the **WH** fills the gap.
There exists either: (a) some process which adds pairs of constituent licensor and constituent licensee arcs to $\overline{X}$s in order to license gap filling operations, e.g. the adjunction operation used in TAG framework (Schabes, Abeille and Joshi 1988, Joshi and Schabes 1991); or equivalently (b) a process of feature percolation which marks $\overline{X}$s meeting certain requirements as intermediate positions for licensing gap filling operations, e.g., a mechanism based on Correa (1987, 1988, 1991, 1992a, 1992b) in which intermediate "landing sites" for long distance dependencies is modeled as a feature on $\overline{X}$s, rather than as an adjunction to $\overline{X}$s—this second solution is very similar to the slash category mechanism assumed in HPSG and GPSG.

In Figure 8.1, a simple example of a GBUG FS representing quantifier scope and wh extraction, the WH which book takes wide scope over the quantifier every as represented by the fact that the feature Scope has the value Root, representing the root of the FS and the feature Internal-Theta Specifier-Case Scope has the $\overline{I}$ as its value. One potential pitfall of this approach is that resulting FS is modeled as a cyclic graph rather than an acyclic graph. This can be corrected by a process of adjunction which would yield the FS in Figure 8.2. Our final formulation, the way in which FSs like Figure 8.2 can be derived, as well as our model of GB’s process for ruling out a reading in which every takes wide scope are all matters for future research.\footnote{...}
A Cyclic Graph/FS Representation of
Which book did everyone responsible read

WH-Operator: Category: WH-Complementizer-Noun
Quantifier: Category: Determiner
Phonology: which
Head-Proj: Category: PH
Phonology: book

Internal-Theta: Category: INFL
Specifier-Case: Category: Noun
Quantifier: Category: Determiner-Noun
Phonology: everyone
Head-Proj: Category: PH
Modifier: Category: Adjective
Phonology: responsible
Head-Proj: PH

Scope: PH
Head-Proj: Category: PH
Head-Proj: Category: PH
Phonology: did
Internal-Theta: Category: Verb
External-Theta: PH
Head-Proj: Category: PH
Head-Proj: Category: PH
Phonology: read
MAC: PH

Scope: Root

Figure 8.1
8.3 Verbal Inflection: Syntactic Category or Lexical Feature

GB Theory and the Minimalist Approach represent verbal finite inflection as syntactic categories independent of the verb. In contrast, most other linguistic theories represent verbal finite inflection as features of the verb or equivalently that finite verbs are a subtype of verb. An interesting topic for future research would be to compare two versions of GB theory formulated in GBUG:

(1) The Independent INFL approach: the version of GB theory assumed above.

(2) The INFL as a Verbal Feature approach: a version of GB theory in which finite inflection is viewed as a set of verbal features
Previous comparisons of (non-GB versions of) these two approaches have concentrated on treatments of modals, auxiliary verbs and infinitival to. The Independent INFL approach assumes that modals, auxiliary verbs and to are all of these are heads of INFL. This approach concentrates on differences between these categories and main verbs, e.g., they have only one form, modals and auxiliaries invert in yes-no questions, etc.

The INFL as a Verbal Feature approach assumes that modals, auxiliaries and to are all heads of some other sort, e.g., following R. Lakoff (1970), McCawley (1975), Pullum and Wilson (1977), Pullum (1982) and others, that they are all types of verbs. This approach concentrates on their similarities between these words and verbs, e.g., the types of possible complements, the raising/control status, their position in phrase structure and the ability to bear tense. They attribute the differences to the fact that different types of verbs have different properties. For example, they assume that modals only occur in finite form, just like the verb rumor only occurs in the passive.

Pollock (1989) carries the Independent INFL approach a step further, dividing INFL up into separate categories for, tense, negation and subject verb agreement. Pollock attempts to account for a variety of additional syntactic phenomena including the interaction of negation, quantification, do-support and other phenomena. Chomsky (1991, 1992) and Chomsky and Lasnik (1991) further distinguish AGR-S and AGR-O as heads of inflectional categories, assuming that case is an LF feature and Ns raise to the specifier positions of AGR-S and AGR-O to receive case.

A comparison of the Independent INFL approach and the INFL as a Verbal Feature approach must examine and compare how both approaches account for the same set of phenomena. Finding objective criteria for a comparison is also an issue. Key questions will be: (1) How do the two approaches account for the fact that inflected verbs are a single unit at S-structure/PF?; and (2) How do the two approaches differ in the way they handle the LF properties discussed in Pollock (1989)?

Chapter 3 of this dissertation discusses the fact that the properties of morphological agreement cannot always be described by the same mechanism as other syntactic phenomena. Data and analyses from Morgan (1972), Fitzpatrick and Sager (1974, 1981), Corbett (1983, 1988) and Ingria (1990) sug-
gests that agreement should be viewed as a co-occurrence restriction rather than a combining operation (unification). This may be viewed as evidence against viewing morphological agreement as a type of constituent relation. Therefore, these data may be crucial evidence for this comparison.\(^3\)

In summary, the Independent INFL and INFL as a Verbal Feature approaches differ in the way they explain the same syntactic phenomena and whether they emphasize the differences or similarities among modals, auxiliaries and infinitival \textit{to}. LF properties, morphological agreement properties, and the means in which affixes and the verb are combined, either in the syntax or the lexicon all bear on this issue.

\textbf{8.4 A Parsing Algorithm}

Formulating an explicit model of a linguistic theory makes it possible to compile a computational lexicon and grammar suitable for use in computational linguistics. For example, a GBUG based grammar and lexicon is well-suited for use with the bottom-up chart parsing algorithm presented in Johnson, Meyers and Moss (1993a,b) (JMM). This section briefly describes that algorithm.

The JMM parsing algorithm operates directly on sets of lexical entries represented as FSs and returns all the possible parses in a chart. The algorithm determines word order based on a partial order on all SURFACE arcs in a FS, where the definition of surface arcs are defined according to the linguistic framework being modeled, the GB definition (see Chapter 7) being assumed here. According to the view most compatible with our assumptions, the partial order\(^4\) is derived from the set of word order rules (see Chapter 4) applicable to each set of same source surface arcs.

The parsing algorithm initializes a chart as a set of state sets, where each STATE SET corresponds to a position \(P\) in a string of words and each state set initially consists of one state for each lexical entry of the word \(W\) which immediately precedes \(P\). For example, given the string of words in (8.4) the numbers standing for positions in the string, the parser initializes a chart in which State-set 1 consists of states representing all possible lexical entries for \textit{John}, State-set 2 consisted of all possible lexical entries for \textit{wants}, etc. Each state includes: (1) a FS representing some constituent (with partially ordered surface arcs); and (2) the initial and final string positions of that constituent.
(8.4) 0 John 1 wants 2 to 3 leave 4

The algorithm traverses the chart from left to right, adding new states to the chart by combining states together. Two states $S_1$ and $S_2$ can combine together if: (1) $S_1$ and $S_2$ adjacent, i.e. the left string position of one state equals the right string position of the other; and (2) word order considerations are satisfied. If $S_1$ equals $[FS_1, \text{Left-Position}_1, \text{Right-Position}_1]$ and $S_2$ equals $[FS_2, \text{Left-Position}_2, \text{Right-Position}_2]$, the resulting state is derived by: (1) unifying one FS with the value of a surface arc A in the other FS provided A satisfies word order constraints, i.e., $FS_1$ is unified with the value of some surface arc in $FS_2$ or vice versa; (2) the resulting state has the left-most and right-most string positions of the original two states, i.e., the lesser of Left-Position$_1$ and Left-Position$_2$, and the greater of Right-Position$_1$ and Right-Position$_2$. Each time two FSs are combined surface arc A in the resulting state is considered to be COMPLETE, where once an arc is complete, no other FS may be combined with its value–arcs anchored by words (e.g., in lexical entries) are also considered complete. A complete parse of the input string is any State $S = [FS, 0, k]$, in which: (1) 0 and k are the initial and final string positions of the input, e.g., $k=5$ for (8.4); and (2) all the surface arcs in FS are complete. See Johnson, Meyers and Moss (1993a,b) for further details.
Notes

1The relations Specifier-Case(I,everyone), Internal-Theta(did,\(\overline{V}\)) and External-Theta(\(\overline{V}\),everyone) hold in Figure 8.1. On our account the specifier-case relation licenses the S-structure position occupied by everyone, but the relative word order of did, everyone and read are determined by the two theta relations. Did precedes the \(\overline{V}\) internal-theta licensee which includes the external-theta licensee of the verb. The \(\overline{V}\) is preceded by its external-theta licensee.

2See Chomsky (1957), Jackendoff (1972) and Akmajian, Steele and Wasow (1977), among others for some arguments in favor of the Independent INFL approach.

3Johnson, Meyers and Moss (1993a,b) assumes a distinction between category features and relational features. Category features include tense, person, number, gender, morphological case, syntactic category, and verb-form (e.g., past participle, infinitive, finite, etc.). Relational features were complex features representing stratified relational levels, e.g., \([3,2,1]\) is a feature label representing that the value of the feature is an initial direct object, which advances to direct object and then is a final subject, e.g., Mary in Mary was given tea by John.

In GBUG, licensor/licensee features correspond to SFG’s relational features for the most part. However from an SFG point of view, GBUG/GB analyses may be seen as confusing category features with licensor/licensee features when we assume that INFL, AGR-S, AGR-O, Tense, etc. are constituents rather than (category) features.

4We assume a partial order in order to allow for some degree of free word order. A strict order, like a set of phrase structure rules, would require us to posit multiple rules, and/or possibly multiple lexical entries, to describe free word order.
References

Abney, Steven (1986). "Licensing and Parsing," unpublished manuscript, MIT.


Barlow, Michael and Charles Ferguson (eds.) (1988a). Agreement in Natural Language: Approaches,
Theories, Descriptions, Stanford: Center for the Study of Language and Information.


Bresnan, Joan (1972). Theory of Complementation in English Syntax, Ph.D. Dissertation, MIT.


Chakravarti, Prabhachandra (1933). The Linguistic Speculations of the Hindus, Calcutta: Publisher Unknown.

(1988).


Fiorella, Josef (1993). NYU Seminar Presentation on Unification and Morphological Agreement.


Johnson, David E. and Lawrence S. Moss (1994). "Stratified Feature Structures for Multistratal Rela-


Stowell, Tim (1978). "What was There Before There was There," in Farkas, Donka et. al. (eds.) Papers from the Fourteenth Regional Meeting of the Chicago Linguistics Society.


Proteus Project ms, New York University.


Ziff, Paul (1964). "On Understanding "Understanding Utterances" in Fodor and Katz (eds)."