Bare Syntax
Dedicated to my wife, Youngmi, whose patience and love have been my main source of comfort and motivation

With a special note of thanks to Juan Uriagereka, whose interest in my work and friendship mean a lot to me
## Contents

*Preface* ix  
*List of Abbreviations* xiii  

### Part I

1 Preliminary considerations 3  
1.1 Basic desiderata 3  
1.2 The framework 5  
1.3 The central why-question 12  
  1.3.1 Locality of selection 12  
  1.3.2 Locality of long-distance dependencies 17  

2 Outline of a General Theory of Locality 28  
2.1 Merge and Move 28  
2.2 Products of Merge and Products of Move 31  
2.3 Unifying Chains and Projections 34  
  2.3.1 Commensurability 34  
  2.3.2 Permutability 37  
    2.3.2.1 Projection by movement 37  
    2.3.2.2 ReproJECTION 39  
  2.3.3 Symmetry transformation 44  
  2.3.4 Additional similarities 47  
2.4 Chains, Projections, and Locality 49  
2.5 Summary 58  

### Part II

3 Unambiguous Merge 63  
3.1 The nature of syntax 63
vi  Contents

3.2 Minimal Interface Requirements 66
  3.2.1 PHON 66
  3.2.2 SEM 71
  3.2.3 The lexicon 74
3.3 On the form of Merge 79
  3.3.1 The symmetry of Merge problem, and the need for labeling 79
  3.3.2 Shaping Merge, or what labels do 84
  3.3.3 Identifying the head, or what labels are 91
3.4 Adjunction 98
3.5 More on projection 106
3.6 Conclusion 118

4 Cartographies and the locality of selection 121
  4.1 Core issues 121
  4.2 The basic pattern 123
  4.3 X-bar everywhere 129
  4.4 Extension by licensing 136
  4.5 Iterated patterns 137
  4.6 Capturing typological restrictions 145
  4.7 How cartographies emerge, and why 149
  4.8 Conclusion: The fractal nature of syntax 159

5 Islands and the locality of chains 163
  5.1 How to approach the issue 163
  5.2 From Last Resort to Bounding 165
  5.3 Checking and Movement 169
  5.4 Generalized C-trace effect 176
  5.5 Avoiding freezing 176
    5.5.1 Anti-agreement 179
    5.5.2 Complementizer-manipulations 185
    5.5.3 Additional remarks 190
  5.6 Subextraction, CED, and QED 193
  5.7 On the robustness of the CED-generalization 198
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.8 Island “repair”</td>
<td>205</td>
</tr>
<tr>
<td>5.8.1 Preliminary remarks</td>
<td>205</td>
</tr>
<tr>
<td>5.8.2 Resumption</td>
<td>207</td>
</tr>
<tr>
<td>5.8.3 Wh-in-situ</td>
<td>213</td>
</tr>
<tr>
<td>5.8.4 Ellipsis</td>
<td>215</td>
</tr>
<tr>
<td>5.8.5 Construal</td>
<td>223</td>
</tr>
<tr>
<td>5.8.6 Pied-piping</td>
<td>228</td>
</tr>
<tr>
<td>5.9 Final considerations</td>
<td>232</td>
</tr>
</tbody>
</table>

### Part III

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Epilogue</td>
<td>243</td>
</tr>
</tbody>
</table>

*References* 251

*Index* 287
This page intentionally left blank
Preface

Syntactic relations in natural language are subject to severe locality restrictions. Selectional requirements (verbs demanding certain kinds of objects, prepositions demanding a given case, etc.) must typically be satisfied under adjacency. Likewise, long-distance dependencies must be established in a step-by-step, local fashion, lest they give rise to well-documented island effects.

The present work seeks to extract a general theory of locality—a theory accounting for both locality of selection and locality of long-distance dependencies—from the bare minimum that must be assumed to properly characterize natural language syntax: a basic structure-building mechanism (the operation “Merge”), minimal lexical specifications (“features”), and basic interface requirements that syntax must meet to relate sound/sign and meaning (“bare output conditions”).

Although my proposal is couched in a specific theoretical idiom (that of the minimalist program), I want to stress right from the start that the issues addressed here arise in all theoretical contexts, and, as I discuss in Chapter 6, the hypothesis put forth here should be of relevance to many frameworks. This said, I think that minimalist considerations are essential to shaping the questions and answers pursued in the pages that follow.

The book is structured like a classical sonata, around three parts. Part I, comprising Chapters 1 and 2, could carry the title Allegro. It introduces background information, which sets the tempo, but also includes a fair amount of development of the main theme. Part II, Adagio, presents a detailed examination of the central hypothesis in this book. Part III, Rondo, is a reprise, and closing.

Readers familiar with foundational assumptions in minimalism (from Chomsky’s work or books like my 2006 Linguistic Minimalism) and with basic issues in locality may skip Chapter 1. This book presupposes a certain amount of background and training in syntax,
but I have tried to make it accessible to students with only limited exposure to the most advanced aspects of current linguistic theory by highlighting difficult passages, and providing references that could be consulted when the going gets tough. Students who have gone through standard introductory texts like Adger (2003), Carnie (2007), or Haegeman (1994) should be able follow the core of the argument without too much difficulty.

I have tried to confine discussion of somewhat arcane technical issues and comparison of alternatives to footnotes. I see this as a compromise between not detracting the reader not in command of all the subtleties of minimalism from the main line of argument, and fairness to the existing literature that the experts are familiar with. For these more challenging passages, I recommend readers keep a copy of Hornstein, Nunes, and Grohmann (2006) at hand.

I began working on bare syntax at the end of my first semester in graduate school, some ten years ago. At the time I managed to state the basic idea on a single page, but quickly had to stop for it was obvious to me that I had not yet assimilated enough material. It took me all these years to assimilate it, and articulate my thoughts to a certain degree of explicitness. As a result of this long gestation period, my intellectual debt has grown so large that I cannot hope to even mention all the people that helped me over the years within reasonable page limits. I have done my best to cite all the works I learned from in the reference list at the end of this volume. Here I merely wish to highlight a few names from my ledger, in recognition of those who played a special role in helping this hypothesis take shape.

For extremely valuable discussions of various issues addressed in subsequent chapters, I want to thank (in no particular order): Danny Fox, Klaus Abels, Misi Brody, Michal Starke, Bob Freidin, Koji Sugisaki, Masao Ochi, Hagit Borer, James Yoon, Abbas Benmamoun, Naoki Fukui, Andrea Moro, James Huang, Gennaro Chierchia, Brent Henderson, Omer Preminger, Tomo Fujii, Robert Chametzky, Michael Wagner, Joan Bresnan, Ivan Sag, Hans Broekhuis, Memo Cinque, Norbert Corver, Riny Huybregts, Henk van Riemsdijk, Daniel Seely, Marc Hauser, Jan Koster, Jan-Wouter Zwart, Norvin Richards, Shigeru Miyagawa, Jackie Nordstrom, Heidi Harley, Halldor Sigurðsson,
Christer Platzack, Wolfram Hinzen, Jon Sprouse, Javier Ormazabal, Bob Berwick, Massimo Piattelli-Palmarini, Jean-Roger Vergnaud, Adam Szczegielniak, and Kleanthes Grohmann.

I am indebted to my teachers Željko Bošković and Howard Lasnik for being so supportive, and for creating an environment in which ideas like those investigated here could be developed. Željko deserves special thanks for preventing me from rushing to finish the manuscript in 2000 and defend it as my PhD thesis. I now realize he was so right that these ideas needed time to mature, and I am forever indebted for his confidence in this work.

Luigi Rizzi, Richard Kayne, Paul Pietroski, Sam Epstein, Norbert Hornstein, Juan Uriagereka, and Noam Chomsky deserve special thanks for producing the works I have relied on so much, and for illuminating discussions at various stages. Norbert, Juan, and Noam read many previous versions of this work, and provided much appreciated comments and encouragement. Without them this work would not exist.

Over the years, I had the good fortune to be offered the opportunity to present versions of this material to audiences that never failed to provide invaluable comments. Let me hereby thank whoever attended talks I gave at (in chronological order) the Harvard Society of Fellows, the University of Illinois at Urbana-Champaign, the University of Maryland, the University of Arizona, the Harvard University Theory Group, the University of Venice, Princeton University, the Universitat Autonoma de Barcelona, Osaka University, Leiden University, Lund University, Bogazici University, the University of Cyprus, Newcastle University, Groningen University, and the Shahid Beheshti University.

I owe a special debt to Steven Franks and to Myriam Uribe-Etxebarria for organizing two intensive courses at Indiana University in 2004, and the University of the Basque Country in 2007, where I could try out previous versions of this work. Thanks also to Lisa Cheng and the Dutch Science Foundation (NWO) for allowing me to spend a sabbatical semester in ideal conditions at Leiden University in Fall 2006, where I could complete a version of this book that I was able to submit to Oxford University Press for review.
My editor John Davey deserves more than words of thanks for his interest in this work, his advice and constant support. I am also grateful to the three reviewers for the Press who provided extremely useful comments, and to Bridget Samuels, Angel Gallego, and Terje Lohndal for a close reading of the final draft. Thanks also to Chloe Plummer and Jess Smith for expert copy-editing help, and Hiroki Narita and Terje Lohndal for editorial assistance.

A superb group of students provided just the right mixture of skepticism and enthusiasm, and played a major role in shaping the final version of this book in a way that is inadequately reflected in the text. I want to thank them here: Bridget Samuels, Dennis Ott, Terje Lohndal, Beste Kamali, Suleyman Ulutas, Hiroki Narita, So-One Hwang, Balkiz Ozturk, Conor Quinn, Taka Kato, Masa Kuno, Hiro Kasai, Clemens Mayr, Jordi Fortuny, Aritz Irurtzun, and Angel Gallego. Clemens and especially Dennis asked tough questions that led me to rewrite the entire manuscript before sending it off to press. During this final rewriting process, Jordi’s and Angel’s dissertations proved extremely valuable.

In fine, I want to thank my wife Youngmi. For six years now she has been my companion, my soulmate, and my refuge. She has provided the sort of encouragement and love than no one can possibly deserve. Even if her faith in me, her patience, and her support must forever remain inadequately acknowledged, I hope that she knows how truly grateful I am to her for everything she is, and everything she does. This book is dedicated to her, with love.
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-I</td>
<td>Conceptual-Intentional system</td>
</tr>
<tr>
<td>CED</td>
<td>Condition on Extraction Domain</td>
</tr>
<tr>
<td>EPP</td>
<td>Extended Projection Principle</td>
</tr>
<tr>
<td>FL</td>
<td>Faculty of Language</td>
</tr>
<tr>
<td>FLB</td>
<td>Faculty of Language in the Broad sense</td>
</tr>
<tr>
<td>FLN</td>
<td>Faculty of Language in the Narrow sense</td>
</tr>
<tr>
<td>GB</td>
<td>Government-Binding</td>
</tr>
<tr>
<td>GG</td>
<td>Generative Grammar</td>
</tr>
<tr>
<td>HPSG</td>
<td>Head-driven Phrase Structure Grammar</td>
</tr>
<tr>
<td>iF</td>
<td>Interpretable Feature</td>
</tr>
<tr>
<td>LCA</td>
<td>Linear Correspondence Axiom</td>
</tr>
<tr>
<td>LF</td>
<td>Logical Form</td>
</tr>
<tr>
<td>PCC</td>
<td>Person-Case Constraint</td>
</tr>
<tr>
<td>PF</td>
<td>Phonetic Form</td>
</tr>
<tr>
<td>PHON</td>
<td>External system for sound/sign</td>
</tr>
<tr>
<td>PLCA</td>
<td>Probe-Label Correspondence Axiom</td>
</tr>
<tr>
<td>QED</td>
<td>Quick Edge Detection</td>
</tr>
<tr>
<td>S-M</td>
<td>Sensori-Motor system</td>
</tr>
<tr>
<td>SEM</td>
<td>External system for meaning/thought</td>
</tr>
<tr>
<td>SI</td>
<td>Strong Island</td>
</tr>
<tr>
<td>SIMPL</td>
<td>Simplify operation</td>
</tr>
<tr>
<td>uF</td>
<td>Uninterpretable Feature</td>
</tr>
<tr>
<td>UG</td>
<td>Universal Grammar</td>
</tr>
<tr>
<td>WI</td>
<td>Weak Island</td>
</tr>
</tbody>
</table>
This page intentionally left blank
Part I
This page intentionally left blank
Preliminary considerations

The aim of the present book is to develop an analysis that explains why syntactic relations have to be established in a local manner, and, more importantly, why the locality conditions on syntactic relations that previous research has revealed in great detail take the form that they do—that is, the aim is to explain not just why syntax is local but why local in this particular way. Before outlining the specific approach which I intend to pursue (Chapter 2), I would like to use this introductory chapter to make a few general remarks concerning the methodology that will guide the present inquiry, and define the basic empirical issues that any adequate theory of locality should attend to.

1.1 Basic desiderata

Minimally, inquiry into the nature of the human language faculty (henceforth, FL, short for Faculty of Language) should address the following questions:

(1) i. How are concepts lexicalized?
   ii. What (necessarily finite) means are responsible for the defining features of FL: discrete infinity and displacement?
   iii. How are syntactic objects, once formed, mapped onto mental components ultimately responsible for their externalization (sound/sign systems)?
   iv. How do syntactic objects, once formed, participate in thought and action?

The questions in (1) are at the heart of the four traditional domains of linguistic analysis: lexicon (i), syntax (ii), phonology (iii), and semantics (iv).
The answers one provides to these questions should rely on a view of FL that is rich and robust enough to counterbalance the well-attested “poverty of stimulus” faced by the child, and at the same time flexible enough to accommodate the surface diversity of languages all too apparent to any student of language. The answers provided should also be compatible with findings in neuroscience and evolutionary biology concerning brain anatomy and physiology, and the evolution of cognitive faculties.¹

Ideally, the ultimate characterization of FL should make perfect sense of the properties discovered. That is, it should render FL perfectly intelligible. The best research strategy to attain (or at least approximate) this (“Galilean”) ideal of explanation is to take FL to be an “optimal” system (in a sense to be made precise in Section 1.2). This is the essence of the *minimalist program for linguistic theory* (see Chomsky 1993, 1995, 2002, 2006a; Boeckx 2006a; Uriagereka 1998 for general discussion).

Since detailed scientific inquiry into this extremely complex phenomenon here referred to as FL is only at a very early stage of development, it should be obvious that the feasibility of a minimalist program in linguistics is presently nothing more than a hopeful gamble, though one that promises great rewards if successfully conducted.² At the very least, the pursuit of the minimalist program will indicate for which domain of FL we can seriously entertain the possibility of a principled explanation.

There is at present nothing like a leading candidate theory, nothing like a clearly delineated path to follow to see the minimalist hope materialize. This is what the term “program” emphasizes. In this essay I will explore an avenue that seems reasonable to me. Because

---

¹ Up to now, compatibility with neuroscience and evolutionary biology has been a rather weak constraint on theory construction in linguistics, given the little we know about the neural and genetic foundations of language. But renewed interest in the evolution of FL in the wake of Hauser, Chomsky, and Fitch (2002) and recent attempts to develop linguistically informed linking hypotheses in neuroscience (see Poeppel 2005; Poeppel and Embick 2005) show signs of new possibilities for a significant rapprochement of these disciplines with traditional linguistics (see Boeckx 2006a: 138–50; Hauser, Barner, and O’Donnell 2007; and Fitch 2005 for relevant discussion).

² I have discussed some of these potential rewards in Boeckx (2006a: ch. 4).
I will make crucial use of notions like “bare output conditions” and “bare phrase structure,” and because I want to deduce locality conditions from a bare minimum, I will refer to the present line of inquiry as an attempt to develop a bare theory of syntax, or bare syntax, for short.

The specific way in which I will approach the questions in (1) will be supported by a certain set of assumptions that must be made clear before we proceed to more empirical matters, and to which I now turn.

1.2 The framework

Sustained scientific inquiry into the nature of FL goes back to the 1950s. Generative Grammar (GG) emerged some fifty years ago at the confluence of (i) the revival of and renewed appreciation for insights from the “first cognitive revolution” that took place in the (extended) Cartesian era (seventeenth to eighteenth centuries), (ii) the solidification of the scientific study of behavioral instincts (ethology), (iii) progress in the domain of computation (nature of algorithms and characterization of notions like infinity, etc.), and (iv) dissatisfaction with the then dominant behaviorist paradigm in psychology, which took external behavior not as evidence for what goes on inside a person’s brain, but as the basic limit of inquiry. Chomsky’s early work contributed to all four strands of research that constitute the conceptual underpinnings of GG (see Chomsky 1955, 1957, 1959, 1965, 1966). Today, Cartesian/Rationalist, biological, computational, and mentalist considerations not only define GG but modern cognitive science as a whole. They constitute the foundation of the computational-representational theory of mind, which many rightfully take to be the best hypothesis we have when it comes to finding out how the mind works (i.e., what the brain does) (see, e.g., Pinker 1997). They certainly constitute the foundation of the best hypothesis concerning how language (more accurately, the language faculty) works. Taken together, they constitute a conceptual core that one can characterize as “non-negotiable.”
In its fifty years GG has made remarkable progress. Although it went through several “conceptual shifts” (here one ought to single out the elaboration of the Principles-and-Parameters model (Chomsky 1981, 1986a) and the exploration of the Minimalist Program (Chomsky 1995)), its historical development is remarkably linear, as I have argued at length elsewhere (see Boeckx 2006a). Specifically, the major conceptual shifts witnessed in its whole history can be seen as parts of a naturally developing organic whole that strictly adheres to the goals of research clearly laid out in Chomsky (1965: ch. 1). These goals consist in addressing the following questions:

i. What is Knowledge of Language?
ii. How is that Knowledge acquired?
iii. How is that Knowledge put to use?
iv. How is that Knowledge implemented in the brain?
v. How did that Knowledge emerge in the species?

One of the most fundamental contributions made by GG to cognitive science is its emphasis on the poverty of stimulus faced by children in their attempt to acquire language. The recognition that one knows much more than one could gather from the environment is a very old one, going back at least to Plato’s Meno. GG has amply demonstrated that in the context of language, the child (unconsciously) attains a degree of linguistic knowledge that cannot be explained unless one posits a substantial contribution to the learning task from a structure that is part of the learner’s biological endowment. That mental structure, which is the focus of attention of generative studies, is called Universal Grammar (UG). Within the GG tradition, UG has been studied in a modular fashion, as a mental organ, a coherent whole that can be studied in relative isolation of other domains of the mind/brain. Indeed studies of various pathological cases, and additional considerations pertaining to the time course of acquisition have shown this mental organ to be biologically isolable, as Lenneberg (1967) had already anticipated.

The exact content of UG is, like all scientific hypotheses, a matter of controversy and intense study. But over the years several “good
leads” have emerged that many (myself included) have taken to be on the right track. Sketching them here will allow me to highlight some methodological principles that have made possible the discovery of these leads and that will prove useful throughout this study.

Any approach to the nature of the human language faculty will have to wrestle with the fact that despite the great similarities holding across languages and the remarkable uniformity of language development in otherwise very diverse populations, there exists an extraordinary amount of linguistic variation on Earth. Clearly, some mechanism must be at work that enables the child to end up speaking English if raised in New York, French if raised in Paris, and both French and English if raised in Montreal. Figuring out what kind of mechanism would be able to resolve the tension between “universals” and “particulars” was one of the major stumbling blocks of early GG. It was only in the mid-1970s that this obstacle could be removed, when linguists proposed the Principles-and-Parameters (P&P) approach. P&P is the immediate result of two intersecting strands of research in GG at the time: on the one hand, the discovery of general constraints on linguistic computations and representations (Chomsky 1970, 1973); and, on the other hand, the detailed study of differences across languages (Kayne 1975; Rizzi 1978). The success of these research agendas led to the hypothesis that UG consists of general principles constraining all linguistic processes (linguistic laws, if you will) and well-defined parameters, understood as options that the child will have to choose from. The theory of parameters can be thought of as a questionnaire with predefined answers that the child will have to choose from on the basis of the specific linguistic input she is given. To use a metaphor due to Jim Higginbotham, parameters are a bit like the switches in a circuit which, depending on their settings, will direct the effects of principles. On this view, the task that the child faces is one of parameter setting.

It is generally agreed now that although the exact number, format, and organization of linguistic parameters, along with strategies of parameter settings, remain a matter of debate (see Baker 2001a, Yang 2002, and Newmeyer 2005 for valuable discussion; for my own take
on the nature of parameters, see Boeckx 2007b, in progress), there is little doubt that something like a P&P architecture must be right for UG.\(^3\) Without P&P, it is simply impossible to imagine how a child equipped with UG ends up speaking the language spoken in her environment.

The P&P model could not have been elaborated without two basic scientific methods: the Galilean method and the comparative method. Of the two, the comparative method is the less controversial one. As its name suggests, the comparative method, well known from biology, studies (in the case of linguistics) different linguistic variants (languages, dialects, idiolects, etc.) very carefully and attempts to delineate their differences in a way consistent with whatever else is known about the language faculty. As Kayne (2005: 277) notes, comparative linguistics necessarily involves work on more than one linguistic system (i.e., language/dialect/idiolect), but it is not simply that. On the one hand, it attempts to characterize and delineate the parameters that ultimately underlie cross-linguistic differences (what is sometimes referred to as the points of variation). On the other, it attempts to exploit those differences as a new source of evidence bearing on the characterization and delineation of the principles of UG. The comparative method has been used, with great success, at various levels in GG: comparison of (superficially) massively different languages (see Baker 1996), comparison of closely related languages (see the contributions in Cinque and Kayne 2005), comparison of adult and child varieties of a given language (“language acquisition”; see Guasti 2002), and even comparison of normal and pathological instantiations of the language faculty (“language disorders”; see Grodzinsky 2000; Terzi 2005). Recently, the comparative method has been extended (as it ought to be) to the comparison of communication systems across species (see Hauser

\(^3\) I urge the reader not to equate P&P with one of its specific implementations in the so-called Government-and-Binding (GB) era (culminating in Baker 2001a). GB-style parameters are but one way of articulating the logic of P&P. As I discuss in Boeckx (2007b, in progress), many linguists who (rightly) criticized GB-style parameters in recent years went on to reject the whole P&P—in effect throwing out the baby with the bathwater.
Over the years the comparative method led to the surprising conclusion that there is much less variation at the core of UG than previously thought. Specifically, variation appears to be confined to lexical properties, as Borer (1984a) originally conjectured. Syntactic principles are thoroughly universal. Put differently, there appears to be no compelling evidence for postulating parameters within the statements of the general principles that shape natural language syntax. Today, the working assumption (adopted in this study) is Chomsky’s (2001: 2) Uniformity Hypothesis.

(3) Uniformity Hypothesis

In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances.

Whereas the comparative method approaches UG from the point of view of the most particular (linguistic peculiarities, ultimately lexical properties), the Galilean method approaches UG from the other end, from the point of view of the most general. In recent years the Galilean method has gained in importance under the rubric of the minimalist program for linguistic theory (Chomsky 1993 et seq.), although it is important to stress that one finds repeated allusions to it in Chomsky’s earlier work (see Boeckx 2006a: ch. 3 for relevant citations; see also Freidin and Vergnaud 2001).

The basic premise of the Galilean method is the commitment to the idea that “nature is perfect and simple, and creates nothing in vain” (see, for example, Galilei 1632 [1962]: 397). At the heart of the Galilean method is the belief, held by all major proponents of modern science, from Kepler to Einstein, that nature is the realization of the simplest conceivable mathematical ideas, and the idea that a theory should be more highly valued if “from a logical standpoint, it is not the result of an arbitrary choice among theories which, among themselves, are of equal value and analogously constructed” (Einstein 1949: 23). In the context of linguistics, the Galilean outlook leads us to expect that language exhibits properties of
optimal design, properties that would not be specific to language, but would have more general applicability (extending to all systems showing signs of “good design”).

The Galilean style can be extremely useful in trying to answer why-questions, which arise at any given stage of inquiry: Why should the language faculty (or, for that matter, any other system) be the way it is? Why should we find the principles that we uncover in our daily investigations?

As physicist Weinberg (2001: 15) observes,

> In all branches of science we try to discover generalizations about nature, and having discovered them we always ask why they are true. […] Why is nature that way? When we answer this question the answer is always found partly in contingencies, […] but partly in other generalizations. And so there is a sense of direction in science, that some generalizations are “explained” by others.

Nature may be the way it is because of some accident that happened at some point in time. Call this the historical answer. Or nature may be the way it is, because that is the simplest way it could be. Call this the Galilean answer. The historical answer may be true in many cases, but it is never particularly deep. Things could well have been otherwise. Rewrite history, rewind the tape of past events, and you may end up with a completely different, yet equally plausible outcome. By contrast the Galilean answer, whenever applicable, implies that things could not have been otherwise. And for this reason the Galilean answer is much more appealing. Precisely for this reason the Galilean style of investigation enjoys a certain privilege in scientific investigation. It should be tried first, because if it succeeds, it is more satisfactory (explanatory).

Although the Galilean stance should be the default one to take, in practice, why-questions must often be delayed, until a certain amount of empirical ground has been covered. This is exactly what happened in linguistics. Once the P&P approach crystallized, and provided a principled way to segregate the universal from the particular, it became possible to address why-questions and embrace

---

4 Here too, principles of the comparative method may prove relevant, as Chomsky notes (2006b: 183).
the Galilean style fully. Minimalism is precisely that: an attempt to 
“ask not only what the properties of language are, but why they are 
that way” (Chomsky 2004: 105) and provide an answer that is 
satisfactory, an answer that makes sense of language.

The strongest minimalist thesis tries to provide a most satisfactory 
answer, one that would make perfect sense of language, by claiming 
that all the properties of FL can be reduced to general considerations 
of computational efficiency, or to properties that any system pairing 
sound/sign and meaning would have to meet to be usable at all 
(what Chomsky (2000: 94) calls “legibility conditions” or (1995: 221) 
“bare output conditions”).

The strong minimalist thesis is probably too strong, but in practice 
scientists often adopt the strongest possible thesis as their working 
thesis. The strongest hypothesis acts as a limiting case, to see 
more precisely where and when the hypothesis fails and how much of 
it may be true. One virtue of the strongest minimalist thesis is that it 
provides us with a certain method of inquiry, a set of guidelines for 
theory construction. To vindicate the Galilean intuition discussed 
above, one must “show that the [richly documented] complexity and 
variety [in language] are only apparent, and that the […] kinds of 
rules can be reduced to a simpler form” (Chomsky 2001: 13).

The strongest minimalist thesis tells us that it is not enough to 
look for some explanation of P (some property of grammar). It 
forces us to look for a particular explanation that reveals P’s contrib-
ution to the “perfection” (optimal organization) of the system. The 
strongest minimalist thesis is meant as a constructive conjecture, a 
challenge to the linguistic community. There is no ready-made defini-
tion of what optimal design is.\footnote{It may be worth stressing that 
the term “optimal” here should be understood in much the same 
way as Chomsky (1986a) intended the term “language” to be under-
stood when he introduced the term “I-language”: from a strictly internalist, inten-
sional, essentially structural/formal perspective. “Optimal” here does not refer 
to functional optimality (which would correspond to Chomsky’s notion of “E-language”), 
measured in terms of communicative fitness (Pinker 1994; Pinker and Bloom 1990; 
Pinker and Jackendoff 2005). There is massive evidence that language is not optimal 
in this “E”-sense. One only needs to think of garden-path sentences, island effects, 
parsing problems with center embedding, and pervasive ambiguities. For discussion of 

Preliminary considerations
efficiency and the legibility conditions are not fixed in advance, they are part of the process of discovery. They will be whatever makes the internal mechanisms of FL look non-arbitrary. It is a major goal of the present study to uncover some of these principles and conditions that determine the character of syntactic principles.

1.3 The central why-question

With this much background regarding language and the methods to be used when studying it as a biological organ, we can now turn to the central empirical issues of this study. Throughout I will be concerned with the nature of locality conditions in syntax. It has been said, with undeniable accuracy, that “locality conditions are at the heart of syntax”; that “given […] the accumulation of evidence […], it seems reasonable to conclude that among the things we have discovered in recent times is that this is in fact how natural language works” (McCloskey 2002: 184–5)—syntactic relations must be established locally.

1.3.1 Locality of selection

Very often, when locality conditions are brought up, it is in connection with long-distance dependencies: How far away from the verb it modifies can an object wh-phrase be pronounced? How many sentential nodes can “movement” cross?

But locality is, in fact, a much broader issue. In addition to constraining displacement, natural language syntax also imposes


6 Not always, though. For an important exception, and a lucid discussion of the broader range of issues falling under locality, see Sag (2005).
locality restrictions on selection. It is a fact that lexical items impose demands on the kind of items they combine with. Ditransitive verbs require two objects, French *vouloir* ‘want’ requires a sentential complement inflected for subjunctive mood, German preposition *mit* ‘with’ requires a complement with dative case, and so on. Typically, such selectional demands must be met very locally. The precise notion of locality needed in this domain has been subject to discussion ever since Chomsky (1965: ch. 2) first addressed the issue, but it is generally agreed upon that notions embodying structural adjacency such as sisterhood (4a) or immediate containment/motherhood (4b) hold in most cases, and, because such notions are so natural and primitive, they are the ideal to strive for.

(4) a. b.

An important development in the context of locality for selection was Chomsky’s 1970 X-bar schema, which, via the endocentricity requirement (the idea that nodes at the phrasal level are projections of a head they contain), made it possible for information contained on a head (say, N) to travel up to the maximal projection level (NP) of the phrase, bringing it as close as possible to the lexical element selecting that phrase (say, P). Hornstein, Nunes, and Grohmann (2006: 178) are right in calling this key feature of X-bar theory (endocentric projection) the periscope property. It’s that which makes available to the “surface” information that would otherwise be “buried” (too deeply embedded), thereby allowing us to maintain the most natural, and strictest version of locality for selection.

---

7 By “natural” and “primitive” I mean that such notions are unavoidable (“come (virtually) for free”) as soon as one attempts to represent the sort of hierarchy at the heart of natural language syntax.
Locality of selection, couched in X-bar terms, is not entirely unproblematic, however. It in fact suffers from its own success. As is well known, a decade after X-bar theory was proposed, the X-bar schema was generalized (with great success) from traditional lexical categories (N, V, A) to functional units like Determiner, Tense, Complementizer, and so on. Soon NPs were renamed DPs; VPs, TPs (or CPs), etc. Sisterhood relations changed accordingly, not always for the better: Once P takes a DP complement, how can we still express the selectional requirement that the complement of P be “nominal” ([+N])?

Grimshaw (1991) (see also van Riemsdijk 1998) proposed a solution to this locality problem in terms of extended projection. She claimed that in addition to their own features, functional elements came equipped with a categorical feature matching that of the lexical projection they embed. Thus, D is (among other things) specified as [+N], since it takes an NP complement. Since D bears the feature [+N], so does DP (endocentricity) (6).

With this extended sense of endocentricity, one can still express the selectional requirement of P without abandoning sisterhood. But

---

8 For introductory coverage of this important development, see Adger (2003) and especially Haegeman (2005).
the problem for locality of selection does not end here. Recent developments in the domain of phrase structure (so-called cartographic approaches; see Cinque 1999, 2002, 2006; Rizzi 1997, 2004b; Belletti 2004a; Manzini and Savoia 2005, 2007) have given rise to a proliferation of ever-more fine-grained functional categories that stretch the notion of extended projection to the limit.

Perhaps the best-known proposal in this domain is Rizzi’s (1997) proposal that the well-established category of Complementizer be rethought as shorthand for a cluster of functional elements specifying the force, finiteness, and discourse contribution (focus, topic) of the clause, as schematized in (7).

\[(7) \quad \text{“CP”} = \text{[ForceP [TopicP [FocusP [FinitenessP]]]]}\]

At a suitable level of abstraction, the shift from CP to an extended left periphery of the clause of the sort proposed by Rizzi should not matter too much for issues of locality, especially if we make use of Grimshaw’s idea of extended projection. Via the latter, information will be passed up from projection to projection in a strictly local fashion all the way up to the topmost projection, say, ForceP, which will thus contain all the relevant selectional information to meet the demands of, say, the embedding verb.

\[(8) \quad [VP \text{ [ForceP [XP … [YP … [ZP]]]]}]\]

But once we get more specific about the nature of the phrases involved, percolation of information along extended projections appears to be problematic. Take, for example, the fact that verbs like ask or wonder select for embedded questions. Traditionally, this is captured by the demand that the CP-complement of such Vs be specified for [+wh]/[+Q]. This is achieved if C° hosts an element like *if* (9a), or if C° enters into an “agreement” (feature-sharing) relation with a [+wh]-element in SpecCP (9b).

---

9 The usefulness of Grimshaw’s notion of extended projection in this domain is also recognized in Shlonsky 2006. For more radical ways of dealing with the locality problem at issue, see Kayne 2000, 2005; Sportiche 2005; Koopman 2005.
But consider now the fact that some languages, like Hebrew, allow a topicalized element at the edge of an embedded question, to the left of the wh-word specifying the whole clause as [+wh] for purposes of selection. (Example from Shlonsky 2006: 2.)

(10) Sa’alta oti et ha sefer le mi le haxzir [Hebrew]
    asked.2sg me ACC the book to whom to return
    ‘You asked me to whom to return the book’

Before Rizzi (1997), one could have said that the topicalized material was adjoined to CP, essentially transparent for the purposes of projection schematized in (9b). Adjunction essentially provided a harmless way to incorporate the topicalized element into (9b).

As soon as Rizzi’s decomposition of CP into distinct functional projections is adopted, it is unclear how the relevant information can be transferred. More specifically, it is unclear what it would mean to allow for [+wh] information to be “passed onto” Topic°, given that [wh] marks new information, and [Topic] old information. Such a semantic clash of feature composition would be expected to bring the percolation of the relevant feature to an end.
(Here I follow Rizzi (1997) and take FocusP to host [+wh] material. Nothing crucial hinges on this choice. The only thing that matters is that the source of [+wh]-information must reside below TopicP.)

\[(12) \quad \text{[VP [ForceP [ TopicP [FocusP [FinitenessP]]]]]}\]

I know of no straightforward solution to this problem, which is not to say that detailed cartographic representations should be abandoned in favor of a return to less elaborate syntactic objects. My intention here was to convince the reader that the issue of locality of selection is less trivial than it might seem at first, and that the nature of extended projections and cartographies will have to be examined closely if one is to maintain that locality of selection reduced to primitive relations like sisterhood and immediate containment. I undertake this task in Chapter 4.

1.3.2 Locality of long-distance dependencies

In addition to locality of selection, any attempt to provide a theory of locality must deal with the fact that long-distance dependencies are subject to severe restrictions. Many of these restrictions are known as “islands”—the discovery of which is arguably the most significant empirical contribution of generative grammar to our understanding of how language works.

Chomsky was the first to observe (in Chomsky 1964) a condition on the application of syntactic transformation that prohibited movement of an element of type A to a position B if the element of type A was dominated by another element of type A. This condition, schematized (in modern form) in (13a), came to be known as the A-over-A principle.

The A-over-A principle straightforwardly accounts for contrasts like (13b–c).

\[(13) \quad \begin{align*}
    a. \ & *[A W]_i \ldots [\ldots [A Z [A t_i]]] \\
    b. \ & [PP From [PP under which bed]] \_i \text{ did John retrieve the book } t_i \\
    c. \ & *[PP Under which bed]_i \text{ did John retrieve the book } [PP from [PP } t_i]]
\]
Ross went far beyond Chomsky’s limited examples motivating the A-over-A condition, and provided us with a detailed map of the major syntactic islands.

In his groundbreaking (1967) PhD thesis, Ross systematically investigated the fact that seemingly minute manipulations of the context of a transformation dramatically affected the acceptability of sentences, as in (14a,b).

(14) a. Handsome though I believe that Dick is __, I’m still going to marry Herman.
   b. *Handsome though I believe the claim that Dick is __, I’m still going to marry Herman.

Of particular interest is Ross’s observation that sheer “distance” (measured in terms of node-crossing) does not seem to underlie islandhood. Thus, the dependency in (15) is longer than the one in (14b), but yields a fully acceptable result.

(15) Handsome though everyone expects me to try to force Bill to make Mom agree that Dick is __, I’m still going to marry Herman.

Although it is perhaps unfair to characterize Ross’s brilliant work as just a list of islands, for the thesis is peppered with insights that have played an important role in subsequent theorizing, it is fair to say that the beginning of our understanding of islands goes back to Chomsky’s (1973) epoch-making paper on “Conditions on transformations.” Chomsky (1973) set out to investigate what several of the domains Ross had identified as islands have in common. Thus began the modern study of locality, which, in many ways, defined the nature of current linguistic theorizing.

Since Chomsky (1973), most conditions introduced and discussed by generative linguists can be characterized as locality principles, that is, principles that limit the space within which transformational rules can apply. Most of these conditions, it should be said, have solidified insights that Chomsky formulated in 1973.

I do not intend to review this vast and very rich literature here, as such a review would require a book-length treatment in itself to do justice to all the subtleties uncovered since 1973. Instead, I will limit
myself to a few remarks pertaining to what McCloskey (2002: 185) calls a “much harder question [. . . ]—what property of language-design determines that this is how things work.” Call this the minimalist \textit{why}-question.

By calling it the minimalist \textit{why}-question, I do not mean to indicate that this is a question that came up only recently. It is very natural to ask why transformations should be constrained in the way they are, and in fact, the question of the naturalness of islands is already raised in Chomsky (1964: 40, 45), and discussed briefly in Chomsky (1973: sec. 11). There Chomsky notes that some of the conditions he discusses such as the A-over-A have the effect of “reducing ambiguity,” or of “increasing the reliability of a reasonable perceptual strategy that seeks the nearest” appropriate element, or of “guaranteeing a correspondence between deep structure position and scope.”

Today one can identify three broad avenues of research into the question of why islands exist. The first, by far the most popular, is firmly rooted in Chomsky’s (1973) Subjacency account, and takes islands to reflect computational limitations. Chomsky’s key intuition here is that islands arise when too many nodes of a specific kind have been crossed. The idea here is that movement is upward-bound. It is forced to apply within certain domains, delimited by so-called bounding nodes. Crossing one bounding node at a time is licit, crossing more than one such node at any given stage of the derivation results in deviance.

What counts as a bounding node has changed over the years (though not dramatically so); bounding nodes have been called “barriers” (Chomsky 1986b), and, more recently, “phases” (Chomsky 2000, 2001, forthcoming), but the intuition has remained the same. The subjacency approach, in any of its guises, emphasizes the idea that well-behaved, “efficient” computations should apply within limited domains; islands characterize these domains. Transgressing such computational limits leads to island violations. Put differently, this conception sees islands as the byproducts of principles that guarantee the \textit{computational efficiency} of grammatical operations. Chomsky himself has consistently favored this view,
from the subjacency account (see Chomsky 1973, 1977) through the Barriers framework (Chomsky 1986b), and, more recently, the phase framework (Chomsky 2000, 2001, forthcoming).

The second conception of the etiology of islands is based on the idea that islands identify conditions on the output(s) of the computational system. Whereas the first approach views islands as reflecting limitations on syntactic processes (rule application), the second views islands as defined over the products of these computations. Under this view islands result from restrictions imposed by the systems with which the syntactic component interfaces.\(^{10}\) In other words, islands amount to representational constraints. (Path-based approaches to locality in the GB era such as Kayne (1984) and Pesetsky (1982) were of this type, as was Koster’s (1978b, 1987) approach—although the issue of interfaces only became prominent with the advent of minimalism.)

Viewed in this light, one can call the first view the derivational view, and the second, the representational view (a similar division is endorsed in Hornstein, Lasnik, and Uriagereka 2007). Both views make islands conditions of the Faculty of Language in the Narrow sense (FLN, to borrow a term from Hauser, Chomsky, and Fitch 2002); that is, islands are conditions imposed on the workings on the syntactic component (“narrow syntax,” as it is now often referred to), or on the mapping from syntax onto the external systems. To put it in yet another way, islands exist due to properties of narrow syntax, or interface conditions.

The third view locates islands outside of FLN, and ascribes them to the Faculty of Language in the Broad sense (FLB). Specifically, the third view treats islands to be the result of processing/memory factors that constrain how linguistic knowledge is put to use. Thus, the third view takes islands to be a matter of performance, not competence (for representative works in this tradition, see

\(^{10}\) The external systems are typically called C-I (conceptual-intentional), and S-M (sensory-motor), which interface with syntax via LF and PF, respectively. But a richer conception of the interfaces may incorporate conditions on discourse-mapping, as investigated by Erteschik-Shir (1973, 2007); see also Goldberg (2006), from a different theoretical perspective.
In practice linguists have attempted to unify all island effects under a single umbrella, although it is conceivable, indeed, quite possible, that each of the three views just outlined captures a portion of the truth. As a matter of fact, the island territory appears to be quite diverse. If no general agreement exists as to how islands should be understood, there is a general consensus regarding the fact that islands come in several varieties and give rise to different degrees of deviance. Over the years, the different kinds of islands have been relatively well circumscribed. There are weak islands (WIs) and strong islands (SIs) (with a further subdivision among SIs to be discussed shortly).

WIs (a.k.a. selective islands) are domains which prohibit extraction of only certain types of elements (adjuncts), allowing other elements (arguments) to extract more freely (resulting in mild deviance). A typical example of a WI is the wh-island. As can be seen in (16), whereas an argument wh-phrase can move out of an embedded question, an adjunct wh-phrase cannot.

SIs (a.k.a. absolute islands) do not distinguish between arguments and adjuncts, and block displacement of both types of elements. Clausal adjuncts are typical SIs. Witness (17).

11 It is fair to say that such theories are often offered as “proof of concept” that island constraints can be explained by processing factors. I don’t know of any processing account that tackles the richness of island data that forms the basis of competence theories.

12 This view is to be kept separate from views that take grammatical constraints like subjacency to have a functional motivation to be found in the structure of the parser (see in particular Berwick and Weinberg 1984; Weinberg 1988).

13 This is true not only in the mainstream generative literature; see, e.g., Levine and Huraki (2006), who offer a unified treatment of islands in an HPSG-framework.

14 For an excellent review of WIs, see Szabolcsi 2006a.
Displaced constituents, such as raised subjects, also constitute robust islands, and are typically grouped with adjuncts as part of SIs. (I do not illustrate extraction of adjuncts out of nominals, since the latter is impossible even out of non-displaced constituents: *From which country did you meet [a professor t].)

Typically, the selectiveness of WIs is characterized in terms of Rizzi’s (1990) Relativized Minimality.15 In a situation like the wh-island, the embedded question word acts as an intervener (i.e., creates an island), prohibiting any relation between the matrix clause and the wh-word we are trying to extract.

The key factor in Relativized Minimality environments is how similar the intervener is to the moving element and the landing site. Schematically, in a configuration like (19), β will act as an intervener for any relation involving α and γ, unless |β| and |α, γ| are of distinct types.

(19) α ... β ... γ

Obviously, the task of the theorist is to find the adequate types of features entering into the schema in (19) that would distinguish between arguments and adjuncts. At present, it seems quite clear that this will necessitate an organization of features in terms of class, subclass (dimensions), and possibly values, a feature hierarchy/geometry not unlike that developed in phonological theory (for relevant discussion, see Rizzi 2001; 2004α; Starke 2001; Boeckx and Jeong 2004).

15 Although alternative treatments, based on rich semantic representations, exist. For a detailed exposition, see Szabolcsi and Zwarts (1997); Szabolcsi (2006α) and references therein.
The type of solution offered by Relativized Minimality to the phenomenon of WIs is very elegant, and makes a lot of sense from a minimalist perspective (particularly, if the set of features over which it is defined is not arbitrary).

SIs have so far received a less satisfactory treatment. Most approaches adopt and seek to refine Huang’s (1982) generalization (see also Cattell 1976) that any extraction out of nongoverned domains is barred (this is Huang’s Condition on Extraction Domain [CED]). For Huang non-governed domains were adjuncts and subjects, i.e., non-complements. Distinguishing non-complements from complements while at the same time unifying subjects (/specifiers) and adjuncts has proven extremely difficult, especially in the context of the minimalist program, where the putatively unifying notion of government is not a primitive.

Three types of solutions to this problem have been explored within minimalism. All share the intuition that SIs emerge derivationally, from the computational dynamics and resources of narrow syntax. That is, SIs are domains that become opaque to extraction in the course of the derivation. The three types of approaches differ in exactly how and why such domains emerge.

The first approach, pioneered by Uriagereka (1999a), pursues the idea that SIs are those domains that must enter the syntactic derivation as units without any visible subparts. Take a complex object consisting of A and B. Such an object can be constructed by combining A and B. One can then combine the resulting unit \{A, B\} with C, and that resulting unit \{A, B, C\} with D, and so on. The derivation is strictly monotonic (each element is introduced one step at a time):

\[
\begin{align*}
\text{(20)} & \quad a. \quad \{A\} \cup \{B\} = \{A, B\} \\
& \quad b. \quad \{A, B\} \cup \{C\} = \{A, B, C\} \\
& \quad c. \quad \{A, B, C\} \cup \{D\} = \{A, B, C, D\} \\
& \quad \text{etc.}
\end{align*}
\]

The diagram corresponding to the abstract derivation in (20) will, no doubt, look familiar to the reader.
Consider now a case where C in (20)/(21) is also complex, consisting of C and E. The corresponding derivation can no longer be strictly monotonic. If the derivation were monotonic, we would obtain the unit \( \{A, B, C, E\} \), but we want C and E to form a subunit \( \{A, B, \{C, E\}\} \).

\[
\begin{align*}
&\text{a. } \{A\} \cup \{B\} = \{A, B\} \\
&\text{b. } \{A, B\} \cup \{C\} = \{A, B, C\} \\
&\text{c. } \{A, B, C\} \cup \{E\} = \{A, B, C, E\} \text{ (crucially: } \neq \{A, B, \{C, D\}\})
\end{align*}
\]

To achieve the right result (i.e., the right constituency), C and E must be combined in a separate derivational space, in parallel to A and B, and only the resulting unit \( \{C, E\} \) will be able to be integrated monotonically, as in (23)/(24).

\[
\begin{align*}
&\text{a. } \{A\} \cup \{B\} = \{A, B\} \\
&\text{a'. } \{C\} \cup \{E\} = \{C, E\} \\
&\text{b. } \{A, B\} \cup \{C, E\} = \{A, B, \{C, E\}\} \\
&\text{c. } \{A, B, \{C, E\}\} \cup \{D\} = \{A, B, \{C, E\}, D\} \\
&\text{etc.}
\end{align*}
\]

Uriagereka’s insight is that SIs are domains that correspond to the unit \( \{C, E\} \) in our toy example. For him, SIs are complex left-branches in trees. It is because \( \{C, E\} \) must enter the derivation as a unit that neither C nor E can be subextracted from \( \{C, E\} \). Over the years, Uriagereka has explored various ways of capturing what it means for a complex left-branch to enter the derivation as a unit.
(see Uriagereka 1999a,b, 2003; see also Nunes and Uriagereka 2000; Zwart 2007; Johnson 2002). I will not go into such details here since nothing at this stage depends on them.16

Another way of capturing the opacity of SIs is to say that SIs are simply not part of the tree. Since they are not attached to the tree, their contents are inaccessible to the elements in the tree (assuming, plausibly, that elements can only search for “partners” from within their own tree), rendering subextraction impossible. This could very well be the case with adjuncts, which are only loosely connected with the rest of the tree. Here too there are various ways of capturing this intuition formally. One could say that SIs are never attached to the rest of the tree (see Uriagereka 1999a: 256f.; Hornstein and Nunes forthcoming for proposals that rely on the idea that adjuncts do not partake in projection), or that SIs are attached to the tree in an “acyclic” fashion, only after the entire tree has been built, and it’s too late to subextract anything because subextraction was required at an earlier stage in the derivation when the SI was not yet part of the tree (see Stepanov 2001, 2007 for such a late insertion account).

A third way of capturing the opacity of SIs is to treat them as domains that have exhausted their derivational potential, and have become deactivated as a result. Once again, there are several ways of implementing this intuition technically (see Boeckx 2003a; see also Rackowski and Richards 2005; Gallego 2007; Gallego and Uriagereka 2007a; Rizzi 2006; Rizzi and Shlonsky 2007). Let me sketch informally the proposal I made in Boeckx (2003a). My point of departure there was the idea that elements are displaced for a reason (Chomsky’s (1986a, 1993) Last Resort condition on movement). I proposed that once an element has been displaced there is no reason left for it to move any further.17 Elements contained in displaced constituents are

---

16 The most influential implementation of Uriagereka’s idea is to be found in Uriagereka (1999a), where complex left-branches are said to have been sent to the interfaces and linearized early. This is the idea of multiple spell-out, which has been incorporated into Chomsky’s (2000, 2001, 2004) phase-based model.

17 Thus stated, the idea appears to run against all the evidence for “successive cyclic” movement. For how to reconcile the two, see Boeckx (2003a: ch. 1) and especially Boeckx (2007a).
unextractable since they have been treated as a unit for purposes of
displacement, and “once a unit, always a unit,” as it were. I argued, in
part following Richards (2001), that the latter condition made sense
since displacement typically leads to pronunciation. Accordingly, what-
ever the cause of displacement is can be understood as an explicit
instruction to the interfaces to pronounce the moved element as a
whole in its moved position.18 Allowing for subextraction out of
moved elements would amount to “ambiguous” instructions: instruc-
tions to pronounce an element in multiple places (once in its displaced
position, and once as part of the moved unit it subextracted from),
which I argued the interfaces cannot tolerate.

To sum up this brief coverage of minimalist approaches to SIs, it
can be said that SIs emerge because at the point when subextraction
must take place, the whole unit containing the element to be moved
is not (yet) part of the tree (approach #2), or it is saturated (ap-
proach #3), or its content had to be obliterated for it to be part of the
tree in the first place (approach #1).

Although I think that all three approaches make a lot of sense, and
thus have the potential to deduce the existence of SIs in natural
language, they face quite a few difficulties. Let me mention the most
salient ones for each. Approach #1 cannot easily accommodate the
fact that not all left-branches are SIs (SIs are typically left-branches
created by adjunction or movement). Approach #2 works best for
adjuncts, which are indeed loosely connected to the “spine” of the
tree, but does not generalize to SIs caused by displacement (for an
element to be displaced, it must have been part of the tree by de-
definition. But if it is part of the tree, one cannot say that it’s not there (yet)
by the time subextraction must take place). Approach #3 is designed
to capture SIs that are the result of displacement, but must add special
conditions to capture SIs that are the result of adjunction.

In sum, approach #1 appears to be too rigid, and approaches #2 and
#3 are incomplete (in a complementary way, in fact). Perhaps this is the
best one can do. Perhaps it is in fact the best one should do. As I already

18 Rizzi’s (2006) notion of (freezing) criterion can be understood in this sense of
interface instruction as well.
pointed out above, it has sometimes been argued (see Stepanov 2001, 2007; see also Sprouse 2007) that SIs should not receive a uniform treatment since they seem to give rise to non-uniform patterns of deviance. Furthermore, some domains (adjuncts) are more robust SIs than others (subjects) cross-linguistically. And, it has also been argued that some operations like ellipsis can “cover up” certain SI-violations (subjects), but not others (adjuncts) (see Merchant 2001). I will examine this type of evidence in due course (see Chapter 4). For now I want to point out that differences among phenomena do not directly argue against unifying them (differences may be due to independent, extrinsic factors). I will in fact propose a unified account of islands in the chapters that follow, although I should stress that I will be less concerned with unifying islands (the topic of what one might call a “special” theory of locality) than with unifying locality of selection and locality of long-distance dependencies (the topic of a “general” theory of locality).

To achieve unification in accordance with minimalist ideals (i.e., to make perfect sense of locality) I will try to depart as little as possible from “virtual conceptual necessity.” Specifically, I will rely as much as possible on the most fundamental operation of syntax, *Merge*, and try to generalize its properties to other aspects of syntax using the mathematical tool of symmetry. Because *Merge* will be so essential to my argument, I will devote an entire chapter (Chapter 3) to explaining all the properties that it has, and that I want to make use of in the context of locality of selection (“cartographies”; Chapter 4), and locality in long-distance dependencies (“islands”; Chapter 5). Because the argument will necessarily be abstract and technically quite involved, I will first sketch the direction of unification I want to push in Chapter 2. This should provide the reader with the basic intuition, which subsequent chapters will try to capture in a formally more adequate fashion.
Outline of a General Theory of Locality

2.1 Merge and Move

It is now standard in the minimalist literature to regard Movement as just one of the possible instantiations of Merge. Merge is the most fundamental syntactic process that takes two elements, $\alpha$ and $\beta$, and combine them into a set $\{\alpha, \beta\}$. Moved elements are simply elements that have been merged again. They first had to be merged to be part of the derivation. The second instance of Merge made it possible for them to enter into a syntactic relation with multiple, non-adjacent, elements in the very same derivation.

In the early days of minimalism (Chomsky 1993, 1995), this view of movement as (re-)Merge was captured in terms of copying. A moved element was seen as a copy of the originally merged element having been merged again (typically) higher up in the tree, as schematized in (1).

1. a. Merge $\alpha$ and $\beta$, forming $\{\alpha, \beta\}$
   b. Keep building the tree, adding elements like $\gamma$, $\delta$, $\psi$
   c. Copy $\beta$
   d. (Re-)merge $\beta$ to $\psi$

In set-theoretic format: $\{\beta, \{\psi, \{\delta, \{\gamma, \{\alpha, \beta\}\}\}\}\}$.
Chomsky (2003: 307; 2004) further simplified the Move-as-Merge view by eliminating the copying step (1c) and allowing the very same element to be merged multiple times, giving rise to representations like (2). (The representation in (2) is an instance of multidominance of the type first explored in a minimalist context in Gaertner 2002.)

Chomsky (2004: 110) called the first instance of Merge of a particular element \textit{External Merge}, and the second instance of Merge of that element \textit{Internal Merge}, emphasizing that both are instances of Merge, solely distinguishable when looked at from their source. Whereas $\beta$ is external to the derivation prior to being merged to $\alpha$ in (2), it is internal to the derivation when merged to $\psi$. Graphically:

By introducing the External/Internal-Merge equivalence Chomsky effectively asserted the \textit{source independence} of the most basic operation
in syntax. That is, Merge is the very same operation irrespective of where the elements being merged come from.

Because this claim will play an important role in what follows, let me repeat it thus:

(4) The Merge-Move Equivalence
    Merge is source-independent

Based on this claim, Chomsky has repeatedly pointed out that the displacement property of natural language follows at once from the most basic operation giving rise to discrete infinity: iterated Merge. That is to say, it would take an extra assumption to rule out Internal Merge (i.e., it would require making Merge source-dependent).

As Luigi Rizzi points out (p.c.) (see also Koster 2007), the equivalence in (4) is in some sense a radical reformulation of Emonds’s (1976) Structure-Preservation Hypothesis. Emonds’s insight was that movement creates configurations that can be independently generated by the fundamental structure-building mechanism.\(^1\) To take a simple example, according to Emonds’s Structure-Preservation Hypothesis, movement into SpecCP would be barred if SpecCP could not be generated by the grammar independently of movement. Today we would reformulate this by saying that there is no node such that it can only be generated by Internal Merge, and not by External Merge. This is essentially reasserting (4).

The equivalence in (4) is the first instance of symmetry encountered in this study. The intuitive idea behind symmetry will be familiar to everyone: It is that which remains unaffected by (is invariant under) any kind of change (transformation). Take your favorite book and rotate it \(360^\circ\), and it will look the same as before the rotation. Your book is symmetric under \(360^\circ\)-rotation.

\(^1\) This statement has to be qualified somewhat. Emonds distinguished between structure-preserving transformations and root transformations (operations confined to root clauses), both of which were called non-local, or major transformations and were further distinguished from local transformations. Today we have come to regard all transformations as structure-preserving, all of which are generated by Merge.
But notice that the symmetry expressed by (4) is a symmetry of a rule. Although Emonds talked about “structure” being preserved, it’s not so much a structure, or an object that (4) is about. What (4) says is that the formulation of Merge does not change under the operation Select, which picks elements for purposes of Merge. (On Select, see Chomsky 1995: 226.) Merge is the same process, irrespective of whether Select reaches into the lexicon or uses elements from the derivation. To put it differently, the source independence of Merge is the reason why movement is structure-preserving.

Having discussed symmetry in the context of (4), let me now try to establish a new symmetry into the grammar.

### 2.2 Products of Merge and Products of Move

If it is now standard to regard Merge and Move as equivalent, it is not standard to regard the product of Merge to be equivalent to the product of Move.\(^2\) Let me clarify what I mean by product of Merge (and Move).

---

\(^2\) The similarity between Move (Chain formation) and Projection to be argued for in this section has not gone completely unnoticed in the previous literature, although I think that it has not been exploited to the fullest extent. But let me mention a few suggestive passages I have been able to find.

Chomsky (1995: 246) refers to the element that projects as the target, “borrowing the [latter] notion from the theory of movement in an obvious way.” Likewise, Speas (1990: 43) coins the term Project \(\alpha\) by drawing an explicit parallelism with Move \(\alpha\), and further uses the term Projection Chain to capture the idea that a word of syntactic category \(X\) is dominated by an uninterrupted sequence of \(X\) nodes, much like a movement chain cannot be interrupted (locality principle).

Perhaps the most explicit parallelism between Move and Project is to be found in Koster (1987: chs. 1–2). While Speas’s Project \(\alpha\) was based on an analogy with Move \(\alpha\) (Speas wanted to let elements project freely, and constrain labeling by independent principles, much like Move \(\alpha\), let Move happen and ruled out unwanted outputs by Filters), Koster’s argument goes beyond the level of analogy. Consider the following passage (Koster 1987: 17):

“It is somewhat accidental, perhaps, that vertical grammatical relations (like the relations between the members of a projection) have hardly been studied from the same perspective as “horizontal” relations like anaphora and movement (an exception is Kayne (1984)). If we abstract away from the distinction related to dominance, it might appear that [i] simply sums up the properties of all local relations of grammar [...]
So far I have used the term Merge to refer to the deceptively simple operation of combining two elements into a set. But Merge is a little bit more complex than this. In its traditional formulation (Chomsky 1995: 243),\(^3\) one of the elements merged is said to project, providing the set with a label, which governs the syntactic behavior of the set. So, if V and N merge, one of them—say, V—projects, making the set \{V, N\} a VP, that is to say, a set that will behave like a V for purposes of further combination (when V(P) merges with T, for example). The label codes the fundamental “is-a” relation of phrase-markers discussed in Chomsky (1955).

Accordingly, if Merge is defined as combining \(\alpha\) and \(\beta\), the product of Merge is defined as the set \(\{\alpha, \beta\}\), labeled by one of the two units combined, say, \(\alpha: \{\alpha, \{\alpha, \beta\}\}\).

The product of Move was defined as a “function chain” (or, simply, “chain”) in Chomsky (1981: 45). A chain was defined there as a sequence composed by all the positions occupied by a moved element in the course of a given derivation. That is, a chain was a device to concisely represent the derivational history of a given element. We can think of a chain as a collection of all the occurrences of a given element. In the Government-and-Binding era, a chain would typically consist of the moved element and all its traces. Within early minimalism, a chain would comprise all the copies of an element. With the advent of the notion of Internal Merge, a chain would be characterized by all the positions into which a given element has been merged. Such positions are standardly represented in terms of their sisters. So SpecCP would be represented as C’. For purposes of this chapter it is not important to go into the details of chain representations. As a matter of fact, because

\[\begin{align*}
\text{i. a. obligatoriness} \\
\text{b. uniqueness of the antecedent} \\
\text{c. c-command of the antecedent} \\
\text{d. locality}^* \\
\text{3 I will return at length to alternative formulations in Chapter 3.}
\end{align*}\]

Koster (1987: 71) further claims that the parallelism between vertical and horizontal syntax is simply obscured by the distinction between isotopic (in-situ) and non-isotopic (at a distance) property-sharing/syntactic relation, which corresponds to the distinction between Merge (local relation) and Move (relation at a distance).
it makes certain properties more conspicuous, I will often use the
(non-minimalist) trace notation in the following pages. My only
concern here is to establish a new symmetry principle, an equivalence
between Projection and Chain, as in (5).

(5) The Product of External/Internal-Merge Equivalence
    Projection and Chain are symmetric objects

Notice that (5) expresses a symmetry of objects. I will argue in Chapter
3 that (5) is deducible from the symmetry of a rule (on a par with (4)),
viz. the very symmetry at the heart of the process of Merge—that
which says that Merge(α, β) = Merge(β, α). We can characterize this
fundamental symmetry of Merge as “Order-independence”. But for
now, let us keep to (5), and see what it would entail.

Recall that the motivation behind Chomsky’s claim that Internal
Merge and External Merge are equivalent was to make displacement
follow from the basic operation of the grammar. Seen in isolation,
displacement is an unexpected property of natural language, one
that is absent from artificial grammars, and that was consistently
seen as an “imperfection,” requiring special motivation (see
Chomsky 2005: 12; 2006c xi). By reducing it to a process that is
arguably part of virtual conceptual necessity (Merge), displacement
can be rationalized. It doesn’t require extra assumptions.

By applying the same logic with (5), I hope to make seemingly
odd properties of chains such as island effects follow from prop-er-
ties of projection, which I think can be more easily explained
because they are associated with a more basic process (Merge). Having made the motivation behind (5) clear, let me examine the
Product of Merge/Move equivalence more closely.

---

4 Rationalization is something that symmetry arguments are extremely good at. What
is symmetric is not seen as requiring special explanation. What is symmetric follows as a
matter of course. This is arguably the main reason why symmetry arguments have figured
so prominently in theoretical sciences. As Brody (2003: 205) correctly pointed out, “it is
the departures from symmetry [...] that are typically taken to be in need of explana-
tion.” I return to the explanatory advantages of symmetry in Chapter 6.

5 This sort of strategy, which consists in solving a more difficult problem by
reducing it to an aspect of a problem already solved, is well known to mathematicians.
Descartes and Abel, for example, often resorted to it.
2.3 Unifying Chains and Projections

The equivalence in (5) aims at unifying two syntactic constructs that are traditionally kept separate. In the more advanced sciences, such as physics, unification typically proceeds in several steps. Entities being unified are shown to be (i) commensurable, (ii) permutable (interchangeable), and finally (iii) traceable to the same underlying process (with diversity resulting from the application of a fixed number of symmetry transformations).

If we follow this example, our task ahead consists in showing that projections and chains are commensurable, permutable, and relatable via a symmetry transformation.

I will show in this section that chains and projections can be unified in this way. After that, I will show that they obey the same kind of locality conditions. Before proceeding let me stress that the purpose of this chapter is to lay out the basic intuition behind bare syntax, and establish certain generalizations. I will not attempt to deduce these generalizations in this chapter, nor will I try to find the most adequate formal implementation of the basic idea here. These are tasks that I undertake in subsequent chapters.

2.3.1 Commensurability

Consider first the commensurability of projections and chains. The simplest way to establish a correspondence between the two is to go back to the roots of our modern understanding of syntactic phrases, Chomsky’s (1970) seminal Remarks on Nominalization. There Chomsky introduced the X-bar Schema. Setting adjuncts aside, the original X-bar schema looked like (6).

---

6 This method of unification goes back to J. C. Maxwell and the unification of electricity and magnetism.

7 The notion of commensurability is already contained in the very term symmetry, whose etymology (Greek σύν + μετρ(ε)ός) is something like “common measure.”

8 Adjuncts are a pesky species, and I will try to keep them out of this chapter as much as possible. I’ll discuss their nature in detail in Chapter 3.
Since Chomsky’s proposal, the structural status of specifier-of and complement-of hasn’t changed much (only the positioning of adjoined material has fluctuated).

From the perspective of projection, we can see in (6) that an element X can be “stretched” to occupy the following three states (“projection levels”): (i) minimal, (ii) intermediate, and (iii) maximal: $X^\circ$, $X'$, and $X''$.

This, as far as we know, is valid for all syntactically relevant categories, lexical and functional alike. It gives rise to the beautiful symmetry of projection achieved in the mid-80s, and represented in (7).

(7) $V''$ $N''$ $A''$ $P''$ $I''$ $C''$
    $V'$ $N'$ $A'$ $P'$ $I'$ $C'$
    $V^\circ$ $N^\circ$ $A^\circ$ $P^\circ$ $I^\circ$ $C^\circ$

All elements functioning in natural language syntax appear to exhibit the same three units of projection.

Turning now to chains, we witness a strikingly similar state of affairs. Elements undergoing movement typically start in a thematic position (for arguments), move through some intermediate position (“successive cyclicity”), and finally reach their ultimate landing site.

---

9 This is most clearly stated in Rizzi’s (2006) characterization of the form of chains. Rizzi claims that chains begin in an s-selectional position (thematic position for arguments) and terminate in a criterial position (a position where criteria like the wh-criterion, or subject criterion are met). In between these two positions, Rizzi argues that a moving element may pass through intermediate landing sites for some feature-checking purposes (crucially distinct from the satisfaction of a criterion) or due to locality requirement forcing short movement steps (as I argued in Boeckx 2003a, 2007a).

10 In many cases, chains contain several intermediate landing sites; this is on a par with Chomsky’s (1970) claim that the intermediate projection level can be iterated to host adjuncts, an issue I am setting aside for now.
Since chains code the history of derivation of moving elements, it is no surprise to see chains contain a maximum of three types of juncture: minimal, intermediate, and maximal. Put differently, chains, like projections, organize themselves along a tripartite schema, consisting of a beginning, a middle, and an end, hierarchically organized. Schematically:

\[
\begin{align*}
(8) & \quad [X'' [X' [X \ldots ]]] \\
(9) & \quad (XP, t'_{XP}, t_{XP})
\end{align*}
\]

All the elements within a chain are copies of one another, just like all projection levels are projections of one and the same element, the head of the phrase.\(^\text{11}\) Chains can thus be said to be headed (all the members of a chain share a certain property, call it their type), and, as we will see momentarily, the members of a chain are strictly ordered, pretty much in the same way as the projection levels are in a phrase.

As has long been recognized, some feature-checking sites (which Rizzi 2006 calls criterial positions; Boeckx 2003\textit{a} and Richards 2001 call them strong positions) constitute freezing points, sites which mark the maximal vertical range of a chain (what Ross (1967: 288) called an element’s “maximal strip”). Likewise, Fukui 1986 argues that some featural relations mark the maximal level of projection. (The term “saturation” is sometimes used to characterize this effect.) For example, according to Fukui, case/agreement relations (which he equates with

\(^\text{11}\) For arguments that projection levels are literal copies of the head, see Boeckx (2006\textit{a}; forthcoming \textit{c}), Harley (2004), and Rezac (2003). All these arguments are based on Chomsky’s (1994, 1995) Bare Phrase Structure.
specifier-head configurations) close off projections. Similarly, Boeckx (2006b) argues that Person-cum-Case feature-checking closes off A-chains (see also Chomsky 1986a, 1993, 2000, and Lasnik 1999a, who argue that Case-checking freezes or deactivates an element).

As we can see, the same descriptive vocabulary, the same phenomenology, can be used to characterize both chains and projections. They can therefore be said to be commensurable.

2.3.2 Permutability

Chains and projections—traditionally kept distinct—have also become permutable/interchangeable in the recent syntactic literature. This is the case in two kinds of situations: (i) situations where the moved element projects, where a chain becomes a projection and (ii) situations where projections “exist” in parallel, sometimes achieving the same result as what previous research had attributed to movement.

Let me briefly illustrate both kinds of situation. (I warn the reader that situations of type (ii) may be a bit more difficult to grasp, and I won’t be able to devote enough space to do justice to them. I encourage the reader to turn to the primary literature (especially Hornstein and Uriagereka 2002) for fuller exposition.)

2.3.2.1 Projection by movement Typically, the element that hosts or attracts the moving element projects. This generalization, so often taken for granted, has recently been re-examined, and counterexamples have been brought up.

To my mind the most persuasive instance of a moved element projecting comes from Donati’s (2006) study of free relatives. There is an intuitively obvious sense in which whatever in (10) has a dual role.

(10) I’ll eat [whatever you ordered last night]

Whatever here is understood as the object of ordered (internal to the relative clause, it is a (non-projecting) dependent), while at the same time dictating the syntactic behavior of the whole relative
clause (functioning as the object of *eat*). By this reasoning, it should be the head of the whole relative clause.

Donati (2006) proposes an analysis of free relatives according to which *whatever* undergoes regular wh-movement internal to the relative clause, but ends up projecting its label, turning the CP into a DP, as depicted in (11). (For a similar proposal, see Iatridou, Anagnostopoulou, and Izvorski 2001: 224f.; see also Chomsky forthcoming.)

\[(11)\]
\[\begin{align*}
\text{a. } &[\text{CP } C^\circ \text{ you ordered whatever}] \\
\text{b. } &[\text{DP whatever } [\text{CP } C^\circ \text{ you ordered } t]] \\
\text{(not: } &[\text{CP whatever } [C^\circ C^\circ \text{ you ordered } t]])
\end{align*}\]

Perhaps more conspicuously:

\[(12)\]
\[\begin{array}{c}
\text{DP} \\
\text{CP} \\
\text{D^\circ} \\
\text{whatever } C^\circ \\
\text{IP}
\end{array}\]

as opposed to:

\[\begin{array}{c}
\text{CP} \\
\text{DP} \\
\text{C^\circ} \\
\text{whatever } C^\circ \\
\text{IP}
\end{array}\]

A similar analysis was proposed by Bhatt (1999) for relative clauses in general. Building on Vergnaud’s (1974) so-called promotion analysis, Kayne (1994) proposed that relative clause formation proceeds as in (13), corresponding to *The boys that you met (were nasty).*

\[(13)\]
\[\begin{align*}
\text{a. } &[\text{CP } C^\circ \text{ you met boys}] \\
\text{b. } &[\text{CP boys } C^\circ \text{ you met } t] \\
\text{c. } &[\text{DP the } [\text{CP boys } C^\circ \text{ you met } t]]
\end{align*}\]

Notice that under Kayne’s analysis, the D-head of the relative clause is not moved from within the relative clause (unlike the D-head in Donati’s analysis above). For Kayne, *the* and *boys* in the sentence under discussion do not form a constituent.

For Bhatt (1999), they do form a constituent since for him *boys* projects under movement (as in Donati’s analysis), as depicted in (14).
Additional cases of moving elements that project are proposed in Citko 2006, Starke 2004, and Jayaseelan 2007. Taken together, the cases discussed by these authors suggest that movement feeds projection, that is to say, the topmost position in a chain morphs into the minimal level of a projection.

2.3.2.2 Reprojection Instances of “reprojections” provide another case attesting to the symmetry of chain and projection, more specifically, the symmetry of XP and X°. As a matter of fact, we’ll see at least one situation where movement was once thought to be needed, but where the same result can be achieved by letting an XP-element in situ project (more accurately, reproject, as the projection in question overrides the existing standard labels). So, this is a case where projection achieves the same result as movement chain, reinforcing the idea that projections and chains are two sides of the same coin.

The most detailed instance of reprojection I am familiar with comes from Hornstein and Uriagereka (2002). Hornstein and Uriagereka’s (2002) original motivation for reprojection were cases where certain dependencies established in the covert component of the grammar (e.g., instance of LF-movement) are subject to island

\[ \text{DP the [NP boys [CP C° you met t]]] } \]

(i–ii)

\begin{itemize}
  \item I drank a bottle of beer
  \item I broke a bottle of beer
\end{itemize}

In (i) beer satisfies the selectional requirement of the verb drink, whereas in (ii) bottle satisfies the requirement of break. That beer is the semantic head of the nominal expression in (i), but not in (ii), is clear from the fact that it may be omitted in the latter, but not in the former.

(iii–iv)

\begin{itemize}
  \item I drank beer yesterday
  \item I broke a bottle yesterday
\end{itemize}

So selectional restrictions may be satisfied by either the content NP or the container DP. Yet agreement is always determined by the container DP.

(v)

\begin{itemize}
  \item the bottles of beer that I drank were/*was big
\end{itemize}

\footnote{Citko (2006) even allows both the host and the moved elements to project in some cases. Starke (2004), and, following him, Jayaseelan (2007), propose that even unmoved specifiers may project.}

\footnote{Another instance of reprojection may be found in complex nominal expressions. Consider the syntax of container-content relations (Selkirk 1977; Castillo 1998, 2001). Selkirk (1977) observed that such relations are ambiguously encoded in the syntax. Witness (i–ii).}

\begin{itemize}
  \item I drank a bottle of beer
  \item I broke a bottle of beer
\end{itemize}

In (i) beer satisfies the selectional requirement of the verb drink, whereas in (ii) bottle satisfies the requirement of break. That beer is the semantic head of the nominal expression in (i), but not in (ii), is clear from the fact that it may be omitted in the latter, but not in the former.

\begin{itemize}
  \item I drank beer yesterday
  \item I broke a bottle yesterday
\end{itemize}

So selectional restrictions may be satisfied by either the content NP or the container DP. Yet agreement is always determined by the container DP.

\begin{itemize}
  \item the bottles of beer that I drank were/*was big
\end{itemize}
effects that are absent from overt syntax. One such case is the quantifier-induced island illustrated in (15).

(15) *nobody gave every child anything

Here a (strong) quantifier like every blocks the licensing of a negative polarity item like anything by negation (no-body). As (16) shows, every does not block the establishment of an operator variable pair by overt wh-movement. (17) shows that it is the strong quantifier that induces the blocking effect. Once every is removed, (15) is licit.

(16) what did nobody give every child?  
(17) nobody gave the two children anything

The basic idea developed by Hornstein and Uriagereka is the claim that “strong” (binary) quantifiers like every (but unlike (the) two) are in some sense transitive predicates. As Larson and Segal (1995: 273) observe, strong quantifiers, like transitive verbs, combine with two hierarchically ordered “arguments,” in this case, a restriction (internal argument), and a scope (external argument). So, in a sentence like every boy likes Mary, boy and likes Mary are arguments of every. This

vi. the bottles of beer that I broke were/*was big  
Pronominalization is also determined by the container DP.  
    vii. I broke two bottles of beer before I could drink them/*it  
    viii. I drank two bottles of beer before I broke them/*it

Castillo (1998, 2001) argues that the dissociation between selection and agreement requires a more dynamic conception of the notion “syntactic head,” which is in the spirit of reproj ection analyses. Without going into the details of his account, we can say that whereas the container determines agreement, the phi-features may be overridden on grounds of interpretability, turning the content NP into a head. The prediction, then, is that subextraction will be possible only when the phi-features of the container can be overridden (when the container NP is not the head for interpretability/selection). The prediction is borne out, as shown in (ix–x).

ix. what did you drink [a bottle of t]  
x. *what did you break [a bottle of t]

If the above analysis is on the right track, it may help us getting around instances of apparent intervention in the domain of selection, of the kind discussed in the context of cartographies in Chapter 1. (For arguments that focused material reprojects, see Herburger 2000 and especially Irurtzun 2007.)
is arguably something we would like to be able to represent at the syntax-semantics interface. One way of capturing this formally, suggested by Hornstein and Uriagereka, is to say that the standard IP-structure for the sentence in (18a) is turned into (18b), via reprojection, motivated by semantic considerations (akin to argument structure). (See Pietroski 2005 and Larson 2005 for further developments of this abstract similarity between scopal and thematic relations; see also Chapter 5.)

(18)  a.  \[ [\text{IP} [\text{DP every [NP boy]]_i I^0 [t_i \text{likes}[\text{Mary}]]] \] \] overt syntax

```
  IP
  └── DP_i
      └── D N(P) IV P
          └── every boy
  I  VP
      └── t_i likes Mary
```

b.  \[ [\text{DP} [\text{D' every [NP boy]]_i [\text{IP} I^0 [t_i \text{likes}[\text{Mary}]]]] \] \] syntax-semantics interface

```
  DP
  └── DP_i
      └── D N(P) IV P
          └── every boy
  IP
      └── t_i likes Mary
```

Apart from purely semantic/interpretive considerations, the reprojected structure has the advantage of predicting the blocking effect induced by the quantifier noted in (15). The basic idea here is that quantifiers like *every* turn the structure surrounding them into a DP, and DPs are islands for various dependencies. Since reprojection happens in covert syntax, only covert dependencies will be affected by this change of structure—predicting the contrast between (15) and (16).
Hornstein and Uriagereka’s idea of reprojection illustrates the symmetry of Merge we are concerned with here: just like the head can project, so can its complement project (reproject).

The second instance of reprojection I am familiar with is presented in Epstein et al. (1998: 70), and concerns the syntax of experiencers. Here we will see how reprojection may have the same effect as movement.

Consider a sentence like (20).

(20) John seems to Bill [t_i to be a genius]

The most puzzling aspect of this sentence is the fact that the experiencer Bill does not block raising of John even though it is expected to do so because it does c-command into the embedded clause. Evidence that the experiencer c-commands (into) the embedded clause comes from a variety of sources. For example, an experiencer pronoun cannot be interpreted as coreferential with a proper name in the embedded clause, as is typically the case when pronouns c-command proper names they are coindexed with (Binding Condition C effect).

(21) They seem to him [t_i to like John]

At this point we face a c-command paradox. Since raising is possible in (20), we may conclude that the experiencer does not block raising
(intervention requiring c-command; Rizzi 1990). But the Condition C effect suggests that the experiencer does c-command into the embedded clause. We are thus faced with two structural possibilities, one where the experiencer is a PP (problematic for binding since him would not c-command into the embedded clause), the other where the experiencer is a DP (with the proposition adjoined to it) (problematic for movement since him should trigger a minimality effect):

\[(22)\]

\[
\text{a. } \left[\text{They}_k \ldots \text{VP seem}_i \left[\text{VP } \left[\text{PP to him}\right]\right] t_i \left[\text{IP } t_k \text{ to like John}\right]\right]
\]

\[
\text{b. } \left[\text{They}_k \ldots \text{VP seem}_i \left[\text{VP } \left[\text{DP to [him]}\right]\right] t_i \left[\text{IP } t_k \text{ to like John}\right]\right]
\]

Kitahara (1997) offers a solution to this paradox that relies on covert movement (in a way similar to how scope is captured via covert Quantifier Raising). (For variants of this analysis, see Boeckx 1999, Torrego 2002, and Epstein and Seely 2006.) In a nutshell, Kitahara proposes that the experiencer does not c-command (into) the embedded clause at the time when subject-raising applies (overt syntax), but at a subsequent stage of the derivation the experiencer DP raises to a position from where it can c-command into the embedded clause. This can be schematized as in (23) (slightly modified from Kitahara’s own proposal).

\[(23)\]

\[
\left[\text{They}_k \ldots \text{VP seem}_i \left[\text{VP } \text{him}_g \left[\text{VP } \left[\text{PP to } t_g\right]\right] t_i \left[\text{IP } t_k \text{ to like John}\right]\right]\right]
\]

As an alternative, Epstein et al. (1998) propose that the experiencer phrase, previously a PP, is turned into a DP in covert syntax (as in (24)), which allows the experiencer to c-command into the embedded clause.

\[(24)\]

\[
\left[\text{PP to him}\right] \rightarrow \left[\text{DP to him}\right]
\]

The reprojection in (24) has the effect of combining the virtues of (22a) and (22b).

\[(25)\]

\[
\text{a. } \left[\text{They}_k \ldots \text{VP seem}_i \left[\text{VP } \left[\text{PP to him}\right]\right] t_i \left[\text{IP } t_k \text{ to like John}\right]\right]
\]

\[
\text{b. } \left[\text{They}_k \ldots \text{VP seem}_i \left[\text{VP } \left[\text{DP to [him]}\right]\right] t_i \left[\text{IP } t_k \text{ to like John}\right]\right]
\]

As in the case of reprojection discussed by Hornstein and Uriagereka, the reprojection in (24) happens after overt movement dependencies
have taken place, and will thus only affect dependencies established at the syntax-semantics interface, such as binding. What was seen as a conflicting c-command requirement can now be interpreted as a property of derivational systems allowing for reprojection. And where movement was thought to be needed (Kitahara’s solution in (23)), Epstein et al. show that (re)projection can be equally successful, providing another case where chain and (re)projection achieve the same result.

The phenomena and analyses just discussed do not exhaust the range of situations where moved XPs and projecting elements become virtually indistinguishable. Of relevance here are cases where a moved specifier is reanalyzed as a head, triggering a change in parameter setting and thus a language change (e.g., the pronominal subject turned copula in languages like Hebrew; see Roberts and Roussou (2002); Whitman (2000) for various examples of this sort).

Although in most studies just discussed the interchangeability or permutability of Merge and Move, or chain and projection, is not explicitly acknowledged, it seems to me that this is exactly what their results point to, which would reinforce the unification of (product of) Merge and (product of) Move in (5).

2.3.3 Symmetry transformation

Perhaps the clearest way of stating the equivalence of product of Merge and product of Move in (5) is to show how a chain can be seen as a projection via a symmetry transformation. This is I think the intuition behind Uriagereka’s (1998: 399) thought-experiment, which I’ll reproduce in this section.

Uriagereka speculated that chains may be describable in terms of “superprojections” that one could easily visualize via a slight

---

14 Perhaps the earliest instance of reprojection is to be found in Epstein (1992), where covert head-movement relabels the structure. Since the status of head-movement is not clear in the current theory, I won’t go into Epstein’s analysis here.

15 Perhaps the equivalence in (5) may be able to make sense of phenomena that suggest that chains have “embedding” effects (as projections do). Thus, Norbert Hornstein (p.c.) observes that (sideward) movement out of adjuncts that are stacked forced the adjuncts to be interpreted as being embedded within one another.
manipulation of graphs (standard tree diagrams). Before reproducing Uriagereka’s original, let me illustrate the idea using a simple abstract example, where an element $\alpha$ is moved from its base position (sister of $P$) to the specifier position of $X$. The effect will be clearest if Move is represented as remerge, as we did in (2). The representation of this abstract example is given in (26).

(26)

```
\begin{center}
\begin{tikzpicture}
    \node (X) at (0,0) {$X$};
    \node (Y) at (1,-1) {$Y$};
    \node (Z) at (2,-2) {$Z$};
    \node (W) at (3,-3) {$W$};
    \node (P) at (4,-4) {$P$};
    \draw (X) -- (Y);
    \draw (Y) -- (Z);
    \draw (Z) -- (W);
    \draw (W) -- (P);
    \draw (X) -- (W);  % Move
\end{tikzpicture}
\end{center}
```

Now suppose we flip $\alpha$ to the other side of the “spine” formed by $X$-$Y$-$Z$-$W$ (a symmetry transformation known as reflection around the axis $X$-$Y$-$Z$-$W$). We obtain (27), where it is now $\alpha$ that dominates the whole syntactic object.

(27)

```
\begin{center}
\begin{tikzpicture}
    \node (X) at (0,0) {$X$};
    \node (Y) at (1,-1) {$Y$};
    \node (Z) at (2,-2) {$Z$};
    \node (W) at (3,-3) {$W$};
    \node (P) at (4,-4) {$P$};
    \draw (X) -- (Y);
    \draw (Y) -- (Z);
    \draw (Z) -- (W);
    \draw (W) -- (P);
    \draw (X) -- (W);  % Move
\end{tikzpicture}
\end{center}
```

Uriagereka’s original example, reproduced in (28)–(29), was slightly more complex in that it included an intermediate step of movement, and made use of the symmetry transformation known as rotation,
as opposed to reflection. It also represented movement more traditionally by making use of traces. None of these differences matter for the point at hand, which is that a slight modification of a typical tree diagram yields an isomorphic representation, another familiar-looking tree.

The symmetry transformation visually captures the symmetry expressed in (5). As I have argued in the previous three sections, products of Move and products of Merge are equivalent. Before turning to what this symmetry means for locality matters, let me mention two more ways in which chains and projections have been treated on a par in the recent literature (though once again, the symmetry has not been explicitly noted in the works cited).
2.3.4 Additional similarities

The first additional similarity between chain and projection I’d like to discuss is their interface status. Both chains and projections are products of syntax that must “disappear” at some point in the mapping process to the external systems. The idea is very clear when we think of the mapping from syntax to sound:\(^{16}\) chains and projections are never pronounced as such, they must be reduced so as to ensure that only one occurrence of a chain or a projection is.\(^ {17}\) (For a well-worked-out mechanism of chain reduction, see Nunes 2004. On a similar projection-reduction mechanism based on Boeckx’s (2006a, forthcoming c) claim that projection levels are literal copies of the head, see Lohndal 2007a.)

A second point of similarity between labels and chains comes from the fact that the very same principle has been appealed to by different authors and independently of one another to argue that neither label nor chain are compatible with minimalist tenets and should therefore be eliminated.

Thus, Seely (2006) argues that projections should be eliminated on grounds of Inclusiveness. The latter is defined in Chomsky (1995: 228) as follows:

\(^{16}\) The status of projections and especially that of chains is less clear on the “meaning” side. Although chains have been used to account for various processes involving scope and binding, particularly in the context of reconstruction, it is, to my knowledge, never the case that multiple occurrences of a given element are interpreted. The use of chains on the meaning side amounts to allowing different portions of an element to be interpreted in different positions as in reconstruction effects (e.g., the operator part of a wh-word is interpreted in SpecCP, but its restriction is interpreted in a lower, “reconstructed” position, as in (i)).

(i) a. I wonder which picture of himself, Mary said that John, likes  
b. LF: I wonder [which picture of himself,] Mary said that John, likes [which picture of himself,]

Hornstein (2001: 5) is to my knowledge the first to suggest an analog of Kayne’s linearization algorithm on the meaning side of the grammar (for scope) (a Scope Correspondence Axiom corresponding to Kayne’s Linear Correspondence Axiom).

\(^{17}\) Moro (2000) and Chomsky (2004) also suggest that labels must be reduced to a unique element in instances where more than one element project, as in Moro’s analysis of small clauses and Chomsky’s analysis of adjunction as Pair-Merge. I come back to the issue of complex labels in Chapter 3.
Inclusiveness Condition
Any structure formed by the computation is constituted of elements already present in the lexical items selected for N; no new objects are added in the course of the computation apart from the rearrangement of lexical properties.

Seely (2006: 194) puts his argument as follows:

As characterized in [Bare Phrase Structure; Chomsky 1994, 1995], Merge takes as input the elements A and B, and Merge gives as output the set \( \{A, \{A, B\}\} \). Note that Merge creates an identity relation between the label \( A \) and the category \( A \). This identity relation is not itself a lexical feature, nor is it a combination of features. The relation itself does not occur in the lexicon, and therefore it violates the Inclusiveness condition.

In a similar vein, Hornstein (1998) argues that chains fall outside the realm of elements that are already present in the lexicon, and so violate Inclusiveness (see also Epstein and Seely 2006).

Hornstein’s and Seely’s argument could be summarized thus: since chains and projections do not exist at the interfaces (the first similarity discussed in this section), chains and projections are suspect from a minimalist perspective. Perhaps the reason chains and projections cannot exist at the interfaces is because such constructs would not be legible by the external systems. Legibility considerations therefore demand that objects never be constructed by the syntax in the first place, if the syntax is optimally designed to meet legibility conditions at the interfaces (a central minimalist tenet). I do not find Hornstein and Seely’s arguments compelling (obviously, for my symmetry of chain and projection argument to go through, chains and projections must exist at some level!), but the fact that the very same condition is used to attack both concepts reinforces the similarity between projections.

---

18 Hornstein and Seely’s argument relies too heavily on a strict reading of Inclusiveness. As I already pointed out in Boeckx 2002 (discussing an earlier version of Seely 2006), if projections (or chains) were banned under Inclusiveness, the mere set resulting from merging A and B (\( \{A, B\} \)) ought to be excluded as well, as the set itself is not part of the lexicon. Clearly, Inclusiveness is not a hard and fast condition, but a guideline prohibiting the use of completely extraneous objects such as traces or bar-levels. If chains and projections are seen as collections (sets) of copies, Inclusiveness is not an issue. Hornstein and Seely’s arguments appear even weaker once chain occurrences and projection levels are collapsed representationally, as in the remerge view on movement (for chains), or in Brody’s (2000, 2003) mirror-theoretic representations of heads as ambiguous/dual minimal-maximal projection points.
and chains (after all, processes violating the same constraint are often unified precisely because they lead to the same kind of violation).

Let me summarize what I have done so far. I have reviewed evidence pointing to the common nature of Merge and Move, or their products: projection and chain. I have shown that chains and projections have become commensurable, permutable, and, indeed, that they can be seen as two sides of the same coin under symmetry transformations. Projections have become micro-chains, and chains, superprojections. Perhaps both chain and projection should be called path, a term that would capture the idea of extending, or stretching an element along portions of a tree.19

The symmetry between projection and chain turns out to have important consequences for our understanding of locality, as I will show in the next section.

2.4 Chains, Projections, and Locality

We saw in Section 2.3.2.1 that chains have the same internal organization as projections. They consist of three hierarchically structured occurrences of the same element, which I characterized as minimal, intermediate, and maximal to emphasize the parallelism with projection levels. Using the trace notation, we can represent it as (30a), and, under the copy notation, as (30b):

\[(30)\]
\[\begin{align*}
a & : (XP, t'_{XP}, t_{XP}) \\
b & : (XP, (XP, (XP)))
\end{align*}\]

We could call this the X-bar schema for chains. I would like to suggest that locality violations could be viewed as amounting to deviations from this schema. That is, island effects may be nothing more than disruptions of the shape of (30). Obviously, the first issue to address in light of this proposal is how many deviations from (30) can there be? (This is basically another way of asking how many distinct island effects there are.) Here the product of Merge/Move

---

19 My choice of the term path is not innocent. Readers familiar with Kayne (1984) will not fail to recognize my attempt to recast Kayne’s insights about path-based locality in a minimalist setting.
symmetry we have explored so far proves extremely useful. If chains really are (at some level of abstraction) projections, we expect bad chains to correspond one-to-one to bad projections, that is, violations of the standard X-bar schema. We are now in familiar territory, for X-bar configurations have been extensively studied, beginning with Jackendoff (1977) and culminating in Kayne (1994). These studies have revealed what makes bad projections.

Consider the configurations in (31)–(35).

\[
\begin{align*}
(31) & \quad *(Z'', X', X^0) \\
(32) & \quad *(X', X'', X^0) \\
(33) & \quad *(X'', X'', X', X^0) \\
(34) & \quad *(X'', X', X', X^0) \\
(35) & \quad *(X'', X', X^0, X^0)
\end{align*}
\]

In tree form:

\[
\begin{align*}
(31') & \quad Z'' \\
(32') & \quad X' \\
(33') & \quad X'' \\
(34') & \quad X'' \\
\end{align*}
\]
If we set aside adjuncts (to which I will return in Chapter 3), it is clear that all the representations in (31)–(35) deviate from the standard X-bar schema. (31) violates endocentricity (as do variants of (31) like *(XP, Z', X°), or *(XP, X', Z°)): no extraneous member (symbolized by Z) is allowed to be part of the projection of X. “Mixed” projections are ruled out. (32) violates the requirement that maximal projections be maximal: they should be on top of intermediate projections (variants of (32) such as *(X°, XP, X'), *(X°, XP, X'), etc. would be excluded by the very definitions of minimal, intermediate, and maximal).

The representations in (33)–(35) are all excluded because they contain multiple instances of the same projection level. Of the three bad projections in (33)–(35), (35) is the least controversial, and is standardly excluded: projections cannot contain multiple heads. Again setting adjuncts aside, we can characterize (33)–(34) as ruling out instances of multiple specifiers within a single projection (depending on whether one chooses to represent specifiers as sister of X’ or daughter of X”—equivalent representations as soon as adjunction is out of the picture). The ban on multiple specifiers was taken for granted until very recently (essentially until the advent of Bare Phrase Structure (Chomsky 1994, 1995)). Empirically, the evidence for multiple specifiers is not overwhelming. Even in the domain where the evidence is strongest (instances of multiple wh-fronting in Slavic languages, for example; see Koizumi 1994; Richards 2001), subtle differences among alleged specifiers suggest that not all specifiers are on an equal footing (see especially Bošković 1999; Boeckx 2003b; Jeong 2006; Krapova and Cinque 2005).20 And alternative analyses that do not rely on multiple specifiers are readily available (see, e.g., Grewendorf 2001).

20 For critiques of the use of multiple specifiers in other domains of the grammar, see Zwart (1997) and Grohmann (2003). As I argue in Boeckx (forthcoming a), the ban
For these reasons, I will continue to assume that multiple specifiers do not exist, and that X-bar relations are unique, with only one type of projection allowed per phrase. As Zwart (1997) notes, multiple specifiers were avoided in GB days for conceptual reasons; specifically, because they violated the following two guiding principles: the unambiguity of structure and the bi-uniqueness of licensing relations.

The unambiguity of structures was most recently defended on various grounds by Hale and Keyser (1993: 66 f.), who built on previous work by Kayne (1984) and Larson (1988).

Hale and Keyser’s discussion is worth quoting in full:

…Each lexical head X determines an unambiguous projection of its category—to a phrasal level, XP—and an unambiguous arrangement of its arguments, as specifier and complement. […] We will speculate further that the unambiguous structure requirement will yield an additional limitation on the projection of categories to types: to wit, the requirement that “intermediate” types (X’) be restricted to just one for any given projection. […] The limitation on types follows, we wager, from the assumption that multiple “intermediate” types would be linguistically (though perhaps not notationally) indistinct.

The bi-uniqueness requirement on licensing was defended most forcefully by Hoekstra (1991), and is standardly incorporated in most versions of “checking theory” in the minimalist program.

I conclude, with much previous work (see especially Kayne 1994), that all the configurations in (31)–(35) are improper projections.

As the above discussion already made clear, improper projections fall into three kinds: they either violate endocentricity, or the rigid hierarchy of projection levels, or the uniqueness of projection level types. To be legitimate, projections must avoid extraneous members, extra (/duplicated) members, and misordered members. Call these conditions on projection Anti-Interference, Anti-Overcrowding, and Anti-Reversal.

on multiple clitics bearing the same feature in clitic clusters known as the Person-Case Constraint (Perlmutter 1971; Kayne 1975; Bonet 1994) is probably another reflex of the uniqueness of X-bar relations.
I will now show that the same requirements hold of chains, emphasizing the symmetry between product of Merge and product of Move. As I have tried to do so far, I will keep to an informal level of discussion, so as to present the intuition in its simplest form. Important details of implementation will be left to subsequent chapters.

The only thing I will be asking of the reader is that she look at the configuration of a chain formed by movement. The reader need not change any of her assumptions regarding how derivations proceed.

The first thing to do to facilitate the discussion and reveal the symmetry is to go back to the X-bar schema for chains in (30) and be clear about the content of “projection” levels in a chain. The most natural way to understand these seems to me to be in terms of checking, or licensing relations. Thus, the chain formed by moving the item what in (36) will consist of a first/minimal licensing relation corresponding to theta-role assignment \( (t^c) \), an intermediate case/agreement-checking relation in embedded SpecTP \( (t') \), and a final/maximal wh-feature-checking relation in matrix SpecCP corresponding to the surface position of what. (Like I did with adjuncts in the representation of projections, I ignore intermediate adjunction steps formed by successive cyclic movement in chains. I will also set aside here chains formed by movement of adjuncts since this would force me to go into issues of adjunction, which I want to delay until Chapter 3.)

(36) What did you say \( t' \) was bought \( t^c \)

The terms minimal, intermediate, and maximal licensing relations for chains are more than terminological convenience. As is well established empirically, argument chains must originate in a theta-position, and cannot proceed beyond a wh- (or more generally, A-bar) checking site.

As is also well established empirically, a uniqueness condition on licensing holds in chains: chains can contain no more than a single theta position (Chomsky’s (1981) theta-criterion), no more than a single Case position (Chomsky’s (1981, 1986a) Chain Condition),
and no more than a single wh-/A-bar position (Epstein 1992; Bošković 2005b, forthcoming; Rizzi’s 2006 “Criterial Freezing”).

Witness (37)–(39).

(37) *John seems [t’ to [t’ like t’]]
(38) *Who [t’ seems [t’ is [t’ happy]]]
(39) *Who did Sue wonder [t’ [t’ was arrested t’]]

There is an obvious sense in which the violations in (37)–(39) correspond to the Anti-Overcrowding requirement discussed in the context of (traditional) projections above. Too much checking of the same kind is going on in (37)–(39).

Anti-Reversal is also at work in chains, corresponding to what is called Improper Movement. Improper Movement was a ban introduced in Chomsky (1973: 242 n. 24) (see also May 1979 and Chomsky 1981: 195–204) to rule out derivations that would disrupt the licensing sequence {Theta ! Case ! Wh} and allow movement to proceed from a theta-position to an A-bar position to an A-position. A typical illustration of the ban on Improper Movement is given in (40).

---

21 I am aware that the ban on multiple theta-assignment and on multiple Case-checking to a single chain is no longer universally accepted (see, among others, Bošković (1994), and Hornstein (1999) on theta-roles, and Yoon (1996) on Case). As in the case of alleged multiple specifier constructions discussed in the text above, subtle differences between seemingly identical relations quickly arise (in the case of multiple Case-checking, one of the cases also functions as a focus/topic-marker, as Yoon himself notes), and there exist alternatives that preserve the uniqueness of the relations at issue (e.g., by relying on null resumptives, as Kayne (2002) does for multiple theta-assignement in control structures, and as Boeckx (2003a) and Moore (1998) do for hyperraising/multiple Case assignments). I return to such considerations in Chapter 3.

22 For recent discussion of Improper Movement, and conclusions that would fit well within the present framework, see Abels (forthcoming). If the approach suggested here is adopted, any account of remnant movement should also tackle the problem of why T is contained in C, and not just focus on why A-bar movement cannot precede A-movement.

23 Clear illustrations of the ban on Improper Movement are hard to construct. By this I mean that Improper Movement examples can often be ruled out on independent
Anti-Interference also obtains in the context of chain, and falls within the rubric of Relativized Minimality effects. Relativized Minimality violations are nothing but cases where an extraneous element disrupts chain formation, just like the “Z”-member disrupted the formation of the “X”-projection in (31).

As we can see, the very requirements constraining projection carry over to chain formation. The symmetry between chain and projection does not stop here, however. So far, I have shown that well-known conditions on chain formation (Improper Movement, Checking Uniqueness, and Relativized Minimality) can be seen as reflexes of Anti-Interference, Anti-Overcrowding, and Anti-Reversal. We expect the same constraints to hold of both movement qua extraction of x, and movement qua extraction from x. This is clearly the null hypothesis: no matter where the moving element originates from, it should be subject to Minimality effects (Anti-Interference), Criterial Freezing effects (Anti-Overcrowding), and Improper Movement (Anti-Reversal). (Note that since the present account collapses conditions on chains and conditions on projections, it collapses Relativized Minimality (a c-command condition) and the A-over-A condition (a dominance condition): both are instances of path interference.25)

grounds (e.g., relativized minimality). Notice also that the example in (40) is not a perfect case, as the A-bar movement step in question is not an A-bar checking site. The reason Improper Movement examples are even hard to construct is probably due to the featural dependence of the Case domain (T) on the C-domain (see Chomsky 2007, forthcoming; Richards 2007). (Thanks to Sam Epstein for discussing this issue with me.)

24 From the present perspective, it makes no difference whether the interfering element could enter into a checking relation with the target of movement (as in superiority effects) (i), or not (as in wh-island contexts) (ii). Accordingly, the present approach does not distinguish between standard instances of minimality (i), and cases of so-called “defective intervention” effects (Chomsky 2000) (ii).

(i) *(guess) what who bought t_{what}
   (cf. (guess) who t_{who} bought what)

(ii) *who did you wonder when Mary kissed t_{who}
    (cf. *when did you wonder t_{when} Mary kissed who)

25 For independent reasons for collapsing Relativized Minimality and the A-over-A condition, see Fukui (1997) and Hornstein (forthcoming).
Although locality conditions on extraction and subextraction are the same from the present perspective, it turns out that, due to the way chain members are defined here, subextraction is quite severely constrained, as I will now show.

Recall that the members of a given chain are defined in terms of licensing relations. The various members of a chain can therefore be associated with (identified by) their licensers. This turns out to be very significant in the context of subextraction, for it means that the point of origin (X/τ) of a chain formed by subextraction will be associated with (identified by) the (head of the) domain from which subextraction takes place. (The final/maximal member of a chain will be identified by the final landing site of the element undergoing subextraction. Since I ignore intermediate steps of movement, the intermediate checking site formed by subextraction may be equated with a “Case”-checking site for the subextractee, as in (a modern reading of) Cinque 1980; see also Ormazabal 1991 and Ticio 2003, 2005.26)

Accordingly, a chain formed by subextraction will inherit some of the properties of the domain/element out of which extraction is taking place. Specifically, since a proper argument chain must start in a theta-position, and the starting point of a subextraction chain is

---

26 Since many instances of subextraction take place out of nominal domains, and Case assignment within nominals is not as well understood as in the sentential domain, intermediate chain members may be hard to identify in some cases. Notice that just like there exist “that-trace” effects for intermediate traces in the context of extraction of x, there appear to be “that-trace” effects for intermediate traces on the context of extraction from x. Such “that-trace” effects—in the guise of “Determiner-trace”—have been argued to give rise to Left-Branch Condition and Definiteness/Specificity effects. (See Uriagereka 1988; Corver 1992; Gavruseva 2000; Boeckx 2003a; and Bošković 2005a.) Thus, the presence of an overt determiner in Chamorro (iib), or of a demonstrative determiner in English (iib) renders subextraction impossible, much like the presence of an overt complementizer does (iiib).

(i) a. Hayi un-ladatdi [τ' patgon-ña τ'] [Chamorro] who agr-scold child-agr ‘Whose child had you scolded?’

b. *Hayi un-li'i [i τ' gā̱'-ña ga'lagu τ'] [i τ' gā̱'-ña ga'lagu τ']
who agr-see the pet-agr dog ‘Whose dog did you see?’

(data from Chung 1998: 282, 286)
identified by the extraction domain, it follows that subextraction chains will only be licit if the extraction domain is thematic. This immediately rules out subextraction from displaced domains (arguments having moved out of the thematic domain), and subextraction from adjunct domains.

(i) a. Who did you see [t' pictures of t']
   b. *Who did you see [these t' pictures of t']

(ii) a. Who did you say [t' was arrested t']
    b. *Who did you say [that t' was arrested t']

The formulation in the text is fully compatible with instances of so-called remnant movement (on which see, especially, Mueller 1998), of the type illustrated in (i).

(i) [How likely to tjohn win]i is John t

Remnant movement is a process that moves the domain out of which subextraction takes place beyond the final landing site of the subextracted element. Crucially, at the point when subextraction takes place (at the point when the subextraction chain is defined), the domain out of which subextraction takes place is thematic.

I assume that any displaced domain constitutes an island—generalizing from the “subject condition” to the freezing condition (Ross 1967; Huybregts 1976; Wexler and Culicover 1980; Takahashi 1994; Ormazabal, Uriagereka, and Uribe-Etxebarria 1994; Boeckx 2003a). I return to apparent counterexamples to the freezing condition in Chapter 6. For relevant discussion, see also Gallego (2007), and Gallego and Uriagereka (2007b).

As I argued in Boeckx (2003a: 100–1), the so-called “adjunct condition” subsumes the ban on extraction out of adjuncts, the ban on extraction out of relative clauses, which are standardly taken to be adjuncts, and the ban on extraction out of nominal complements, which, following Stowell (1981), I take to be appositives. (Sentential subjects may also be part of the list if they are treated as “satellites” (Koster 1978a). However, there are reasons to believe that they are genuine arguments, much as nominal subjects. See Bošković 1995.) The various bans just listed are illustrated in the order they were mentioned.

(i) *Who did John arrive [after Bill kissed t]
(ii) *Who did John meet [the woman [that said that Bill kissed t]]
(iii) *Who did John listen to [rumors [that Peter kissed t]]

The consequences of the inertness for adjuncts extend to the paradigm discussed by Zwicky (1971) and Kayne (1984), who observe that verbs of manner of saying are impervious to movement.

(iv) *Who did John grunt [that Mary likes t]

Following Stepanov (2001), who in turn follows Snyder (1992), I analyze such verbs to consist of a (hidden) complex NP. More precisely, an NP to which a CP is apposed, as schematized in (94).

(v) John grunted that Mary left = John gave [a grunt [that Mary left]]

As originally observed by Zwicky, almost all verbs of saying have homonymous nouns. For example, to holler corresponds to a holler, to grunt to a grunt. Zwicky also points out
In effect, the subextraction/extraction domain symmetry under discussion captures Huang’s Condition on Extraction Domain (CED) mentioned in Chapter 1. The contrast between (41a) and (41b,c) follows at once.

(41)  
   a. *Who did you see [pictures of e']?  
       Domain is $\theta \rightarrow \sqrt{e'}$
   b. Who did [pictures of e'] cause Bill to fight  
       Domain is $\theta \leftrightarrow e'$
   c. *Who did Bill laugh [after Mary saw e']?  
       Domain is $\theta \leftrightarrow e'$

Note that the present analysis of CED-effects does not collapse displaced elements and adjuncts, it simply says that subextraction out of either domain gives rise to an improper chain. As such, it leaves open the possibility that something about displaced elements or about adjuncts (processing factors) may render the locality violation more severe in one case than in the other. It is therefore fully compatible with “asymmetric” CED-effects of the type discussed in Sprouse (2007).

2.5 Summary

Let me summarize what I have done in this chapter. The aim was relatively modest. It was basically an attempt to motivate a framework for addressing the issue of locality. Taking as my point of

that given this striking symmetry between verbs of saying and their corresponding nouns, it is always possible to paraphrase any of these verbs with a phrase like give a N, e.g., give a grunt. It is thus plausible to assume, as does Stepanov, that manner of speaking verbs have an underlying structure that involves a corresponding noun. This suggestion is in line with the approach put forth by Hale and Keyser (1993).

A similar reasoning applies to languages like Polish (and many other languages), where extraction out of finite complement clauses is prohibited.

(vi) *Kogo ty wiesz [ze Janek lubi t]  
    whom you know that Janek loves
    ‘Who do you know that Janek loves?’

To rule out such cases, I assume that indicative complements in such languages function as appositives to a (sometimes null) correlative pronoun, along the lines proposed by Torrego and Uriagereka (1993), among others, and schematized in (vii).  
(For Polish, Gieigo (1981) shows that the correlative pronoun to can be overt.)

(vii) know [that …] = know [N IT [CP that …]]

The structure in (vii) essentially turns indicative complements into adjuncts.
departure the symmetry between External Merge and Internal Merge (movement), I proceeded to argue in favor of an equivalence between the product of External Merge (projection) and the product of Internal Merge (chain). I have shown that chain and projection are very similar objects. They are commensurable, interchangeable, and one can easily map one onto the other via symmetry transformations.

Once projection and chain are seen as symmetric objects, the nature of the locality issue discussed in Chapter 1 changes significantly. Instead of treating chains as separate entities being subject to very specific locality conditions, it becomes possible to entertain the possibility that locality conditions on chains are equivalent to the locality conditions on projection. I have sketched a possible way in which this may be so. But I stress that I haven’t even begun to explain these locality conditions. To do so requires a careful examination of Merge and its product, projection. Because Merge and projection are so basic, my hope is that it will be easier to make sense of (locality) conditions in this domain. Once we do that, the symmetry between chain and projection argued for here will enable us to generalize our explanation of the locality of Merge to Move. In a certain sense, there will be no special theory of locality for movement. (The present chapter said nothing about locality of selection, but the issue here is clear: just like I hope to deduce the locality of long-distance dependencies from conditions on Merge by treating chains as projections, I also hope to deduce the locality of selection from conditions on Merge by treating “cartographies” as projections. This is essentially the direction suggested by Grimshaw 1991, who viewed functional sequences as extended projections—although special attention will have to be paid to the intervention situations discussed at the end of Chapter 1.)

As far as I can see, the major explanatory challenge ahead lies in reformulating the conditions on projection we found in terms of a more minimalist theory of phrase structure. All the conditions I made use of in this chapter are couched in X-bar terms. To make perfect sense of them, we would like to deduce them from considerations of computational efficiency (“economy”) and interface legibility conditions (“bare output” conditions). This is no easy
task, given that recent developments in the realm of phrase structure, starting with Chomsky’s (1994, 1995) bare phrase structure (reviewed in the next chapter), have questioned central tenets of X-bar theory, to the point that some have argued that projections should be dispensed with (see Collins 2002; Seely 2006; Bowers 2001; Collins and Ura 2001). Since projection is so central to the argumentation in this chapter, I will have to find a way to recapture this concept, and all the other concepts needed (such as uniqueness of specifiers) in the context of a bare theory of syntax—whence the title of this volume.
Part II
This page intentionally left blank
3

Unambiguous Merge

3.1 The nature of syntax

At the most basic level of description natural language syntax can be characterized as an interface system, providing the meeting ground between “sound/sign” (more accurately, the mental systems responsible for externalization, henceforth PHON) and “meaning” (the mental systems giving rise to thought, hereafter SEM). Natural language syntax operates on units that are standardly characterized as bundles of features. Such features are lexicalized concepts. Syntax creates ever-larger molecules by combining featural atoms through iterated use of Merge. Such molecules, the expressions generated by syntax, provide instructions to PHON and SEM.

I doubt that it will be regarded as controversial if I say that a well-designed interface system should provide instructions to PHON and SEM in a format that is usable for them. Chomsky (2000: 94) calls these legibility requirements. Thus, syntax is often said to provide a Logical Form (LF) representation to SEM and a Phonetic Form (PF) to PHON. In addition to being readable by PHON and SEM, such instructions coming from a well-designed system should be unambiguous and non-contradictory. Instructions meeting these basic design requirements should ensure an efficient mapping, and fast processing of the data (what is known in the image-processing literature as “feature extraction”). An optimally designed interface system should consist of nothing more than this: a minimal set of operations (ideally, just one) generating expressions that facilitate
feature extraction by the external systems, minimizing resources required for processing data structures. The minimalist program for linguistic theory is an attempt to determine how closely natural language syntax approximates such an ideal of architectural design. It stands to reason that looking at syntax alone won’t do in this context. A significant part of such a minimalist enterprise will consist in determining the shape of PHON and SEM, the format(s) of the information they can handle, their processing resources, etc. Another important aspect of the minimalist program will deal with the content of the lexicon, and the format of lexicalized concepts, since the type of molecules formed by syntax and transferred to PHON and SEM will depend in large part on the types of features that syntax has at its disposal. Accordingly, determining the nature of syntax is akin to solving an equation with three unknowns (the minimalist would contend that the equation requires no other term, and can be solved in the simplest fashion):

\[
\text{Syntax} = \text{Merge} + x_{\text{Lex}} + y_{\text{LF}} + z_{\text{PF}}
\]

This could be represented schematically as in (1).

(1)

The above remarks are meant to highlight the fact that minimalist inquiry must necessarily address issues touching on mental components traditionally falling outside the purview of studies focusing on “narrow syntax.” And because we know significantly less about such external systems, the number of avenues to explore is very large, and the conclusions tentative. However, to the extent that the initial settings of inquiry yield results that we have independent reasons to believe are true of narrow syntax, they can themselves
be legitimized, and parts of the nature of the external systems (Conceptual repertoire, SEM, and PHON) thereby revealed. It should be said that although the problem at hand is vast, it is no different from the task faced by the experimenter in her attempt to determine any property of the natural world. It is the defining feature of empirical science. (The minimalist challenge consists in showing that this is all that we need.) It should also be borne in mind that success of the research at hand does not depend on figuring out all the properties of the lexicon, PHON and SEM—only those properties relevant to the interface that syntax must establish (what Chomsky (1995: 221) calls the “bare output conditions”). And although we know significantly less in these areas, we are not completely in the dark. As we will see momentarily, I will be able to rely on a rich literature that has identified important aspects of the systems surrounding syntax. As a matter of fact, very little will have to be added to existing proposals (apart from combining them all together) to achieve the desired results.

Within this general context of inquiry, the present chapter explores an avenue of research that seeks to deduce the X-bar properties discussed in the previous chapter, which I have argued provide the basis for understanding locality. Ideally, such X-bar properties should fall out from legibility conditions imposed by PHON and SEM and/or properties of the features syntax is forced to work with.

I will begin by stating as clearly as I can what I take to be the basic properties of SEM and PHON that syntax must meet. Then I will turn to the features that syntax operates on. And, finally, I will show how these properties and these features go a long way, perhaps all the way, toward determining the form of Merge (resulting in projection). I should point out that it will be impossible for me to do justice to the richness of the literature I draw from. Accordingly, rather than illustrating the basic ideas with toy examples that would inevitably oversimplify things and thus beg even more questions, I will simply state the results I will rely on, and urge the reader to consult the works cited. Accessible sources for the various sections will be given in the notes.
3.2 Minimal Interface Requirements

3.2.1 PHON

Perhaps the easiest place to start is PHON. At some point, perhaps at various points, in the course of the syntactic derivation, material is transferred to the P(honetic) F(orm)-component. This is the operation Spell-Out in Chomsky (1993). Three major things happen post-spell-out. First, syntactic hierarchies are flattened; syntactic trees must be linearized. Syntactic objects are at least two-dimensional: there is the horizontal dimension defined by combining $\alpha$ and $\beta$, and the vertical dimension defined by projection, as represented in (2).

The physics of speech, however, cannot entertain the possibility of 2-D representations. Each unit (morpheme, word, etc.) must be pronounced one after the other. Speech is rigidly one-dimensional. Although this fundamental difference between syntactic and PF-representations was already stressed in Higginbotham (1983), it wasn’t until Kayne (1994) that the importance of a linearization algorithm was recognized. Kayne proposed such an algorithm based on his Linear Correspondence Axiom (LCA), which says that asymmetric c-command maps onto precedence. Chomsky (1995) was the first to explicitly suggest we view the LCA as a property of the PF-mapping, perhaps part of the operation

---

1 As in Chomsky’s (2000, 2001) phase-based model, or Uriagereka’s (1999a) multiple spell-out approach.

2 It is an open question whether the same requirement holds for sign languages. Because the characterization of the syntax-PF interface in sign language studies I am familiar with appears to be isomorphic with the one assumed for spoken languages, I will tentatively assume that the linearization requirement holds of all languages, irrespective of their modality.
Spell-Out itself (Chomsky 2004). Epstein et al. (1998), Uriagereka (1999a), Fox and Pesetsky (2005), and Richards (2004) explore various ways of making linearization take place cyclically—all based on Kayne’s original algorithm. These works deal with linearization of (standard) projections. On linearization of products of movement (chains) that also rely on Kayne’s LCA, see Nunes (2004), Hornstein (2001), and Bošković and Nunes (2007).

In addition to assigning a 1-D representation to syntactic objects, PF must also assign a prosodic structure to them, responsible for intonational properties of sentences. This is the process of intonational phrasing/stress assignment at the sentential level. That prosody reflects aspects of the hierarchical syntactic representation of an expression has long been recognized (see already Chomsky and Halle 1968). What remains controversial is the extent to which it is necessary to refer to syntax in stating generalizations about prosody, and to what extend the hierarchical organization of prosody replicates hierarchy in syntax. I do not intend to review this vast literature here (see Elordieta 2007). I will simply assume Wagner’s (2005, 2007) model, which provides an algorithm for converting syntactic representations onto metrical grids in a cyclic fashion—a compositional mapping very much like the one standardly assumed at the syntax-semantics interface. Wagner’s central idea is that prosodic phrasing can be derived from the way syntactic derivations work by the simple assumption that the output of cycles gets mapped to a foot in prosodic representations. Call this the Prosodic Correspondence Axiom. (Note that because it crucially reflects syntactic hierarchies, prosodic mapping must precede linearization.)

I find Wagner’s approach particularly attractive because, much like Kayne’s LCA, it allows us to maintain an isomorphism between syntactic representations and PF-representations. Wagner’s approach in fact assumes that prosodic structure building requires nothing more than basic operations like Concatenate, Project, and Embed—all of which are part of the standard definition of Merge (combine + project)—to derive effects of flat prosody like \((A \& \& B)\)

---

3 For discussions of the basics of linearization under Kayne’s LCA, see Cinque (1996) and especially Hornstein, Nunes, and Grohmann (2006: ch. 7).
Wagner shows that, contrary to what is often assumed, syntax is quite in tune with prosody; mismatches are only apparent. Put differently, the prosody of an utterance reflects syntax to a much higher degree, and in a much more direct fashion than previously thought. As a matter of fact, Wagner’s approach recaptures Liberman’s (1975: 258) insight that the metrical grid is not a separate representation, completely independent from syntax, but just another way of representing syntactic information (see also Steedman 2006 and Vergnaud 2003). In Liberman’s words,

Thus the basic assumptions of our theory depend on the idea that the phonological component is not so much a destruction of structure, which maps a complex tree onto a simple serial ordering of segments, as a transmutation of structure, which maps a structure suitable for operations in one domain (syntax and semantics) onto a structure suitable for use in another domain (the motor control of articulatory gestures, and its perceptual analogue).

Let me close this brief discussion of prosody by saying that just like chains and projections appear to be organized around an X-bar schema, so do metrical grids. Phonologists have long recognized the need to distinguish three stress-relations: stressed vs unstressed, primary stress, and secondary stress. These correspond rather straightforwardly to head (minimal projection) vs non-head, intermediate, and maximal projection levels. Prosodic structure may then be seen as a reflex of the product of Merge. The third operation that must take place at the syntax-PF interface is vocabulary insertion. Here I follow much recent work in realizational morphology (Halle and Marantz’s (1993) Distributed Morphology model; Borer’s (1998, 2005), and Di Sciullo’s (2005) Parallel Morphologies; Beard’s (1995) Lexeme-Morpheme-based morphology). The central assumption in realizational morphology is that there is no specific syntax of words. In fact, words don’t exist, they are the phlogiston of linguistics. It’s (the same) syntax all the way down to morphemes (bundles of features). In Marantz’s (1997) terms, “there is no escape

4 For actual examples and derivations, see Wagner (2007). For a more comprehensive overview, see Wagner (2005).
from syntax.” (For a particularly strong attack on the notion of “word,” see Julien 2002, 2007.)

It should be obvious that realization morphology is the null hypothesis. Everyone must say that sentences are built out of morphemes. Lexicalism (the alternative to realizational morphology) says that sentences are built out of morphemes, but only indirectly: words intervene between morphemes and phrases, and have their own unique properties. In realizational morphology, pieces of inflection are ornamental, they are like bits of flesh competing to attach around syntactic bones. They do not in any way constrain the syntax; they simply “spell-out” syntactic structures. We therefore expect a high degree of isomorphism, of the type expressed by Baker’s (1985) Mirror generalization; although various affixal properties may lead to readjustment rules that end up masking syntactic structures; see, e.g., Bošković (2001) and Embick and Noyer (2001). The extent of such readjustment rules enriches the morphological component of the grammar, and makes it difficult to order vocabulary insertion with respect to linearization. Some (see Nunes 2004, building on a proposal by Chomsky 1995) assume that the structure of words is opaque to the workings of the LCA (i.e., words are built before the LCA applies); others assume that the LCA can look inside words. Because word formation (vocabulary item selection) is sensitive to various linear adjacency conditions (cf. Chomsky’s 1957 do-support operation, revived by Bobaljik 1995, and Lasnik 1999a, 2000, 2003), I will tentatively assume that vocabulary insertion takes place after linearization, although I stress that a more complex organization of the morphological component in terms of cyclic

---

5 For a clear overview of Distributed Morphology, the dominant paradigm in realization morphology, see Embick and Noyer (2007).

6 On the mechanisms of competition, see Halle (1997) and Caha (2007).

7 As I discuss in Boeckx (2007b, in progress), realizational morphology models strengthen Chomsky’s Uniformity hypothesis (cf. (3) in Chapter 1), since they essentially render syntax immune to most parametric effects if these are seen as part of lexical items.

8 In my view, the strongest kind of argument for a strictly realizational morphology comes from Di Sciullo’s (2005) extensive demonstration that morphology, unlike syntax, does not show signs of symmetry(-breaking). Its structure and organization is purely asymmetric, much like linear order.
rules and post-cyclic rules may prove empirically necessary, and wouldn’t affect the core arguments of the present work.

Since I just mentioned cyclic rules, it is interesting to note that in recent work, Marantz (2000, 2006) has advocated a cyclic approach to “word” formation within Distributed Morphology. The model meshes very well with approaches that assume cyclic linearization (see references above) and Wagner’s cyclic prosodic mapping. These three cyclic mapping hypotheses concerning linear order, word structure, and prosodic structure ensure maximal transparency between syntax and PF. It may well be that various readjustment rules (“interface strategies,” to borrow Reinhart’s 2006 term) may be needed, but already now it seems safe to conclude that the basic PF-skeleton is determined by the syntax on the basis of simple mapping algorithms (essentially, strict correspondence rules). Taken together, the results of the three approaches adopted here argue strongly for a syntactocentric organization of the PF-side of the grammar (contra Jackendoff 1997, 2002; Culicover and Jackendoff 2005). When we turn to the LF-side of the grammar, and the mapping from syntax to SEM in the next section, we will see that syntactocentrism is even stronger in this domain.

To sum up this section, syntactic representations on their way to PHON (the external systems responsible for the externalization of syntax) must be mapped onto prosodic structure, word structure, and linear order. Following Wagner, I assume that prosodic structure is essentially isomorphic with syntactic structure. Word structure (vocabulary insertion) is an embellishment of linearized syntax. The only significant change to syntactic structures taking place in the PF-component is thus the flattening of 2-D syntactic objects onto a 1-D, linear sequence.

Let me stress again that these are the bare minimum that PHON requires of syntax. It is extremely unlikely that syntax will be allowed to get away with less. The works I have relied on in this section claim that these are sufficient to derive the various facts and generalizations that have been brought to bear on the issue of the syntax/PHON interface. I will assume that they are basically correct, but let me stress that even if it turns out that more is needed, at least
nothing more will be needed to derive the properties of Merge that I need for the argument sketched in Chapter 2 to go through.

3.2.2 SEM

Although no one denies the existence of a mapping between syntax and SEM (products of syntax enter into thought and action), the nature of SEM is more obscure, and consequently, the format required by SEM is more controversial than in the case of PF.

Here I will adopt Pietroksí’s hypothesis, developed in a number of works (Pietroski 2002, 2005, 2006), that the mapping from syntax to SEM is much more homogenous than standardly assumed. It is almost universally assumed in the semantics literature that the meaning of an expression is compositional: it depends on its constituent words and how they are combined. But it is assumed in most semantic models that there are several modes of combination, hence distinct contributions to SEM.9

Pietroski entertains the minimalist (and, I’d argue, null) hypothesis that the semantic contribution of combining expressions is simple and uniform across constructions.10 According to him, each and every complex expression of natural language is the concatenation of two simpler expressions. These two constituents, together with the meaning of concatenation, determine the meaning of the complex expression. For Pietroski, all constituents are understood as monadic predicates, and concatenation always signifies conjunction. Accordingly, from a semantic perspective, every complex expression is a conjunction of predicates. Put differently, all of semantics boils down to cases like “red ball”: a red ball is something that is both “red” and “ball.”

---


10 Pietroski’s view contrasts with Chomsky’s (2004: 118) claim that there exists a strong interface condition at SEM that requires “sufficient diversity at SEM.” If Pietroski is right, the intrinsic expressive power of SEM is much more limited than Chomsky grants to it. For a different argument in favor of SEM—“downsizing,” see Hinzen (2006, 2007).
Pietroski departs from most semantic models, which take expressions to be associated with functions, or with some elements in the domain of these functions (think of a Verb Phrase consisting of a Verb requiring one argument, and of an argument satisfying the Verb’s requirement).

Crucial to Pietroski’s enterprise is a (neo-)Davidsonian eventish semantics, which provides a model to analyze complex expressions such as causative constructions like *Brutus stabbed Caesar with a knife* in terms of predicate conjunction, amounting to (in the case at hand) something like “there was an event of stabbing, the event involved Brutus as the executor, and Caesar as the victim, and the event involved the use of a knife. More formally:

\[
\exists e [\text{Stab} (e) \& \text{Agent} (\text{Brutus}, e) \& \text{Patient} (\text{Caesar}, e) \& \text{With} (\text{knife}, e)]
\]

As Pietroski (2003) has noted, event representations like (3) mesh very well with current minimalist representations of the thematic domain (Verb Phrase) in terms of Larsonian shells. For instance, (3) translates straightforwardly into a tree like (4), with a Root (\(\sqrt{\text{Stab}}\)) element that gets categorized (i.e., modified) by functional heads introducing the internal argument/patient (V), the external argument/agent (v), and the applied argument/quasi-adjunct (vAppl).

\[
(4)
\]

\[\text{Brutus} \quad \downarrow v^o \]
\[\text{with-a-knife} \quad vAppl^o \]
\[\text{Caesar} \quad V^o \quad \sqrt{\text{Stab}}\]

In many ways, Pietroski’s analyses can be seen as generalized versions of Davidson’s (1967a) original analysis of causatives.
Crucial to Pietroski’s project as well is his rejection of the idea that a theory of meaning (for natural languages) is a theory of truth (see Pietroski 2005: 248; and Pietroski forthcoming b). Pietroski correctly points out that the Meaning-Theory-as-Truth-Theory (at the heart of Davidson’s 1967b program, and Montagovian semantics) ends up relying on either (i) an unduly superficial conception of semantics, or else (ii) an unduly truth-focused and implausible conception of grammar, according to which every factor relevant to the truth or falsity of a sentential utterance is tracked by some aspect of the relevant sentence. Pietroski is right to point out that abandoning a truth-theoretic approach to natural language semantics allows semanticists to set aside many (messy) facts about the truth conditions of utterances, and in so doing (and most importantly for us), many apparent difficulties for simple hypotheses about semantic composition.

I regard Pietroski’s hypothesis as the default characterization of natural language semantics. Though controversial, it makes a lot of sense from the point of view of the minimalist program. If Merge is the only combinatorial operation of the grammar, it makes perfect sense to assume that its contribution to meaning is constant. This is Pietroski’s Conjunctivism. And it makes perfect sense to take adjunction (predicate concatenation) to be the simplest kind of operation there is. We need not look any further to find an operation that is syntactically recursive, and semantically compositional.

From the present perspective, Pietroski’s claim that semantic representations are homogenous and uniformly concatenative appears to be the LF-equivalent of Kayne’s LCA. Just like all syntactic formatives must be mapped onto a succession at PF (linearization), all syntactic formatives must be serialized at LF, forming a string of conjuncts. Complex syntactic objects will therefore require dimension reduction (forming a complex *monadic* predicate) for purposes of LF-integration, just like complex specifiers in Uriagereka’s (1999a) system had to be “atomized” for PF-integration.

Elsewhere (Pietroski forthcoming a), Pietroski has discussed this “monadicizing” process from another angle, focusing on the nature of the lexicon, to which I now turn.
3.2.3 The lexicon

In addition to interfacing with SEM and PHON, syntax feeds on our human conceptual repertoire, via the lexicon. The lexicon provides syntax with the units to operate on. In this sense, syntax is constrained by the nature of lexical representations. The term “lexicon” is here meant in its narrow sense, as the repository of grammatical formatives. Lexicon is sometimes used in a broader sense, which includes the narrow lexicon, but in addition contains information about usage and information about pronunciation. I here follow Distributed Morphologists in distinguishing the narrow lexicon from the encyclopedia (the repository of information about usage) and the dictionary (the repository of information about pronunciation). On this view, a lexical item in the broad sense is nothing but a highly specific interface rule, a minimal idiom, combining information from the narrow lexicon, the encyclopedia, and the dictionary (see Marantz 1997, 2000; see also Jackendoff 1997, 2002). In other words, the broad lexicon is distributed over various components of the grammar. My main reason for distinguishing the narrow lexicon from the broad lexicon is to highlight the fact that the lexicon narrowly construed is not the “repository of idiosyncracies,” as the lexicon is often characterized (see Chomsky 1995: 235). Such a “Bloomfieldian” view of the lexicon (see Bloomfield 1933) fits the lexicon broadly construed: an idiosyncratic, to a large extent, arbitrary linking of sound and meaning; something that must be learned. The narrow lexicon strikes me as much more well behaved. Although not completely predictable in the absence of relevant exposure (due to the existence of specific lexical parameters pertaining to the presence and richness of lexical items, of the type discussed in Boeckx 2007b, in progress), the narrow lexicon is very systematic in character. This systematicity led to the investigation of “l(exical)-syntax”—the isolation of universal, law-like patterns of lexicalization of the type pioneered by Hale and Keyser (1993, 2002), and pursued by Marantz (2000, 2006), Borer (2005), Baker (1997, 2001b), Ramchand (forthcoming), Travis (forthcoming), and Reinhart (2002), among others.
Such works show how properties that used to be encoded by brute force into lexical entries, such as the argument structure of lexical items like verbs, can actually be made to emerge from the syntax. Perhaps the clearest expression of this move toward an “impoverished lexicon” (a bare theory of lexical entries, as it were) comes from the development of a bare phrase structure (Chomsky 1994, 1995)—a matter of utmost importance for the present study.

Any adequate theory of phrase structure will have to capture the fact that there exist three levels of projection: minimal, intermediate, and maximal. The issue that Chomsky addresses in bare phrase structure is how such projection levels should be understood. As Hornstein, Nunes, and Grohmann (2006: ch. 6) point out, there are basically two ways to do it: a rigid way, and a flexible, or relational way. The “rigid” way of conceptualizing projection levels would be to claim that they have different intrinsic features, the way, say, Nouns and Verbs have. That is to say, under this view, projection levels are categorical features, part of lexical entries. The alternative, “relational” view advocated by Chomsky (see already Speas 1990), who builds on Muysken (1982), is to simplify lexical entries, and say that projection levels are “extrinsically” defined.

Specifically, Chomsky claims that projection levels could be defined as follows:

(5) a. Minimal Projection (X^\text{0})
   A minimal projection is a lexical item selected from the lexicon
b. Maximal Projection (X^{''})
   A maximal projection is a lexical item that doesn’t project any further
c. Intermediate Projection (X')
   An intermediate projection is a lexical item whose status is neither minimal nor maximal

The relational view on bar-levels immediately eliminates “spurious” or “vacuous” projections that were standard under the rigid view. Under the relational view, unless a lexical item combines with another lexical item, there is no way to tell whether we are dealing with a minimal or maximal projection.\footnote{11 The absence of an intrinsic minimal/maximal distinction turns out to have desirable empirical consequences; see Boeckx (2006a: 175ff.).}
Under the relational view, projection levels are not lexical features, they are ways of looking at syntactic representations—they emerge as the syntactic derivation unfolds.\textsuperscript{12} A similar logic was applied by Hale and Keyser (1993, 2002) to the domain of thematic roles. Hale and Keyser wanted to find a reason for why the number of thematic relations is so small (a verb can only have a very limited number of arguments—two, maybe three, but not more). One reason could be because theta-features are few in number, but natural, i.e., non-stipulative restrictions in this domain are virtually non-existent. Another, favored by Hale and Keyser, is because the number of licit syntactic configurations available for thematic licensing is very small. In other words, thematic relations are few because syntactic representations are severely constrained. Once this move is made, it becomes possible to eliminate thematic requirements from lexical entries, and let thematic roles “emerge” from syntactic representations the same way projection levels do in bare phrase structure. In other words, the range of thematic relations boils down to the X-bar schema (see Baker 1997: 122). Put differently, thematic relations may be nothing other than the way SEM “reads off” syntactic relations like Specifier of VP or Complement of V. Paraphrasing Baker (1997: 126), X-bar structure and argument structure may be the same representation seen from different perspectives (as Wagner, following Liberman, claims for metrical structure, as discussed above).\textsuperscript{13}

The research program initiated by Chomsky and Hale and Keyser allows for a drastic simplification of what a lexical entry must contain for lexical items to function in the syntax—even more so if Marantz (2000, 2006), and Borer (2005) are correct in claiming that categorial information is contextually determined. According to Marantz and

\textsuperscript{12} Notice that projection levels thus construed do not violate the Inclusiveness condition (cf. Chapter 2), as the syntax makes no reference to them: the syntax of X’s and XP’s is the same, but the information can be accessed by the external systems, where the X’/XP distinction makes a difference (see, e.g., Wagner 2005, 2007; Kayne 1994).

\textsuperscript{13} On the emergence of rich semantic relations from (bare) phrase structure, see Hinzen (2007); Fortuny (2006); Gallego (2007); Uriagereka (2002, forthcoming). See also Chapter 4.
Borer, *race* is not listed in our mental lexicon as +N(oun) or +V(eral). It is a category-neutral Root (√/race), which is verbal (i.e., behaves verbally) if it merges with a functional head like ν or T, nominal if it merges with a functional head like n or D.

In this sense, minimalist syntax is much less dependent on the lexicon than previous models (such as GB) were. In GB, the range of lexical features was very rich (one had to resort to lexical features to define empty categories, reflexives, pronouns, etc.). Within minimalism, an attempt is being made to let these properties emerge from bare phrase structure and the way these are interpreted by the interfaces.

The major lexical distinction that survives this minimalist drive is the distinction between open-class items (“lexical categories”) and closed-class items (“functional categories”). Open-class items are now analyzed as roots; closed-class items as bundles of features that enter into checking relations. Currently, checking relations amount to a process of valuation (Agree) (see Chomsky 2000). Agree is a process whereby the lexical item L called the Probe finds a lexical item L’ (the Goal) with a matching feature F whose value is lexically fixed. Valuation of F on L happens under feature-sharing.14 The reader will have noted that Agree is an operation distinct from Merge. It operates at the sublexical level (at the level of individual features), and for this reason does not require lexical contiguity/adjacency. In other words, Agree can take place at a distance.15 The interplay between Agree and Merge will be a major concern in subsequent sections of this chapter. Let me close this section by pointing out that the view of the lexicon developed within minimalism leads to a rather natural characterization of the process of lexicalization, particularly the lexicalization of roots, as monadicization. Following Pietroski (forthcoming a), one could say that when a concept is lexicalized, it becomes a monadic (or simplex)

---

14 For various formulations of Agree as feature-sharing, see Frampton and Gutmann (2000), Pesetsky and Torrego (2007), Boeckx (2003b). See also Shieber (1986).
15 On Long-distance Agree, see Boeckx (2007c, forthcoming b); Boeckx and Niinuma (2004); Polinsky and Potsdam (2001); Bhatt (2005); Soltan (2007); Etxepare (2007); Wurmbrand (2006).
concept. For instance, the dyadic concept CHASE(x, y) is lexicalized as $\sqrt{\text{chase}(_{\_})}$. The empty slot of the root will allow the lexical item to be integrated into an event structure (i.e., the empty slot will be filled by e). One could think of this empty slot as Chomsky’s (forthcoming: 6) edge-property: the basic property of a lexical item that allows it to enter into a Merge relation.

Pietroski’s suggestion amounts to saying that much of the content of a concept becomes inaccessible at the stage of lexicalization (which fits well with today’s most compelling view of concepts as “hard” atoms; see Fodor (1998)). Lexicalization imposes a unique format (monadic predicate), causing many concepts to lose their original adicity. This view of things fits very well with Pietroski’s conjunctivist project discussed in Section 3.2.2, and may provide insights into the character of natural language semantics. As Pietroski notes, by imposing a unique format onto all concepts, lexicalization allows lexical items to combine in virtue of their shared format, as opposed to a less “syntactic”/formal criterion such as semantic/conceptual affinity. As Pietroski speculates, this would allow for concepts originally residing in distinct mental modules and therefore opaque to one another, to combine and form new concepts (through predicate conjunction, and SEM-mapping). It is quite possible that what is at first a formal restriction on lexical items is the source of a cross-modular syntax of thought—giving rise to a full-blown language of thought, arguably the source of our Great (mental) Leap Forward at the evolutionary scale.

---

16 I here depart from Chomsky, who uses the term edge-feature. I’d rather avoid the use of the term feature, since, as Chomsky makes clear, the edge-“feature” is unlike other features that enter into checking relations: it does not “Agree”/Value, it does not “erase,” etc.

17 Accordingly, lexicalization, like vocabulary insertion under Distributed Morphology, flatly violates Inclusiveness, which is to say that the latter holds only of narrow syntax (see also Chomsky 2000: 118).

18 For speculations along similar lines, see Hinzen (2006), and especially (2007).

19 It has often been noted that modern humans distinguish themselves from other animals by their use of a “floodlight” (i.e., cross-modular, as opposed to laser-beam-like, i.e., highly modular) intelligence. See, e.g., Hauser (2000); Spelke (2003).
In other words, by being able to resort to a common lexical format, post-lexical thought (characterized as the conceptual molecules created by syntax and mapped onto SEM) is more varied, and more powerful than pre-lexical thought, where concepts are trapped inside modules (constraining the thoughts of our ancestors and other creatures).

Be that as it may, the upshot of this discussion of lexicalization is that the minimalist (narrow) lexicon is very impoverished. All it provides to the syntax are items with “open” slots (monadic predicates, unvalued features). The role of syntax is to take these lexical atoms and form molecules (merge these predicates, and value these features under Agree) before shipping them to the interfaces, where they will be converted into sequences (a linear string, or a string of monadic predicates). Having established this much, the next question for us in this study is what form Merge must have if it is to meet these basic, arguably most minimal interface requirements.

### 3.3 On the form of Merge

#### 3.3.1 The symmetry of Merge problem, and the need for labeling

With Chomsky (2004: 117), let us define Merge as a symmetric, unlimited/unrestricted (“free”), binary operation forming a simple set.

\[(6) \quad \text{Merge } \{\alpha\} \text{ and } \{\beta\} =_{\text{def}} \{\alpha, \beta\}\]

The basic operation defined in (6) provides the basic concatenative procedure that must exist to string together terminals at PF and monadic predicates at LF. Basic Merge is also sufficient to define the basic associative process that place elements on a par on the first tier of metrical structure (see Wagner 2005, 2007), and, arguably, also sufficient to characterize the basic process of vocabulary insertion, which essentially combines a bundle of morphophonological features and a terminal symbol (bundle of syntactic features).
The open-slot/edge-property imposed on concepts upon linear-
ization provides the minimal combinatoric potential to allow Merge
to operate.

The problem with (6) is that so defined Merge does not yield a
rich enough representation to be able to map syntactic objects onto
ordered, asymmetric representations such as a linear string defined
over precedence \( a > b \neq b > a \), or a structured representation of
arguments or scopal elements (John bit Fido \( \neq \) Fido bit John), or
articulated prosodic structures defined over prominence \( ([A \& B] \& C) \neq [A \& (B \& C)] \). To capture these relations, the interfaces
must be provided with representations properly formatted. That is
to say, some asymmetry must be built into (the product of) Merge.

It is tempting to rely on the order of applications of Merge to
build the relevant asymmetry in the syntax (see Chomsky 2005: 14,
2007: 11), where complement formation is defined as “first-Merge,”
and specifier-formation as “second-Merge”; see also Fortuny 2006).
Under this view, the history of derivation would be key. Thus, \( \delta \)
would be in an asymmetric relation with both \( \alpha \) and \( \beta \) by virtue of
having been merged to the set \( \{\alpha, \beta\} \) already merged: \( \{\delta, \{\alpha, \beta\}\} \).

But although derivational time could have the right atomizing
property, freezing \( \{\alpha, \beta\} \) for purposes of subsequent applications of
Merge, the ordering problem within \( \{\alpha, \beta\} \) remains.

Several solutions have been proposed, but I think they are all
unsatisfactory. But to appreciate the problem at hand, it is useful to
discuss a few prominent proposals briefly.

Solution 1: Zwart (2003) and Fortuny (2006) suggest we solve the
asymmetry problem by first merging one of the Merge members to
the empty set (corresponding to time 0 of the derivation), adding
the second Merge member to the set \( \{\phi, \alpha\} \) already formed. I think
one should resist such a move, and keep to the idea that all instances
of Merge combine lexical items (Inclusiveness within narrow syn-
tax), and that Merge is, at some level, fundamentally symmetric (if
only to be able to derive the equivalence between the product of
Merge and the product of Move from the symmetry of Merge, as
I hinted at in Chapter 2.)
Solution 2: Guimaraes (2000) and Kayne (2007) suggest we let one of the Merge members merge to itself prior to combining with another lexical item. This effectively preserves the idea that Merge combines lexical items, but at the cost of allowing gratuitous (self-)Merge, something which I will suggest below we should do well to avoid (to generalize Last Resort to all instances of Merge; see already Watanabe (1996: 142) and Collins (1997: 66)).

Solution 3: Yet another solution to the problem, suggested by Chomsky (1995: 337) and explored in Moro (2000, 2007) is to force an asymmetry between the two Merge partners by moving one of them out of the set formed by Merge. Since traces of movement are typically not pronounced, they need not be linearized, hence for interface purposes the set containing α and a (trace of) β is equivalent to the singleton {α}. I think this view is problematic for several reasons: First, assuming a strict reading of movement as copying/remerge, how does the interface know not to pronounce the copy/occurrence of β merged to α? Second, if mapping to the interface is strictly cyclic, and takes place upon Merge (as in Epstein and Seely 2002)—before any Merge member has had a chance to move—the lack of asymmetry will still be present. Third, the solution concentrates on the problem posed by the lack of asymmetry of Merge for linearization. But the problem is more general: prosodic structure and semantic structure also require asymmetries, and evidence suggests that “traces” participate in stress assignment (see already Bresnan 1971) and interpretation (reconstruction effects).

In addition to the problem posed by the lack of asymmetry within the object formed by Merge, the lack of asymmetry internal to Merge poses a problem for the next application of Merge. Typically, when δ combines with {α, β}, it really combines with one of them (think of a Tense head combining with a Verb Phrase (say, {V, DP})—T really combines with V). This was the asymmetric is-a relation of Phrase-Structure representations that was coded in labels (see Chomsky 1955). As soon as Merge comes unlabeled, as in (6), the selection problem surfaces. Building on Collins (2002), Chomsky (2005: 10) (see also Mayr 2007 and Moro 2007) suggests that the selected element is designated by a search algorithm akin to
the one underlying agreement (*Agree*). In his terms, “the label [...] is [the element] selected.”

As Jayaseelan (2007: 9) observes, the search required arguably involves complexity. Unless there is an asymmetry built internal to Merge, both Merge members must be inspected for purposes of *Agree*. For this reason, Chomsky’s (forthcoming: 10) complete formulation of the labeling algorithm is that “the label selects and is selected.” The suggestion that the label be equated with the selector is already present in Chomsky (2000: 133) (see also Collins 2002, who calls the label/selector the “locus”\(^{20}\)), and is, I think, close to correct. But it should be pointed out that the suggestion is incompatible with Chomsky’s (2004: 112) own suggestion that s-selection considerations be eliminated from the syntax and relegated to SEM.\(^{21}\) Clearly, letting the is-a relation be defined at SEM is inadequate, as the

\(^{20}\) Collins differs from Chomsky in taking the locus to be a very derivational notion. For Collins, locus and head/label differ in two crucial respects. One is that there can only be one locus per stage of derivation (whereas there can be many labels per stage of derivation). The other is that the locus is a derivational notion (for Collins, once its selectional requirements have been met, the locus ceases to exist), whereas labels persist throughout the derivation.

Hendrick (2007) takes Collins’s description of the locus as an argument against the elimination of labels, as the highly derivational character attributed by Collins to the locus appears to predict many more instances of “projection” reversals than are empirically attested.

For additional problems with Collins’s approach, see Boeckx (2002), and Irurtzun (2007). See also next note.

\(^{21}\) In addition, assimilating selection to *Agree* is not unproblematic from a technical point of view, as pointed out by Seely (2006) in his discussion of Collins (2002).

The crucial assimilatory step according to Collins consists in extending the notion of Minimality to subcategorization. (For selection as an instance of feature-checking process, see also Svenonius 1994.) Collins understands Minimality as in Chomsky (2000): a function of the Agree relation, whereby X and Y enter into a successful Agree relation if no matching element Z intervenes between them (intervention being defined in terms of c-command).

Collins notes that if c(ategorical)-selection is indeed reducible to s(emantic)-selection, as argued for in Pesetsky (1982), then, the task of a label-free theory is simplified, as it is c-selection that makes crucial use of (category) labels. But if the reduction is not feasible, then statements like (i)/(iii) must be reformulated as in (ii)/(iv).

i. *destroy*\(_v\) subcategorizes for a DP

ii. *destroy*\(_v\) probes for a D-feature
very same is-a relation is used at PF: the same element acts as the head of a phrase for PF and LF-purposes. Letting SEM decide the

Likewise, (iii) must be reformulated as (iv).

iii. *theD subcategorizes for an NP
iv. theD probes for an N-feature

The Minimality condition on Agree rules out (v), as destroy intervenes between the and city.

Although extending the Agree/checking relation to selection is desirable (it reduces the number of operations in the grammar), there are several problems with Collins’s proposal.

First, it crucially relies on the existence of categorial features such as +N, +V. The explanatory inadequacy of the latter has recently been emphasized by Baker (2003). Chomsky (2001: 7, 14), following work in Distributed and Parallel Morphologies (see Marantz 2000, 2006; Borer 2005), argues for the elimination of categorial features. With these gone, it is not clear how to state the Agree relations that Collins needs.

But setting aside the conceptual status of categorial features (which may be replaced by other features over which Agree could be stated, leaving the essence of Collins’s account intact), there are important technical problems with the extension of Agree to subcategorization, as pointed out by Seely (2006). Seely considers two problematic cases.

The first concerns (v). In order to exclude the example, destroy must intervene. But, as Seely correctly observes, in order to intervene, destroy must have a matching feature, which it lacks. (*The probes for an N-feature.) Seely concludes that for the desired result to obtain, minimality for subcategorization must be stated so as to allow intervention by any (matching or not) lexical category. But that entails that the Agree-relation for subcategorization is distinct from that of Agree involved in Case/agreement-checking. Seely notes that (v) can be excluded by the locus principle, as it contains two unsaturated elements at a given stage of the derivation (the and destroy). Once destroy is introduced into the computation, the insertion of a second argument must take place (assuming, for simplicity, that city qua NP can saturate one theta-role of destroy). Introducing the fails to meet that demand.

Although the locus principle appears to solve the problem posed by (v) for Collins’s theory of subcategorization, careful consideration reveals that it is not enough. The second problem raised by Seely against Collins’s theory pertains to (vi).

vi. *the the picture

It is not clear how the Locus principle excludes (vi). After saturation of the selectional property of the, what prevents inserting a second the which would probe and agree with picture? Appealing to minimality (one instance of the intervenes between the second instance of the and picture) won’t do, as minimality would also (incorrectly) exclude (vii).

vii. see the picture

As Seely notes, the flattening of phrase structure resulting from the absence of labels (the and picture are now on an equal footing) results in picture intervening between see and the, blocking the desired subcategorization/Agree relation between them.

viii. (see (the picture)) = (see (picture the)) →

(see (picture the))

Saying that picture does not intervene in (viii) in the absence of a matching feature forces one to appeal to the Locus principle to exclude (v), which leaves (vi) unaccounted for.

22 Higginbotham (2002: 582 n. 9) points out that a few head mismatches may exist. But to the extent they exist (which I doubt), some explanation for why they are so rare is required.
label is similar to letting the label be determined upon linearization (as in Moro 2000). Both mechanisms increase the computational power of the external systems. Specifically, both mechanisms depart from the idea that syntax provides instructions to the interfaces, the latter being merely interpretive/selective.\(^{23}\)

To get around this problem, Chomsky (2005: 11) takes the labeling algorithm to “apply freely,” letting deviant outputs be filtered out by the external systems with which syntax interfaces. I think we can provide a more deterministic labeling algorithm (one that guarantees interface legibility). For now, let me stress the upshot of this discussion concerning the form of Merge: some asymmetry needs to be coded into Merge (the traditional role of labels). That labels are needed is not a particularly novel claim (see already Chomsky (1995: 244), and more recently, Boeckx (2002), Hornstein (forthcoming), Fukui (2005), Irurtzun (2007), Koster (2007), Hendrick (2007), Hinzen (2006, 2007), Langendoen (2003)), but Collins (2002) and Seely (2006) are right to emphasize that we would like to know what labeling really is (for). I have suggested that labels are needed to ensure a proper mapping to the external systems. Unlabeled Merge products are not properly formatted for SEM or PHON. In addition to figuring out what labels do, we would also like to know what the specific mechanism of labeling/projection is, and be able to understand what it is about the head of a phrase that allows it to project; put differently, what makes a head a head?\(^{24}\)

I try to answer these questions in the next two sections.

3.3.2 \textit{Shaping Merge, or what labels do}

Let me begin by saying, with Chomsky (1995: 244), that I take the label of the set formed by \{\(\alpha\), \(\beta\)\} to be either \(\alpha\) or \(\beta\). Considerations of

---

\(^{23}\) Hinzen (2006, 2007) repeatedly points out the perils of this minimalist tendency to relegate syntactic problems to the interfaces.

\(^{24}\) Chomsky’s (forthcoming: 10) claim that the label is the head begs the question. Taking a head to be a simplex terminal is problematic when two simplex terminals merge (the case of symmetry discussed by Chomsky 1995: 337, Moro 2000, Donati 2006, and others).
Inclusiveness ban the introduction of an element extrinsic to Merge itself (cf. the bare phrase-structure claim that formatives like X’ or X” don’t exist). Chomsky rules out other logical possibilities for labeling: intersection of \{α, β\}, and union of \{α, β\}. According to Chomsky, the intersection and union options are immediately excluded: the intersection of α, β will generally be irrelevant to output conditions, often null; and the union will not be irrelevant but “contradictory” if α, β differ in value for some feature, the normal case. We are left with [only one labeling option]: the label K is either α or β; one or the other projects and is the head of K. If α projects, then \( K = \{α, \{α, β\}\} \).

Chomsky thus concludes that labeling is asymmetric projection. Although several authors have taken issue with Chomsky’s reasoning just quoted (see especially Koster 2007 and Citko 2006), his conclusion is exactly what we need to render Merge asymmetric, hence legible to the external systems. Other options (union/intersection) would fail to make Merge asymmetric enough, since they would emphasize what is common between α and β, not what makes them distinct.

Arguments against Chomsky’s logic go in three directions. First, Chomsky’s claim that the intersection of α and β will often be null is now standardly taken to be incorrect. Most current studies in minimalism take Merge to be subject to Last Resort (see Collins 1997; Hornstein 2001; Chomsky 2000; Boeckx 2002; among many others), or what Pesetsky and Torrego (2006) call the Vehicle Requirement on Merge: whenever α and β merge, a checking relation (aka a Probe-Goal relation) must exist between them. If this is a fact about Merge, it follows that the intersection of α and β will never be null, since for a checking relation to exist between them, some feature F on α must match some feature F on β (see Chomsky (2000: 122) on Match as being a prerequisite for Checking).

Second, Chomsky’s claim that the union of α and β would often be contradictory is true only if we let feature values project. If we let the matching feature project (irrespective of value), no contradiction will ensue. (Elsewhere (Chomsky 2000: 124), Chomsky claims that various syntactic processes (such as minimality) are blind to feature values.)

Third, as Koster (2003) notes, the list of options Chomsky considers in his characterization of projection is too narrow. Koster observes that if what projects is a subset of the intersection of α and β (the matching feature between them), Merge becomes a subcase of pied-piping (feature percolation), which offers interesting possibilities when it comes to unifying Merge and Move.

All these objections are sound, but they fail to affect labeling/projection per se. If the union or intersection of \{α, β\} amounts to a feature, and features can’t be manipulated independently of lexical items, it follows that features won’t be able to project. Accordingly, the label can’t be the union or intersection of \{α, β\}.
Furthermore, there appear to be empirical arguments for the claim that labels must be asymmetric (unambiguous/unique). The strongest argument I have been able to come up with comes from a comparison of Baker and Stewart’s (1999) and (2002) analyses of serial verb constructions like (7) (from Edo).

(7) Òzò ẹ̀ hí ẹ̀ ọ̀ wè [Edo]
   Ozo FUT buy yam eat
   ‘Ozo will buy yams and eat them’

A certain kind of double-headedness was proposed for SVCs as early as in Baker (1989), but that view was not in practice widely adopted, perhaps largely because of the ternary branching structure it made use of (as shown in (8)).

(8) FP
    ZP
      F’
        F
          VP
            V
              YP

Baker and Stewart (1999) defend a doubly headed analysis of verb serialization, roughly as in (3). (The representation in (9) is taken from Stewart (2001), where arguments against alternative analyses of verb serialization are provided.)

(9) FP
    ZP
      F’
        F
          VP
            V’
              V’
                V
                  YP
                    V
                      V
                        WP

(9) shows that binary branching can be assumed alongside a doubly headed structure for serial verbs. Baker and Stewart note that (9)
would be excluded under Kayne’s (1994) Antisymmetry Hypothesis, but that not all current theories of phrase structure exclude it. Of relevance here is that Chomsky’s (1995) conception of labeling under discussion appears to allow double-headedness structures in at least some contexts.

Baker and Stewart agree with Chomsky that ultimately, the label must be a set of features associated with the lexical entry of one or more words that the syntactic object is built from, given the assumption that the computational system never introduces new material into the representation (the Inclusiveness Condition). They make the crucial observation that if Chomsky is right about excluding options a (intersection) and b (union), then within the logic of the Minimalist Program we do not want to explicitly ban labeling a phrase by way of Union or Intersection of feature sets, since that can be left to general principles (ultimately, perhaps, Full Interpretation/Legibility at the Interfaces). Once this is granted, it appears that there is at least one important special case where Chomsky’s reasoning does not apply: namely the case in which the features in the labels of the two phrases combined by Merge are identical. In that case, the intersection will be no less than the features of one of the subparts, and the union will be no more than that; either would give a coherent label for the new category that could support further computation. Thus, doubly headed phrases emerge as a natural possibility in the Minimalist Program. Indeed, as Baker and Stewart emphasize, it would require an extrinsic stipulation to exclude this possibility.

It is important to realize that for Baker and Stewart options a (intersection) and b (union) arise only when the two phrases combined by Merge are identical in syntactic features (which they take to be the case in verb serialization constructions). In other contexts, they adopt option c (asymmetric labeling).

Interestingly, Baker and Stewart (2002) reject their earlier (1999) double-headed structure for serial verbs on empirical grounds. They show that there is at least one context in which the two verbs forming a serial verb construction don’t behave alike (which is unexpected if they are co-heads). Specifically, they show that in
Nupe verbs can be doubled for focusing reasons (a cross-linguistically common process known as predicate cleft). Importantly in Nupé, only the first verb can be doubled; the second one can’t.

(10) a. Musa du etsi du kun  [Nupé]
    Musa cook yam cook sell
    ‘Musa DID cook the yam and sold it’
b. *Musa kun etsi du kun
    Musa sell yam cook sell
    ‘Musa cooked the yam and DID sell it’

Baker and Stewart take (10) to show that the double-headed structure is not viable. Now, recall that in their (1999) article they had justified such a structure on conceptual grounds, specifically, by appealing to legibility at the interfaces. I take this to mean that labeling cannot be determined by interface mechanisms. Asymmetric labeling, of the type argued for by Chomsky, must be a fact about narrow syntax.

And yet we have reasons to hold on to the claim that Merge is symmetric. First because symmetric formulations of rules are always simpler, and second, if Merge is symmetric, we can understand why labeling is in some sense available to both Merge partners (the chain/projection symmetry argued for in Chapter 2).

I’d like to reconcile these two conflicting requirements on the form of Merge by making use of the concept of spontaneous symmetry breaking. Specifically, I would like to propose that the rule of Merge applies symmetrically, but that the output of Merge breaks the symmetry of rule application (as it must if the product of the rule is to be mapped onto SEM and PHON) as soon as the rule has applied.26 Put differently, merging $\alpha$ to $\beta$ is equivalent to merging $\beta$ to $\alpha$; either application of Merge must result in either $\alpha$ or $\beta$ “projecting,” as represented in (11) (where $\alpha$ projects).

(11) $\text{Merge}(\alpha, \beta) \rightarrow (\alpha, \{\alpha, \beta\})$

By projecting I mean that either $\alpha$ or $\beta$ must be identified as the head or type of the phrase. This is necessary if the product of Merge

26 By claiming that symmetry breaks spontaneously upon Merge, my use of the notion of symmetry-breaking differs from that of Moro (2000).
(a 2-D object) is to be mapped onto a 1-D representation at PHON (a linear string) and SEM (a string of monadic predicates). Dimension reduction (what is known as “projection” in the mathematical literature, not to be confused the linguist’s sense of “projection”) necessarily requires orientation. If we insist on syntax providing instructions to the interfaces that allow for a fast, efficient mapping, we must say that syntax provides the orientation. In other words, the product of Merge must have a prespecified orientation—it must be chiral.

The need for orientation is nothing more than a distinctness condition imposed on Merge partners upon Merge. Concretely speaking, upon Merge, one of the Merge partners must be identified as minimal, the other as maximal. This is what labeling does.

Schematically:

(12) Merge \{X, Y\} → \{X, YP\}

A more accurate representation, one that does not require the X/XP notation, would be an ordered pair, as in (13).

---

27 Chomsky has repeatedly argued against a Pair-Merge notation for (basic) Merge, on grounds that linear order plays no role in narrow syntax (or the mapping to SEM); see, e.g., Chomsky forthcoming: 5). I agree with his contention that linear order plays no role in syntax (contra Saito and Fukui 1998), but from this it does not follow that the product of Merge should not be thought of as a pair (asymmetric object). Using pairs in narrow syntax does not entail the linear order is built into syntax, it simply means that a way of mapping an otherwise symmetric object onto a sequence (which I stress is not restricted to linear order; SEM requires it too if Pietroski’s Conjunctivism is correct) must be specified. With Hinzen (2006, 2007) I think we should resist the temptation of letting the external systems format the objects they must work upon, especially if—as I argue in the text below—narrow syntax already provides the mean to format the information appropriately.

28 It is often pointed out in the mathematical literature that dimension-reduction (“projection”) causes information to be lost (think of the information loss when a human body is “projected” onto a shadow). But no such information is lost in the context of Merge, given that Inclusiveness ensures that the objects of higher dimensions (the labels) are identical copies of the terminals that get mapped onto 1-D representations. This may be taken as another sign of particularly good design.

29 Not only does labeling break the symmetry between X and Y, it also breaks the symmetry between \(X^{\text{min}}\) and \(X^{\text{max}}\) (or \(X^-\) and XP), for prior to merging X and Y, X and Y are both minimal and maximal at the same time.
The pair \(<X, Y>\) is equivalent to the set notation used by Chomsky to indicate labeling: \(<X, Y> = \{X, \{X, Y\}\}\). The pair notation is used by Chomsky (2000, 2004) to characterize adjunction, and is crucially distinguished from Set-Merge, the type of Merge under discussion (I return to adjunction in Section 3.4). At the risk of confusion, I will use the pair notation to refer to the products of Set-Merge because the pair notation captures the idea of asymmetry. A pair is thus a better way of representing the idea that Merge produces a vector, where the asymmetry is one of orientation.\(^{30}\)

Vectors are objects that have both a magnitude, and an orientation (they are richer than scalars, which lack orientation). They are like arrows; they have a clear point of origin and an end point. Just like traditional X-bar phrases, they have a clear beginning (minimal projection) and a clear end (maximal projection). Like traditional X-bar phrases, their point of origin is unambiguous (only one head per phrase). In other words, they are asymmetric, like we want products of Merge to be (call this Vector-Merge). Since they provide an orientation, vectors are ideal objects to map a representation onto another (especially if the representation to be mapped onto is of lower dimensionality). To make the mapping even faster, and more efficient (particularly from the point of view of economizing working memory resources), let me add to the vector in question the requirement that its end point be adjacent to the point of origin. In other words, the arrow should have minimal length; both beginning and end should be quickly detectable. This can be done by demanding that the vector produced by Merge be of magnitude 2. This is in fact the familiar binary branching requirement on Merge, which Kayne (1984) imposed on phrase structure to reduce ambiguity, and which Chomsky has repeatedly mentioned in the context of efficient computation (see Chomsky (2004: 115), and especially Chomsky (2005: 16), where “minimization of search in working memory” is hinted at). A vector of magnitude 2 provides a maximally unambiguous object to the interfaces—ensuring

---

\(^{30}\) Ideally, vectors should be represented with an arrow on top of them, which I will omit for typographical convenience.
a most efficient mapping, one aspect of which I conjecture takes the form of a “Quick Edge Detection” (QED) requirement.\footnote{Something like Quick Edge Detection is standardly assumed in the image-processing literature.}

(14) Quick Edge Detection (QED) conjecture
To ensure efficient, optimal mapping, processing decisions must be made as quickly as possible. Quick edge detection contributes significantly to mapping efficiency.

I use the acronym in part to suggest that if edges (i.e., extremities of the product of Merge; origin and end) can be detected quickly, mapping is optimal.\footnote{See also Chomsky (2000: 111), where “computational complexity” is said to be reduced if (among other things) there is “a guarantee of ‘quick decision’ for all derivations.”} Given that any type of working memory is standardly said to be best at remembering beginning and end (the privileged “anchor points”), it strikes me as reasonable to conclude that if we manage to specify beginning and end unambiguously (via minimal vectors), the system processing the output of Merge will be able to determine the properties of the object immediately (as soon as possible), and will not forget any of these properties as the amount of information is minimized to the maximum. (The need to keep working memory resources is particularly important if mapping to SEM/PHON is strictly cyclic, as recent studies, repeatedly alluded to in this study, have claimed.)

Having established what I take “labeling” to amount to (the specification of an unambiguous orientation), let me now turn to the question of which element is identified as the label.

3.3.3 Identifying the head, or what labels are

Given that labeling amounts to imposing a distinctness condition on Merge, the optimal way of identifying the label—the most efficient labeling algorithm—would be one that capitalizes on an asymmetry already present in the grammar. Here I will follow what I take to be Collins’s (2002) key insight that an unsaturated element counts as the locus (/label) (see already Koopman and Sportiche
(1991: 215)), and claim that the operation Agree, with its inherently asymmetric Probe-Goal relation, is exactly what we need to designate the label of the product of Merge. Recall that under Agree, an element with an underspecified/unvalued feature F (the Probe) seeks the closest matching element that would be able to provide a value for F (the Goal). The Probe constitutes the point of origin of the search, and the Goal provides the end point. Probing, like labeling, is thus vectorial in character.

I thus propose a **Probe-Label Correspondence Axiom** (PLCA), stated in (15).

(15) **Probe-Label Correspondence Axiom** (PLCA) (first version)

The label of \{\alpha, \beta\} is the Probe, where the Probe = Lexical Item bearing \(uF\)

The statement in (15) is identical to Chomsky’s (2005: 7) proposal that “the label […] is the [P]robe that seeks a [G]oal for operations internal to the S[yntactic] O[bject]: Agree or I[nternal] M[erge].” Chomsky enriches (15) by providing a second labeling algorithm for instances of External Merge, where the label is the element that “selects and is selected” (Chomsky 2005: 7). As already pointed out above, appealing to selectional considerations for labeling purposes (an operation which I insist is part of narrow syntax) constitutes a departure from the claim that “s-selection is dispensable [within narrow syntax]” (Chomsky 2004: 112). Contrary to Chomsky (2004: 112), I would like to avoid claiming that “properties of [SEM] determine generally the application of External Merge.” Instead, I propose that the same Agree relation underlies all instances of Merge.\(^{33}\) In other words, we must regard “selection” as an instance of Agree/Probing. To do so, one need not resort to

\(^{33}\) This is not a particularly new proposal. I provided a long argument in favor of a feature-sharing requirement on Merge in Boeckx (2002), a requirement adopted in Gallego (2007). Chomsky (2000: 134) notes that “Set-Merge of (\(\alpha, \beta\)) has some of the properties of Agree: a feature F of one of the merged element (say, \(\alpha\)) must be satisfied for the operation to take place.” Recently, Pesetsky and Torrego (2006) have imposed a “Vehicle Requirement on Merge,” according to which if \(\alpha\) and \(\beta\) merge, some feature \(F\) on \(\alpha\) must probe \(F\) on \(\beta\).

The conclusion appears to have been anticipated in Lebeaux (1988). Thus, in the published version of his 1988 thesis (Lebeaux 2000: xv–xvi), Lebeaux notes that:
s-selectional features in the syntax. We can simply endow the relevant element with an appropriate formal feature that will be read off in selectional terms at SEM, the same way as VP-configurations are read off in thematic terms at SEM under Hale and Keyser’s (1993, 2002) “bare argument structure” proposal.

What I am proposing is similar to what Chomsky did in Chomsky (2000) when he avoided talk of interpretable/uninterpretable features (a distinction syntax should be blind to) used in Chomsky (1995) and replaced them with valued/unvalued features (a purely formal distinction visible to syntax).34

Unlike Collins (2002), I don’t think we need to resort to categorial features if we are to generalizing Agree to “s-selection” situations. Categorial features are notoriously stipulative in character (see Baker 2003), and very quickly dissolve into semantic criteria (see again Baker 2003). I suggest we understand “selectional” features in terms of the familiar $\phi$-features.35 Accordingly, a traditional “selector”

“The proposal here [Project $a$ as a projection of an open-class structure into a closed-class frame] is thus more thoroughgoing than that in the minimalist literature, in that both the composition operation, and the movement operation are triggered by Agreement, and the satisfaction of closed class feature. In the minimalist literature, it is simply movement which is triggered by the satisfaction of closed class elements (features); phrase structure is done simply geometrically (bottom-up).”

34 Whether there is a one-to-one mapping between uninterpretable and unvalued on the one hand and interpretable/valued on the other, as Chomsky (2001: 5) assumes, is disputed (see Pesetsky and Torrego 2007), but it seems to me that some connection between interpretability and valuation is a must, if the syntax-semantics mapping is to be transparent.

35 It seems to me that my proposal is to be preferred over Chomsky’s appeal to an undifferentiated “edge-feature,” which is unlike “standard uninterpretable features,” in that it isn’t valued or “deleted” when satisfied (see Chomsky 2007: 11). Thus defined, the edge-feature is just like a semantic property in Chomsky (2000: 134), which, being interpretable, does not delete (“another difference between theta-theory and checking theory”).

Note that by involving $\phi$-features in “selectional” domains, I cannot adopt the idea that some languages may lack $\phi$-features (see Fukui 2006). For me, the presence of $\phi$-features cannot be a parameter, much like the presence of abstract Case could not be a parameter if abstract Case plays as fundamental a role as Vergnaud claimed it did in his famous 1977 unpublished letter. I assume that the presence of overt Case-markers in languages with virtually no morphological agreement is enough of an indicator (a learning “cue”) of the presence of $\phi$-checking relations (cf. (16)).
(say, V)\textsuperscript{36} is to be understood as a feature bundle containing unvalued ϕ-features: \( uϕ \).\textsuperscript{37} Being unvalued, such features will force probing. The matching Goal will be the closest element with valued ϕ-features. This fits well with the fact that arguments tend to be nominal in character. Invoking ϕ-features to capture selectional relations also allows us to incorporate Joseph Aoun’s suggestion (adopted in Chomsky 1981; see Aoun 1985) that arguments must be “visible” for theta-assignment, with visibility defined in terms of Case. A similar visibility condition (there called “activity condition”), also defined in terms of Case, was imposed by Chomsky on DPs valuing ϕ-features. We can subsume both the visibility condition on theta-assignment and the activity condition on ϕ-feature valuation under a general Accessibility condition demanding that all Goals bear an unvalued Case-feature in all ϕ-Probing relations.

(16) Accessibility condition
ϕ-Probing requires \( uCase \) on Goal

The accessibility condition in (16) appears to pose a problem for the PLCA in (15). If the presence of an unvalued feature on a lexical item identifies the label of Merge, \( uCase \) on the Goal DP mandated by (16) would appear to be sufficient for the Goal to Project. Collins (2002: 47) discusses a similar problem for his definition of what counts as the locus at a given derivational stage (which he takes to be the element with an unsaturated feature), and simply stipulates (as does Chomsky 2000) that “Case cannot probe.” Collins here relies on the fact that Case stands out as a sore thumb in Chomsky’s (1995, 2000) system: it is the only feature that is unvalued on both the Probe and on the Goal—the only feature that is never interpreted at the interfaces on any element. Clearly, this begs the question of why Case should be part of the system in the first place, especially if that system is optimally tuned to interface demands, as the (strong) minimalist thesis contends.

\textsuperscript{36} Following much work in Distributed Morphology, I assume that selectors are functional heads (roots being introduced via adjunction).

\textsuperscript{37} Involving ϕ-features captures the idea that only DPs are genuine arguments (as Angel Gallego explores in work in progress).
Faced with this challenge posed by Case, various researchers have tried to “eliminate” Case, by either reducing it completely to a morphological notion (McFadden 2004, building on Marantz 1991), or by showing that Case is indeed interpretable. The hypothesis I favor (and support elsewhere; see Boeckx 2003a, 2007a) is the one put forth by Pesetsky and Torrego in a series of work (see Pesetsky and Torrego 2001, 2004, 2006, 2007). According to them, Case is in fact a bad name for a much better behaved tense feature, interpretable on verbal elements, but (typically) uninterpreted on nominal elements.

If Pesetsky and Torrego are correct, the clash between (15) and (16) cannot be avoided the way Collins claims. Stipulating that $uT$-Tense features cannot probe would beg the question of why $uf$ can. I should stress at this point that the issue under discussion is not limited to $f$- and Case/Tense-features. Typically, Merge situations will involve two elements both of which contain some unvalued feature. Accordingly, if (15) is adopted, such situations would not be able to give rise to unambiguous labels. We could let the labeling algorithm apply freely (as Chomsky (2005: 11) does), and simply state that if $uf$ provides the label, a verbal projection results; if $uT$ provides the label, a nominal projection ensues.

However, a closer examination of Agree reveals that focusing on the presence of $uf$ on lexical items makes Agree more symmetric than it actually is. Recall that the presence of $u$Case ($uT$) on DP makes it able to value $uf$ on V. It does not ensure that $u$Case/$u$T will be valued under Merge. Indeed, if Case is thought of as a Tense feature, it is plausible to assume that V lacks the relevant Tense feature (it may lack a T-feature entirely, or it may contain an unvalued T-feature). It follows from this situation that the only Probe satisfied by merging V and DP is V. Accordingly, (16) ceases to conflict with (14) if the Probe is understood as the element whose $uf$ gets valued under Merge. We can thus revise (15) as in (17).

(17) **Probe-Label Correspondence Axiom (PLCA), revised version**

The label of $\{\alpha, \beta\}$ is the Probe, where the Probe = Lexical Item whose $uf$ gets valued
The adequacy of (17) depends on whether there are situations where both Merge partners value each other’s uFs. I claim that such situations don’t exist. Although all elements bearing uFs get valued in the course of the derivation, I claim that there are no situations of mutual valuation under Merge (this leaves open the possibility of mutual valuation taking place at a distance, under Agree). To make this very explicit, let me restate (17) as in (18).

(18)  *Probe-Label Correspondence Axiom (PLCA), final version*

The label of \{α, β\} is whichever of α or β probes the other, where the Probe = Lexical Item whose uF gets valued

(18) expresses the idea that even if both Merge members must enter into checking relations, syntax imposes a certain checking complementarity to ensure interface legibility.

The labeling algorithm adopted here depicts natural language syntax as a computational engine that ensures that only one relevant thing happens at any given time. This observation is not new. Rizzi (2004b: 9) notes that “one driving factor/underlying assumption, CB] of the cartographic endeavor is a fundamental intuition of simplicity […]. Complex structures arise from the proliferation of extremely simple structural units: ideally, one structural unit (a head and the phrase it projects) is defined by a single syntactically relevant feature” (emphasis added, CB).

Rizzi (2004b: 10) goes on to point out that “local simplicity is preserved by natural languages at the price of accepting a higher global complexity, through the proliferation of structural units. […] Recursion is cheap; local computation is expensive and to be reduced to the minimum.”

38 This checking complementarity may be what underlies Chomsky’s (1995: 313) intuition that “theta-relatedness is a ‘base property,’ complementary to feature checking, which is a property of movement.” The approach developed in the text captures this complementarity without requiring that theta-relatedness be seen as a distinct process from feature-checking, nor that checking be equated with movement. It also leads one to view instances of Agree/Valuation at a distance (so-called long-distance Agree) as a spandrel—a natural, inevitable consequence of ensuring unique label identification (*Locally Unambiguous Merge*). If this is correct, the existence of long-distance agreement should not be seen as an imperfection (contra Chandra 2007).
What I would add to this is that it is not so much local computation for syntax that is expensive (syntax computes elements at a distance). What is expensive (and must therefore be reduced) is computation of (necessarily local) units mapped to SEM and PHON. It is the complexity of computation required at the interfaces that must be reduced at all costs. In other words, mapping must be efficient.

Let me close this section on labeling by Agree by pointing out that by generalizing Probing to all Merge relations, External Merge and Internal Merge look even more alike than they did when we first discussed their (formal) equivalence in Chapter 2. They now share a substantive equivalence, being both subject to Last Resort. Although the claim that what we used to call selection is actually probing increases the amount of determinism in narrow syntax,\(^{39}\) Merge (internal or external) remains blind to semantic considerations. Replacing s-selection by f(eature)-selection (i.e., Probing) is done solely to ensure that Merge be QED-compliant (at both SEM and PHON), it does not guarantee semantic appropriateness/well-formedness.

In sum, seeking a natural labeling algorithm appears to lead us to a greater symmetry of rule: both Merge and Agree, underlying both External Merge and Internal Merge, are composed of a symmetric part (Match), and an asymmetric part (Value/Project). The resulting asymmetry between Probe and Goal is not intrinsic (we saw that both Probe and Goal can bear uF), it is a symmetry-breaking process imposed locally (spontaneously) for purposes of efficient mapping. Globally, we can see the symmetry between Merge members (Probe and Goal) restored if we view chains as projections, as I suggested in Chapter 2—a global symmetry to which I return in Chapter 5.

It should be obvious that the approach to labeling pursued here is geared toward ensuring that Merge be unambiguous/asymmetric

---

\(^{39}\) Let me stress that by “determinism,” I mean that the syntax ought to provide legible (i.e., properly formatted) instructions to SEM and PHON. I do not mean that interpretability is guaranteed (many outputs of syntax may end up being deviant at SEM; think of colorless green ideas sleep furiously). For valuable discussion of the legibility/usability vs interpretability (in the context of Frampton and Gutmann’s (2002) notion of “crash-proof syntax”), see Epstein (2007).
locally. Since the symmetry of Merge is restored globally, the approach demands that syntax be derivational (at the representational level/globally, syntax would be too symmetric for SEM and PHON to be able to deal with).

3.4 Adjunction

Before building on the labeling algorithm argued for in the previous section, to revisit the locality issues raised in Chapter 2 I must make an excursus into the nature of adjunction, an issue intimately related to labeling.

As many have noted (see Chomsky 1995: 382 n. 22), current syntactic theory lacks a compelling theory of adjunction. Put bluntly, we know that adjunction exists; that is, we know that some elements relate syntactically in a way that is not amenable to complementation or specification, but we don’t know how to characterize that third relation precisely (formally).

As Hornstein and Nunes (forthcoming) point out, the default strategy when it comes to analyzing adjunction is to treat it as special, subject to highly specific operations or configurations (Chomsky-adjunction, Late-Merge, Pair-Merge, etc.). Put differently, the null hypothesis in mainstream generative grammar is this: complementation is the norm; adjunction, the exception.

Paul Pietroski has argued in recent work (see Pietroski 2005 and references therein) that this appears to get some features of the language faculty exactly backward. According to him, adjunction ought to occupy a more central place in syntactic theorizing, as it gives rise to arguably the simplest syntactic relation there can be: association. Arguments, by contrast, require a more complex syntax.

Pietroski’s claim is based upon the fact that in Neo-Davidsonian representations of events, adjunction translates into the most natural mode of composition: conjunction. Adjuncts (understood as the result of adjunction) directly modify the event variable (e), and combine with other event-modifiers, as shown in (19).
(19) It rained heavily = ∃e [rain (e) & heavy (e)]

By contrast, arguments require a theta-role before they can be conjoined like other predicates/modifiers of events. Compare John and heavily in (20).

(20) John breathed heavily = ∃e [breathe (e) & heavy (e) & Agent (John, e)]

On the basis of this difference, Pietroski notes that it is odd to treat adjunction as more complex than complementation. The result of the mapping to SEM suggests just the opposite.

As an example of the standard treatment of adjunction as special, consider Chomsky’s (2000: 133, 2004: 117) claim that complementation takes the form of “Set-Merge,” characterized as “simple” set formation ({α, β}), whereas adjunction takes the form of “Pair-Merge,” a more complex construct. Unlike Set-Merge, Pair-Merge creates an ordered pair <α, β>, and requires a special spell-out operation, SIMPL, which converts the pair <α, β> into a simple set {α, {α, β}} (which could have been achieved by Set-Merge) (Chomsky 2004: 118). Notice that since <α, β> = {α, {α, β}}, adjunction for Chomsky amounts to a set with an explicit coding for the label. This makes it clear that adjunction and labeling are intimately related under standard approaches to adjunction, and may explain why label-free theories have remained silent on adjunction (see Collins 2002: 48). Setting aside the fact that I think Chomsky’s conception of Merge for complementation is incorrect, for reasons that were made clear earlier in this chapter, there are two interesting properties in Chomsky’s analysis of adjunction. First, Chomsky chooses the pair notation to capture the fact that there is an intrinsic “asymmetry” in adjunction that is missing in complementation (see Chomsky (2000: 133, 2004: 117)). Adjunction is a process that merges α to β, unlike complementation which is order-invariant (cf. Chapter 2): {α, β} = {β, α}. Perhaps the major reason for thinking that adjunction is asymmetric is that it is optional, unlike complementation. Without a direct object, there is no transitive verb. Indeed, without an (internal) argument, there is no argument-taking verb (unaccusative verb). But the status of a host
is unaffected by adjunction: a verb remains a verb, irrespective of whether a manner adverb is adjoined to it.

Second, Chomsky’s analysis makes it clear that upon spell-out, adjunction must be “downgraded” or “simplified.” In other words, it requires a special mapping rule. The clearest piece of evidence suggesting that adjunction is special at spell-out is the fact that adjuncts have a much more flexible mode of attachment than complements: John left quickly and John quickly left are both acceptable, in a way that John kissed himself and John himself kissed aren’t. Although adjunction is not completely unconstrained (witness Cinque 1999), there is no denying that adjuncts are less rigidly ordered with respect to their hosts than complements are with respect to their theta-assigners (see especially Ernst 2000).40

I think that Chomsky is quite correct to insist on both the optional and flexible character of adjunction, but if I am right about the nature of labeling above, complementation must be Pair-Merge (Vector-Merge), and (therefore) adjunction cannot be Pair-Merge. For me, Pair-Merge (understood as a labeled set) gives unambiguous instructions to the interface, and therefore should not result in flexible mapping/linearization. Second, Pair-Merge comes about via Probing, which requires a Probe-Goal relation, so both Merge members are required.

As a matter of fact, I think that Chomsky’s Set-Merge notation fits adjunction like a glove.41 First, the absence of labeling (resulting in Set-Merge) would be due to the absence of a Probe-Goal relation. It seems abundantly clear from the detailed studies of adjunction like Ernst (2000) that adjunction does not involve any featural transaction of the

---

40 For a very suggestive outline of an argument on how the internal hierarchy discovered by Cinque could be derived from simpler considerations, see Fortuny (2006). See also Chapter 4 and Butler (forthcoming).

41 Hornstein and Nunes (forthcoming) make a similar proposal. Building upon Hornstein’s (forthcoming) proposal that Merge is the result of the combination of two primitive operations, Concatenate + Label, they suggest that adjunction amounts to mere concatenation, i.e., grouping without labeling. By contrast, arguments require labeling to be integrated phrase-structurally. The discussion of adjunction included in this chapter was inspired by Hornstein and Nunes’s analysis.
Agree/checking sort (see also Boeckx 2003a: 100).\footnote{On the lack of feature-checking in (and concomitant optionality of) intermediate steps of movements created by adjunction, see Boeckx (2003a: ch. 1), Takahashi (1994), (Bošković 2002a, 2007), and especially Boeckx (2007a).} Adjuncts merge with their hosts under Match (like any instance of Merge), but the matching relation here is distinct from the matching giving rise to valuation. The latter takes the form of a “lock and key” situation (complementarity). Matching under adjunction is like the relation that exists between a key and its duplicate (identity/parallelism). The absence of checking would account for the optional character of adjunction.

Second, due to the lack of explicit mapping orientation in the absence of a label, attachment of $\beta$ to $\alpha$ is left unspecified, hence gives rise to greater flexibility. Furthermore, adopting a Set-Merge notation for adjunction would immediately capture Pietroski’s observation that adjunction is a simpler mode of combination than complementation. If complementation is Pair-Merge, and adjunction Set-Merge, adjunction is in a subset relation with complementation, reflecting the complexity of the latter.\footnote{The idea that arguments require a more complex integration in the guise of labeling is similar to Hale and Keyser’s (1993) intuition that theta-roles are “configurational.” By that they mean that theta-roles (qua properties of arguments) are in some sense the result of hierarchical Phrase Structure. That theta-role assignment depends on projection captures Pietroski and Uriagereka’s (2002) idea that theta-roles “lift” predicates into argumenthood (“type-lifting”) (see also Irurtzun 2007). Since projection and chain-formation are two sides of the same coin, one could say that as the head projects, it pied-pipes (“lifts”) its Merge partner into argumenthood (a higher type).} And because Set-Merge is the simplest product of Merge, adjunction can be iterated the same way Merge is. All in all, the Set-Merge relation appears to be designed to capture all the defining characteristics of adjunction.

To the best of my knowledge, the idea that adjunction does not result in labeling was first expressed by Chametzky (2000) (see also Chametzky 2003). As Chametzky noted, the lack of labeling under adjunction is already explicit in GB treatments of adjunction. In the traditional X-bar schema, adding a complement to a head $X$, or a specifier to $X'$, results in the creation of a new label ($X'$ and $X''$, respectively). By contrast, adding an adjunct to $X'$ or XP returns the same label ($X'$ or XP). So, already in GB linguists expressed the idea...
that adjunction does not result in (new) labeling. What Chametzky did is take adjunction outside of Phrase-Structure grammar, as it were, and claim that adjuncts are not structurally integrated in the way arguments are. As Uriagereka has later claimed (Uriagereka 2001), Chametzky’s analysis may capture the intuition (originally due to Goodall 1987) that adjuncts live on a “separate plane” (Chomsky 2004: 118).

Together with Pietroski, and Hornstein and Nunes (forthcoming), I believe that adjunction is the most primitive syntactic operation. It may in fact be a “vestige,” evolutionarily speaking. As Ernst (2000) makes clear, adjuncts do not have an interesting syntax: they do not involve feature-checking, they are not hierarchically organized, and they relate to their hosts largely on the basis of semantic affinity. This likely characterizes the thought processes of our closest relatives. (As a matter of fact, Kanzi’s “beads-on-a-string”-like utterances (see Anderson 2004: ch. 10), with its loose syntax (lack of rigid ordering), would most readily be characterized as adjunction structures.)

But viewing adjunction as a primitive operation does not make it the default computational process of our modern language faculty. There is a sense in which complementation is more constitutive of natural language as humans now have it. If the key function innovation when it comes to the language faculty is syntax (the pairing of SEM and PHON; cf. the first section of the present chapter), complementation is perfectly suited for this. By contrast, in the absence of explicit labeling (Set-Merge), adjunction poses a mapping problem, and for this reason, should be “avoided.”

Now, since adjunction results in a syntactic object that fails to provide explicit mapping instructions (adjunction is ambiguous Merge), some other mapping mechanism must be used (as a last resort) to disambiguate adjoined structures. To achieve this, I suggest a return to the idea that adjunction necessarily amounts to Late-Merge (Lebeaux 1988; Chomsky 1993). More specifically, I claim that adjunction takes the form of Later-Merge.44 That is, unlike complementation,

44 Note that Later Merge does not mean that adjuncts are attached at the very end of a derivation. Later Merge is not Latest Merge (of the kind defended by Stepanov 2001).
where both Merge members are introduced at the same time (symmetry/order-independence of Merge), adjunction attaches an element to an already present element. Accordingly, adjunction should be conceptualized as in (21) (where the index $t$ represents timing of insertion).

\[
\{\alpha\}_n \{\beta\}_{t+1}
\]

The temporal asymmetry seen in (21) is, I claim, what allows the Set-Merge representation to be mapped onto SEM/PHON.

The absence of explicit labeling under adjunction has important consequences in a variety of domains. First, since they don’t participate in probing/labeling, they won’t be part of the representation of projection. Nor will they be part of the representation of chains. This has the welcome consequence that the intermediate projection levels and intermediate steps of movement created by adjunction will not be part of the projection/chain-representations, which will allow me to maintain the claim made in Chapter 2 that there is at most one $X'/t'$ per projection/chain.\footnote{Excluding intermediate steps of movement created by adjunction is also necessary if one is to reconcile Grohmann’s (2003) claim that an element cannot move internal to a functional domain like the theta-domain with Boeckx’s (2003a, 2007a) claim that successive cyclic movement targets all accessible projections within a given functional domain. For discussion of Grohmann’s approach in the context of successive cyclicity, see Boeckx (2007a: ch. 5).}

The absence of labeling under adjunction will also affect extraction and subextraction possibilities. Concerning extraction, if adjoined structures are not part of projection/chain-representations, they won’t form constituents with their hosts, hence they will not be pied-piped under movement. Witness (22).\footnote{Examples like [eat a cake in the yard] John will $t$ could be analyzed as resulting from adjunction after VP-fronting.}

\[
\text{(22) a. John will eat a cake in the yard}
\]
\[
\text{b. [eat a cake] John will $t$ in the yard}
\]

Concerning subextraction, if adjoined structures are not part of projection/chain-representations, subextraction/chain-formation from within an adjoined structure will inevitably introduce an illicit chain
member in the resulting chain: a violation of endocentricity (cf. Chapter 2). This captures the general ban on extraction out of adjuncts.\(^{47}\)

Lack of labeling also predicts that adjoined elements will be immobile (since they don’t project, they don’t create chains). This appears to be a correct prediction, as adjoined materials have been reported to be unable to participate in long-distance scrambling (see Bošković and Takahashi 1998). It has also been claimed that prototypical wh-adjuncts like why are inserted in their base position in SpecCP (as opposed to being moved there) (see Bromberger 1986; Rizzi 1990; Law 1991, 1993, and much subsequent work). Of course, some adjuncts move long-distance, as in long-distance why-questions (why did you say that Peter left t). Some adjuncts even agree (as in locative inversion constructions in Bantu; cf. Baker forthcoming), and even allow for subextraction (?Who did you go there [to visit <who>]). It is precisely to account for some of these facts that syntacticians have found it necessary to introduce the notion of “quasi-argument” (alternatively, “quasi-adjunct”) into the theory (see Huang 1982 and many others since).

To capture such facts, I follow Hornstein and Nunes (forthcoming) in claiming that the term “adjunct” is ambiguous. It can designate an object that lacks a theta-role (non-argument), or an object that is adjoined. So far, I have been talking about the latter sense of adjunct. The former sense need not involve adjunction. If the process involves some feature-checking,\(^{48}\) the adjunct will be structurally integrated (under labeling) like an argument, and behave as such for purposes of projection/chain-formation. (In such situations—but in these only—

\(^{47}\) My proposal captures the spirit of Chomsky’s (forthcoming) claim that adjuncts cannot be probed because they are on a different plane. See also Uriagereka (2001) and Hornstein and Nunes (forthcoming).

\(^{48}\) There may be various features that an adjunct may possess that would lead to its phrase-structural integration, and I won’t attempt to characterize them here. Focus may be one of them, as Hornstein and Nunes suggest. Miyagawa (2007) claims that focus may be regarded as a phi-feature, on a par with case, person, number, etc. (an “A-bar” phi-feature, so to speak; see also Ouali (2006), Watanabe (2004), Simpson and Wu (2000)). The fact that agreement underlies the integration of adjuncts, and that agreement prototypically involves nominal elements, may capture the oft-expressed intuition that quasi-arguments are nominal-like; see Huang (1982); Reinhart (1995).
adjuncts may become indistinguishable from complements (Larson 1988) or specifiers (Cinque 1999, 2004).

Crucially for us, the apparently optional character of adjunct-integration would then reduce to the optional assignment of extra features like focus upon lexical selection. The syntax can be kept fully deterministic: if there is some featural transaction, we get Pair-Merge. If there is no featural exchange, we get Set-Merge.

Evidence that adjuncts behave like arguments when they must have moved (e.g., when they are construed long-distance) comes from a variety of sources. Let me discuss one such case, from Spanish. (The paradigm discussed here is from Uriagereka (1988). Data from Haegeman (2003) concerning the interaction of fronted adverbs and that-trace effects, and discussed in Chapter 5, point to the same conclusion.)

Consider (23)–(26).

(23) a. Qué vió Juan [Spanish]
   What saw Juan
   ‘What did Juan see?’
   b. *Qué Juan vió

(24) a. Qué pensaste tú que Juan vio
   What thought you that Juan saw
   ‘What did you think that Juan saw?’
   b. *Qué tú pensaste que Juan vio

(25) a. Por qué Juan vio a María
   why Juan saw A María
   ‘Why did Juan see María?’
   b. Por qué vio Juan a María

(26) a. Por qué pensaste tú que Juan vio a María
   why thought you that Juan saw A María
   ‘Why did you think that Juan saw María?’ (embedded interpretation of why intended)
   b. *Por qué tú pensaste que Juan vio a María

(23)–(24) show that subject-verb inversion is obligatory with wh-arguments. (25)–(26) show that inversion is optional in matrix questions introduced by por qué ‘why,’ but not in the long-distance cases (under the long-distance reading of why). This paradigm can be accounted for if we assume that the optionality in (25) is due to
the fact that *por qué* may enter the derivation in two ways, via adjunction (which would correlate with absence of inversion, if inversion is tied to feature-checking in C; cf. Pesetsky and Torrego 2001), or via feature-checking (with inversion). Since insertion of *por qué* in its surface position would be incompatible with its long-distance construal, only movement is possible. Crucially, in this case, the adjunct *por qué* behaves like an argument.

Examples like these suggest that adjuncts should not always be seen as the products of adjunction. They also suggest that there are two kinds of products of Merge: labeled and unlabeled Merge. Since the distinction boils down, in my view, to the presence vs absence of a Probe-Goal relation, with consequences for narrow syntax and both SEM and PHON, it provides another argument against incorporating selectional considerations into the labeling algorithm, as Chomsky (2000, forthcoming) does, since such considerations would not be reflected in adjunct/argument contrasts within narrow syntax and at PHON.

### 3.5 More on projection

We have seen so far that in its simplest (unrestricted, symmetric) formulation, *Merge* would be too ambiguous to meet interface demands (the QED-conjecture in (14)). A disambiguating process has to be built into it. I have suggested that in the core cases (complementation), *Agree* does the job. (For adjunction, asynchrony must be imposed on Merge, as Agree is not involved.)

Using Aristotelian terms, one could say that my proposal concerning the form of *Merge* amounts to saying that the reason projection exists is because *Merge* is unambiguous. *Agree* is the formal cause of unambiguous *Merge*, the existence of the valued/unvalued feature in the lexicon is the material cause, the mapping to SEM/PHON provides the efficient cause, and interface considerations (QED) constitute the final cause. Coupled with Probe-Goal asymmetry, *Merge* yields a labeled set, by which I mean that one of the *Merge*-members orients the set. Accordingly, labeling is more like a “start-as” relation than the more traditional “is-a” relation.
“Is-a” is achieved by combining “start-as” with Minimal Search (Minimality/Economy), which says that the element γ that merges with the set that starts with α (\(<\alpha, \beta> / \{\alpha, \{\alpha, \beta}\}\)) does not look past α—hence equates the whole set with α (the set is (an) α).

It is worth stressing again that by taking the product of Merge to be a vector, I am not building linear order into narrow syntax. Linear ordering is but one of the possible interpretations of the asymmetry created by orientation. It is, so to speak, the horizontal interpretation of orientation. Dominance would amount to the vertical interpretation of orientation. The 2-D character of syntax (cf. (1)) is thus the result of the fact that products of Merge get mapped onto different dimensions at SEM and PHON. Put differently, vectorial Merge is the most concise “(context-free) Phrase-Structure grammar” syntax could create. It has a designated initial symbol (Σ), which is the initial symbol (the head/label) and an arrow connecting the initial symbol on its left to terminal symbols on its right. (Since α is also mapped onto SEM/PHON, it has a dual status as a non-terminal/terminal symbol, as do all labels.)

\[
(27) \quad <\alpha, \beta> = \\
\begin{align*}
&\alpha \max \\
\text{a. } \Sigma : \alpha \\
&\beta \\
\text{b. } \alpha \max \rightarrow \alpha \beta \\
\end{align*}
\]

In tree form:

```
α
  \max
β
```

The discussion so far has focused on labeling. But recall that for Merge to be QED-compliant, labeling/orientation is not enough. The magnitude of the vector must be kept to a minimum, so as to ensure that both the point of origin of the vector and the end of the vector be processed unambiguously. If there is only one other member in the Merge set, the end point of the vector is unambiguously.

\[49\] From this perspective it is not surprising to find that theories of phrase structure that only recognize dominance, like Brody’s (2000) Mirror Theory or Starke’s (2001, 2004) and Jayaseelan’s (2007) specifier-free theories, claim that “syntax is linear” (by which they really mean 1-D).

\[50\] For an illuminating discussion of the properties of Phrase-Structure grammars, see Lasnik (2000).
In concrete terms, this means that the very next thing merged to the label closes the set formed by Merge. QED therefore imposes an adjacency condition—the very locality condition that every linguist takes to be basic. (Unique) labeling and adjacency have thus a common explanation: efficient mapping, understood as minimal search in the service of disambiguation (“quickest detection of anchor points”)/QED). QED thus imposes a minimality condition on Merge that will, by symmetry, percolate to domains that are products of Merge too: full X-bar projections, extended projections, and chains. Unambiguous Merge forms the shortest path. Seen in this light, Merge takes a somewhat “perverse” form: it brings together two elements and featurally relates them, but only to impose a distinctness on them. But this is exactly what we expect under Inclusiveness. Lexical integrity/identity should be preserved under Merge. Recall that the key property of natural language syntax is discrete infinity. Merge is not a blending system. It doesn’t lose any information.

Armed with such considerations we can now go back to the locality considerations laid out in Chapter 2. The bulk of this task (issues pertaining to large syntactic objects like extended projections and chains) will be left for the next two chapters. Here, I want to set the stage by examining the structure of full-size phrases, captured by the traditional X-bar schema. A fully expanded phrase is the result of several Merge applications. According to the definition of Merge advocated here, phrases will be limited to two Pair-Merge products (products of adjunction won’t be so limited). Here is why.

The first application of (unambiguous) Merge will result in a head-complement structure. But this does not exhaust the range of QED-compliant structures. Given that products of Merge are mapped onto a 1-D representation (a line), what the first application of Merge does is define the right edge of the head of a projection by associating it with the Merge partner. QED does not rule out remerging the head with another element, for the latter can then define the left edge of the head. This can be schematized as in (28).

On minimality as disruption of adjacency (at appropriate levels of representation), see already Wilkins (1977, 1980). On minimality as a condition reducing ambiguity (of government), see Rizzi (1990: 1).
(28) a. first Merge: \( <\alpha, \beta> \)
b. second Merge: \( <\alpha, \gamma> \)
\[ \therefore \text{mapping: } \beta = \text{right-edge of } \alpha; \gamma = \text{left-edge of } \alpha \]

1D-representation:
\[ \gamma \rightarrow \alpha \rightarrow \beta \]

Temporal representation:

\[
\begin{align*}
\text{time } t+1 & \quad <\gamma \rightarrow \alpha> \\
\text{time } t & \quad <\alpha \rightarrow \beta>
\end{align*}
\]

Set representation: \{\alpha, \{\gamma, \{\alpha, \beta\}\}\}\}

Familiar tree representation:

It should be obvious from the tree representation just given that the traditional X-bar schema can be easily recovered from a bare phrase structure of the sort developed in this chapter. The first application of Merge creates the first substantive level of projection (\( X' \)), and the second application of Merge, the second and maximal substantive level of projection (\( X'' \)). Since Merge maps onto 1-D representations, any other application of Merge would have to involve adjunction, or would have to treat \( \alpha \) as the non-head. Other options would be indistinguishable from either the first pair created by Merge, or the second pair created by Merge, and as such, would be ruled out by QED. In a 1D-world, there can only be two edges.

Note that we need not build these constraints into Merge. Merge remains symmetric throughout. The restrictions on Merge (more accurately, on the product of Merge) can “emerge” from efficient mapping considerations.

As I keep stressing, left-edge and right-edge should not be understood solely in linear terms. We could, without loss of information, call them first and second edge, or inner edge and outer edge. The latter terminology is reminiscent of “inner” object and “outer” object (Huang 1997), or “internal” and “external” argument (Baker 1997). Indeed, efficient mapping consideration (QED) ensures that a given
predicate can only take at most two arguments. Any extra argument will have to be adjoined, “non”-core, or, as Baker (1997) calls “applied” arguments, “oblique.”

Note that if we understand left-edge and right-edge in linear terms, and assume that, by default (transparency), the orientation of Merge maps onto precedence (see also Kayne 1994: 36), (27) amounts to saying that universally, (pair) Merge maps onto a head-complement order. Since, by QED, the left-edge and the right-edge must be distinct, it follows that specifiers (produced by the second application of Merge) will be on the other side of complement. We thus capture Kayne’s (1994) universal Specifier-Head-Complement order.

Left-edge and right-edge also translate straightforwardly onto prefix and suffix, with suffixes forming a closer bond with the head/stem by virtue of having merged with it first. This correctly predicts Julien’s (2002) finding that prefixes are not as tightly connected to the stem as suffixes are (i.e., suffixes are not “postfixes” (the mirror image of “prefixes”)).

What we are witnessing is essentially a distinctness condition on non-heads within a phrase, which follows (by symmetry of Merge) from the distinctness condition imposed on Merge by QED. Not only must Merge members be distinct within the Merge set, all Merge members must be kept distinct across Merge sets.

It follows from this that one cannot remerge the complement of X into the specifier of XP. This would produce the same pair, which would be either vacuous \((\alpha \rightarrow \beta) = (\beta \leftarrow \alpha)\), or contradictory \((\beta \text{ both precedes and follows } \alpha)\). We thereby exclude any movement from Complement to Spec, or any movement internal to a phrase, instances of movement that have been characterized as too local in the recent literature on “Anti-locality” (see Murasugi and Saito 1995; Bošković 1994; Kayne 2005; Grohmann 2003; Abels 2003; Pesetsky and Torrego 2001; Boeckx 2007a, and references therein).

This restriction holds even if we impose a “one-argument-per-phrase” restriction, as Larson (1988) does. If we do so, the restriction noted in the text will have to be rephrased in terms of extended (thematic) projection, which is straightforward (see Chapter 4).
We may want to extend the logic just used to capture Anti-locality to situations that have been analyzed as violating a [*XX*-filter (van Riemsdijk forthcoming)—cases that Richards (2006) characterizes as “situations where two elements are [in a way to be made precise] ‘too similar’ and ‘too close’ ” (a syntactic version of the Obligatory Contour Principle in phonology).

Richards notes that in language after language situations of “syntactic stuttering” are excluded. For instance, many languages exclude sequences where both the subject and object DPs follow the verb (see Alexiadou and Anagnostopoulou 2001). A relevant example, from English quotative inversion, appears in (30d).

(30) a. “It’s cold” said John
b. “It’s cold” John told Mary
c. “It’s cold” said John to Mary
d. *“It’s cold” told John Mary

Richards also observes that numerous languages disallow two object DPs from bearing special markers for animacy. For example, Spanish marks [+animate] objects with the preposition a, as shown in (31).

(31) Ana levantó a un niño [Spanish]
    ‘Ana lifted a child’

The same preposition a is used to introduced indirect/propositional objects, as shown in (32).

(32) No (le) hablé a nadie
    ‘I talked to no one’

Interestingly, Spanish disallows a sequence of two a-phrases in ditransitives, as (33) illustrates.
As Richards notes, it is not the case that any sequence of linearly adjacent elements of the same morphosyntactic type is excluded. For example, English allows double DP-object constructions without any problem (34).

Richards uses cases like (34) to suggest that the “identity-avoidance” situations at issue cannot be a mere linear constraint. He goes on to argue that the proper characterization of the [*XX]-ban is this: “Within a spell-out domain, no two elements with an identical set of features [feature types—CB] may exist.” Richards argues that this ban should be understood in the context of linearization. If PHON receives instructions of the sort {...XX...}, such instructions are rejected as “uninterpretable” or “uninformative” (such instructions would amount to dictating that X precedes itself). (For Richards’s analysis to go through, it is important that morphophonetic pieces, which could distinguish between X(DP₁) and X(DP₂), be added after syntax/linearization, as in the realizational morphology frameworks discussed above).

Richards’s analysis is very much in the spirit of the Unambiguous Merge approach pursued here. I would in fact like to offer a minor amendment of it to make it follow from the logic of Anti-locality discussed above.

As far as I could determine, Richards’s analysis takes the relevant spell-out domains to be full phrases. Under the definition of full phrase pursued here, this means that the relevant spell-out domains consist of two products of Pair-Merge, one creating a head-complement pair, the other a specifier-head pair.

Now, it follows under our approach that specifier and complement must be featurally distinct for the same reason that the complement of

53 Technically, for Richards, these count as phase complements. But since phase complements are the maximal projections whose sisters are phase heads (see Richards 2007), phase complements can be equated with maximally expanded phrases (full XPs).
X cannot move to the specifier of X, since both types of situations would instruct PHON to regard the same element as both the left edge and the right edge of X: a contradiction that would make PHON “gag.” In other words, the featural distinctness condition on Merge gives you an “exclusivity condition” on syntactic relation within the phrase. (Since no such featural distinctness condition applies to adjunction, it also follows that adjoined material can iterate.) I am not suggesting that all instances of “haplology” be treated in the way suggested by Richards. I take it that at least some instances of haplology are purely morphophonological. For example, Bošković (2002b) and Bošković and Nunes (2007) provide evidence that the ban against identical wh-phrases in multiple wh-fronting languages is a purely PF-matter (Bošković (1999) shows that syntactically/featurally, the two wh-phrases are distinct).

(35)  a. Cine ce precede [Romanian]
     who what precedes
     “who precedes what”
     b. *Cine precede ce
     c. *Ce ce precede
     what what precedes
     “What precedes what”
     d. Ce precede ce

Following Boeckx (2003b, forthcoming a), I would like to argue that the pattern of featural exclusivity identified by Richards extends beyond the cases he discussed. For instance, languages with multiple object agreement (associated with a single head) impose a distinctness

54 Accordingly, the ban on $<$ Verb DP$_1$, DP$_2$ $>$ sequences discussed by Alexiadou and Anagnostopoulou (2001), Richards 2006, and Mayr 2007 can be seen as the result of incompatible instructions sent to PHON. If each Merge pair is sent to PHON, $<$ Verb DP$_1$, DP$_2$ $>$ is combination of two instructions of the form $<$Verb DP$>$.

55 Jeong (2005) and Clemens Mayr (p.c.) have pointed out that the cases discussed under the proposed extension would be problematic for Richards’s specific formulation of distinctness, which does not take into account the way features of arguments relate to the features of a head.

56 Situations that Hiraiwa 2001 dubbed “Multiple Agree” (see Boeckx 2003b; Hiraiwa 2005; van Koppen 2005; see also Bošković 1999 and Ura 2000 on multiple feature checking).
condition on agreement markers—a condition known as the Person-Case Constraint (PCC) (Perlmutter 1971; Kayne 1975; Bonet 1994).

\[36\] Person-Case Constraint (PCC)
If Dative agreement/clitic, then Accusative agreement/clitic = 3rd person

The constraint in (36) covers cases like the multiple agreement constraint in Basque (37), and the clitic cluster constraint in French (38).

\[37\] a. Azpisapoek etsaiari misilak saldu d-Ø-izki-o-te
    \hspace{1cm} \text{traitors-ERG enemy-DAT missiles-ABS sell ABS3-DAT3-ERG3}
    \hspace{1cm} [\text{Basque}]
    \hspace{1cm} ‘The traitors sold the missiles to the enemy’

b. *Azpisapoek etsaiari ni saldu na-i-o-te
    \hspace{1cm} \text{traitors-ERG enemy-DAT me-ABS sell ABS1-DAT3-ERG3}
    \hspace{1cm} ‘The traitors sold me to the enemy’

\[38\] a. Jean le lui présentera
    \hspace{1cm} \text{Jean it him will-present}
    \hspace{1cm} ‘Jean will introduce it to him’

b. *Jean me lui présentera
    \hspace{1cm} \text{Jean me him will-present}
    \hspace{1cm} ‘Jean will introduce me to him’

Bejar (2003), Bejar and Rezac (2003), Anagnostopoulou (2003), Boeckx (2000, forthcoming b), Richards (2005), and Rezac (forthcoming), among others, have provided evidence showing that PCC effects reduce to the following conditions:

(i) PCC arises when two (or more) DPs, $D_1$ and $D_2$, relate to the same Phi-Probe/Case-licenser.

(ii) Person agreement obtains with the closest DP (the one DP that asymmetrically c-commands the other DP before movement).

(iii) The asymmetry among DPs relating to a single head is preserved after movement.

The latter condition stresses the fact that even if all the DPs target the same spell-out domain, some asymmetry exists among them. That is, the PCC amounts to a constraint on “feature-checking uniqueness” and cannot be reduced to a ban on a single DP per spell-out domain.
A similar condition on feature-checking exclusivity holds in so-called multiple specifier constructions. The featural asymmetry among multiple specifiers is very clear in the context of multiple wh-fronting (see Bošković 1999).

Consider the following examples from Bulgarian.

(39) a. Koj kogo kakvo e pital? [Bulgarian]
   who whom what is asked
   'Who asked whom what?'
b. Koj kakvo kogo e pital?
c. *Kogo kakvo koj e pital?
d. *Kakvo kogo koj e pital?
e. *Kakvo koj kogo e pital?
f. *Koko kajvav koval e pital?
g. *koj e pital kogo kakvo
h. *kogo e pital kajvav kajvav
i. *kakvo e pital kajvav kajvav
etc.

To analyze cases like (39), I adopt Bošković’s (1999) claim that the first wh-phrase in the wh-sequence enters into a feature-checking relation with C\textsuperscript{0} that no other wh-phrases participate in. In part

57 McGinnis’s (2004) Lethal Ambiguity condition, which bans the establishment of an anaphoric [i.e., identity/symmetry] relation between two specifiers of the same head, arguably belongs here too. The lack of relevant asymmetry would here be the cause of ill-formedness at SEM, not PHON. As Norbert Hornstein suggests (p.c.), perhaps other binding constraints, such as strict subject orientation (ban on an object reflexive being bound by another object in the same clause) found in many languages, may be analyzed in terms of a distinctness condition imposed at LF. I will leave this as an idea to explore in future work.

58 Although Bošković has shown in a series of work (Bošković 1999, 2002b) that Serbo-Croatian appears to lack the asymmetric aspect of multiple wh-fronting, there is reason to believe that the Bulgarian pattern is more basic than the one found in Serbo-Croatian (among other languages). Bošković in particular has conclusively argued that the lack of superiority is only true in matrix questions in which there is no overt evidence for a complementizer. Elsewhere (e.g., embedded questions, long-distance questions, correlatives, matrix questions with overt complementizers), Serbo-Croatian behaves exactly like Bulgarian. Basque offers an even more dramatic example of the point I am making on the basis of Bulgarian in the main text, as it displays superiority effects not only with respect to the first wh-element in the wh-cluster, but to all wh-phrases, rigidly ordering them all. See Jeong (2006) for data and discussion.
following Jeong (2006), I reformulate this as follows: only the first wh-specifier is a true specifier. All other wh-phrases are adjoined (hence their iterativity and lack of ordering). Since they are adjoined, they have no (negative) effect on featural exclusivity. In other words, multiple specifier constructions are only apparent (cf. Chapter 2). In making this claim, I follow Zwart (1997), who claimed that genuine instances of multiple specifiers should be excluded, not for phrase-structural reasons (in this, he and I concur with Chomsky 2000), but because they would run against the spirit of bi-uniqueness of syntactic relations, here expressed in terms of featural exclusivity/Unambiguous Merge.59

Featural exclusivity may also account for the obligatory movement of one of the members of small clauses consisting of two members of the same category. As Moro (2007) notes, Italian prohibits structures of the form in (40).

(40) *pro copula [Small clause DP DP]

Whenever a structure like (40) arises, one of the DPs must raise, as shown in (41).

(41) a. molte foto del muro sono t la causa della rivolta
    many pictures of-the wall are the cause of-the riot
    [Italian]
    ‘many pictures on the wall are the cause of the riot’
  b. la causa della rivolta sono molte foto del muro t
  c. *sono molte foto del muro la causa della rivolta

Moro (2007) refers to structures of the abstract type in (40) as “unstable,” or “too symmetric” (see also Moro 2000). Moro (2000) argues

59 Zwart (1997) and Grohmann (2003) provide arguments from various constructions in Germanic that multiple specifiers fail to adequately capture instances of multiple subject constructions. But their arguments differ from the one I made on the basis of multiple wh-fronting. Their arguments amount to showing that some facts are better captured by involving multiple functional projections hosting a specifier each, as opposed to fewer functional projections hosting multiple specifiers. The argument offered here points to a more subtle reason for why multiple specifiers don’t exist: far from denying the fact that multiple elements check their features against a single head, I show that whenever such checking situations arise (situations that fall under the rubric of Multiple Agree), one element is always singled out.
that the reason (40) is too symmetric is because both members of the small clause project—creating a point of symmetry that cannot be tolerated by PHON. According to him, movement breaks the symmetry, and saves the structure (39a,b). Moro (2007) offers a different account, namely, that small clauses of the type involved in (40) don’t project. Because they do not project, they fail to designate one of the two elements of the small clause as the Goal for Probing from the outside.

I think that Moro’s intuition that structures like (40) are too symmetric is on the right track. But I don’t think that the cause of symmetry is due to co-labeling (Moro 2000), since, following Chomsky (1995), I have argued that labels are unique, even in situations where they appear not to be (cf. the discussion of serial verbs above). I also do not think that the reason (40) is unstable is due to the bi-uniqueness of probing (situations of multiple Agree, of the type discussed by Hiraiwa (2001, 2005), Boeckx (2003b), and van Koppen (2005) suggest that probing can target multiple goals). Instead, I would like to combine parts of Moro (2000) and Moro (2007) and argue that the reason (40) is excluded is due to the fact that merging a DP to another DP results in a structure like \texttt{<DP DP>}, an instruction that causes PHON to gag, as it translates (in the present system) as something like “DP is the right edge of DP.” That is, the labeling in a structure like (40) fails to give rise to an informative, unambiguous output. Since no labeling option internal to Merge will succeed in (41), I follow the spirit of Moro (2000) and claim that movement is required. Since movement/chain-formation is equivalent labeling/projection (cf. Chapter 2), movement in (41a,b) will achieve the result that labeling internal to Merge couldn’t do. It will provide an unambiguous ordering of the two DPs. Movement is here superior to labeling, not intrinsically (they are equivalent), but because movement will have the effect of separating the two DPs by (at least) one other element, making them sufficiently distinct.

Let me conclude this discussion on distinctness by restating the most important result of this section. By construing labeling/projection as required to render Merge unambiguous, it follows that the maximum number of distinct edges and projection levels will be two.
The X-bar schema can thus be thought of as the maximal degree of diversification or expansion that syntax can achieve. Anything extra will result in ambiguity. Put differently, under QED-considerations advanced here to understand labeling, the X-bar schema amounts to the most optimal way of packaging information.\textsuperscript{60} The analysis of Merge developed here may thus have given us a concrete sense in which natural language syntax may be said to be an Optimal Conceptual Packager—a clear sign of good design of the type minimalism strives for.

The next chapters will build on the results achieved here to show how Optimal Packaging may be said to give rise to the locality conditions on chains and the cartographies that we find in natural language.

### 3.6 Conclusion

I began this chapter by stating that the role of syntax is that of an interface system, a hinge-like device\textsuperscript{61} that relates thoughts (SEM) and sounds (PHON). I spent some time drawing on the existing literature on PHON and SEM to establish that, in terms of dimensionality, PHON- and SEM-representations are poorer than syntactic objects. The mapping from syntax to either PHON or SEM must therefore take the form of “projection” (in the geometrical sense), i.e., dimension reduction. Such a mapping is known to require orientation. I went on to argue that for the mapping to be efficient, and to avoid attributing to the external systems the kind of computational power that we suspect they do not have, I have argued that

\textsuperscript{60} Wagner (2005, 2007) equates this with the Super-Catalan number. Boeckx, Carnie, and Meideros (2006) relate maximal expansion of the X-bar schema to the Fibonacci series, a well-known case of optimal packaging, as Kepler had already noted (see Uriagereka 1998 for extensive discussion of the relevance of the Fibonacci series in a linguistic context).

\textsuperscript{61} I owe the hinge metaphor to Juan Uriagereka. Uriagereka (2007) uses the image of a hinge to characterize clitics, which, like syntax as a whole, are minimal, simple objects whose shape has remained remarkably constant over language change. This is perhaps because the range of possible solutions to the problems clitics solve is so restricted that good design leaves us with just one possibility.
syntax itself provides properly oriented objects in the form of labeled Merge products, what I called *Unambiguous Merge*. Although the objects being spelled-out must be asymmetric, it would be wrong to stress the asymmetric character of syntax to the exclusion of the fundamental symmetry of symmetric rules like Merge (the key to locality, if I’m right in Chapter 2). Asymmetry is in the service of Anti-ambiguity. Syntax must ensure that if more than one element gets mapped onto PHON or SEM, the set of elements is aligned (/ordered) properly. I claimed that the distinctness condition on Merge products is achieved by Agree. Feature-checking is thus at the heart of the syntactic engine of our language faculty.

Chomsky has often remarked (see, e.g., Chomsky 2000: 101) that unlike Merge, Agree appears to be a fundamentally language-specific operation, and I think he is exactly right in speculating on the basis of this fact that Agree “relates to the design specifications for human language.” But we need not see Agree as an operation that appeared *de novo*, “without significant analogue elsewhere.” Genuine novelty, in the sense of emergence of completely new processes, is extremely rare in the biological world. Nature, as Jacob famously pronounced, is a tinkerer. By this I think Jacob really wanted to emphasize the idea that novelty in the organism’s physiology, anatomy, or behavior arises mostly by the use of conserved processes in new combinations, at different times, and in different places and amounts, rather than by the invention of completely new processes.62 This is what Darwin meant by his term “descent with modification.” Germans would say that novelty is characterized by *Um-bildung* (“re-formation”), not by *Neu-bildung* (“new formation”)—topological variations, not introductions of novel elements. As Gould (1977: 409) said, “there may be nothing new under the sun, but permutations of the old within complex systems can do wonders.” This is exactly what I have in mind regarding Unambiguous Merge, the product of two generic operations (unrestricted Merge, a grouping process, and Agree, a pointer), whose

---

62 I do not think that tinkering necessarily implies that biology is the kingdom of the “law of the higgledy-piggledy.” See Boeckx (2006a: 131ff.) for relevant discussion.
combination may be rare because the problem facing human lan-
guage—facing sound and meaning—may be rare. As Ernst Mayr
never tired of saying (see especially Mayr 2004: 34), the locus of bio-
logical specificity very often finds its source in physiology. Combining
Merge and Agree is the solution to optimize the parsing of products of
Merge. From this perspective, what had to evolve was the unvalued/
valued distinction in the lexicon (concepts had to become digital),
which enabled the grafting of Agree onto Merge. Once this was
achieved, syntax could become the great concept-integrator, unleashing
the power of cross-modular combinations that characterize our spe-
cies’s mental world.

Evolutionary considerations aside, my goal throughout this chap-
ter has been to rationalize some fairly well-established aspects of
Merge, traditionally coded in the X-bar schema. Building on
Chomsky’s bare phrase structure, I have suggested that X-bar con-
straints are not intrinsic to Merge, but follow from the topography
of mind (the architecture of the language faculty) coupled with
considerations of mapping efficiency. It is true that I have had to
enrich Merge (incorporating Agree into unrestricted Merge), but
the road to unification taking us beyond explanatory adequacy often
involves adopting a slightly richer set of basic primitives that allow
for greater symmetry in the formulations of fundamental relations.
The next two chapters show that the enrichment of Merge argued for
here allows one to unify basic projections, cartographies, and
chains—offering the basis for a general theory of locality.
The goal of this chapter is straightforward: show that the rich arrays of functional projections that the syntactic literature continues to reveal under the rubric of “cartographies” organize themselves as standard phrases do. Put differently, my goal in this chapter is to show how cartographies are nothing more than “blown-up” projections, subject to the same constraints as regular projections. Once they are seen as projections, cartographies can be said to function as the basic units of selection without us having to depart from the claim that the simplest kind of locality—adjacency—is the right one in this domain. In other words, if cartographies reduce to projections, the deduction of locality suggested in Chapter 3 in terms of QED for basic Merge products can be carried over: locality will be preserved under symmetry.

As I did for basic Merge, I will present my argument in two steps. I will first present the relevant generalizations in familiar X-bar terms, and then (beginning with Section 4.7) resort to the more stringent bare phrase-structure terminology in an attempt to provide a more minimalist motivation for the generalizations.

4.1 Core issues

I regard the cartographic approach as one of the great success stories in modern linguistic theory, a success story still very much in the making. It all started with Chomsky (1986b) adopting proposals by
Tim Stowell, David Pesetsky, and Abdelkader Fassi-Fehri about extending the X-bar schema to elements like Inflection and Complementizer, giving rise to IP and CP. This first step was soon followed by Pollock’s (1989) influential split-Infl hypothesis and Larson’s (1988) shell-analysis of VP. The past decade has seen a proliferation of functional projections, a growth of the exponential sort that took its mature form in Rizzi (1997) for the CP-space, Baker (2001b) for the VP-space, and even more dramatically in Cinque (1999) for the IP-space.

The empirical coverage of the cartographic project cannot be overlooked. Recent works like Julien (2002) and Cinque (2005) manage to uncover morphosyntactic regularities holding of hundreds of historically unrelated languages. But the cartographic approach raises non-trivial questions, which I list here because they will guide the investigation in this chapter, together with the questions raised about cartographies in Chapter 1.

The first question is whether all projections have the same status. Beyond the lexical/functional divide (Abney 1987; Fukui 1986), it is not clear whether any further subdivision is needed. Den Dikken (2006) suggests that some projections act as Relators (projections headed by categories mediating a predication relation between two terms), others as Linkers (projections headed by categories that host moved relators, rendering predicate inversion possible). But aside from its descriptive virtue, the linker/relator subdivision within functional projections lacks a solid conceptual basis (Why two? Why these functions?).

Baker (2003: 325) claims that functional projections subdivide into “transparent” and “opaque” projections. Transparent functional projections sit right on top of lexical projections and extend these, whereas opaque functional projections cause a category shift (from, say, a verbal projection series to a nominal projection series). Like den Dikken’s, Baker’s claim is also very descriptive at the present stage, but I will come back to his proposal later on in this chapter, where I will show how it fits conceptually with core propositions made here.

Another question that arises in the realm of projections is whether there are “basic” or “core” projections. Chomsky has repeatedly used
the term (see, e.g., Chomsky 2000: 102) to refer to traditional projections (including CP and IP), which he uses as “shorthand” for the novel projections that have been argued to emerge once CP and IP are split (TopicP, FocusP, etc.) (see Chomsky 2001: 43 n. 8, 2002: 123, forthcoming: 10). Chomsky, however, doesn’t provide any definition of what a core projection may be. So it is not clear whether the term has any theoretical meaning. But if we agree that it has intuitive content (as I think it does), it may be interesting to see if it can be formalized. Pursuing the consequences of exploded functional domains like CP and IP, it has become harder to determine the boundaries of certain syntactic spaces or zones. For example the content of lowest projection in Rizzi’s exploded CP-domain, FinitenessP, matches that of TP/IP (the topmost functional domain immediately below the CP-space) fairly closely (see Rizzi 1997). One would certainly like to know what this means theoretically. That is, do we need both FinitenessP and TenseP?

All these questions really revolve around whether there are organizational principles behind the series of projections we find. And if there are, how many principles do we need to characterize cartographies? And, are the principles confined to this domain of syntax?

Addressing these questions now presupposes that our understanding in the domain of projection has progressed enough so as to make these questions ripe for a serious investigation. I believe that this is the case, although I cannot fail to mention that we don’t yet have a good sense of how many projections there are. If our current guesses are widely off the mark, they may bias us toward answers that will prove to be equally misguided. But this is a common state of affairs in empirical science. One must start somewhere.

4.2 The basic pattern

My starting point in my attempt to make sense of projection series will be the X-bar schema, understood as emerging from bare considerations of the type discussed in Chapter 3.
Because it makes it easier to talk about the relevant generalizations, I will use the traditional X'-notation first. Then I will revert to the more austere framework developed in Chapter 3. Until I do so, the reader should bear in mind that whenever I talk about the projection line \{X^\circ, X', X''\}, the latter is “emergent”: the result of the establishment of two basic oriented relations. That is, (1a) is really (1b).

```
(1)  a. \[
\begin{array}{c}
\text{YP} \\
\text{WP}
\end{array}
\quad b. \[
\begin{array}{c}
\text{YP} \\
\text{WP}
\end{array}
\end{equation}
```

The mechanism of projection at the heart of the X-bar schema makes room for three linguistically significant relations: complement-of, specifier-of, and adjoined-to. In Chomsky (1970), these relations were rigidly defined structurally: Complement-of was the sister of \(X^\circ\) and daughter of \(X'\); adjunct-to was the sister of \(X'\) and daughter of \(X''\); and specifier-of was the sister of \(X'\) and daughter of \(X''\). The figure in (2) reproduces the original X-bar schema as just described.

```
(2) \[
\begin{array}{c}
\text{Spec} \\
\text{(Adj)} \\
\text{X}
\end{array}
\quad \begin{array}{c}
\text{X'} \\
\text{X''}
\end{array}
\quad \begin{array}{c}
\text{Compl}
\end{array}
\] (Chomsky 1970)
```

The structural status of specifier-of and of complement-of is very rigid, but the positioning of adjoined material fluctuates. While Chomsky took adjoined material to be the sister of \(X'\) and daughter of \(X'\) (2), May (1985) argued that adjoined material should be added to the \(X''\) level, not the \(X'\) level, as represented in (3). (A proposal adopted in Chomsky (1986b).)
Later on Larson (1988) famously argued that adjoined material occupies the innermost “complement” position, and that traditional complements are specifiers of a projection (shell) dominating the projection hosting adjuncts. (Since Larson limited the number of relation per phrase to one, his representations take us beyond the domain of the single phrase, and will not be reproduced here.) (For a recent reappraisal of his original position on adjuncts, see Larson 2004.)

While the uncertainty concerning the positioning of adjoined material is often seen to reflect the fact that we still lack an adequate theory of what adjunction amounts to, the mobility of adjuncts can be captured if we adopt the idea that adjuncts are in some sense in a parallel dimension (see Chomsky 2004; Lasnik, Uriagereka, and Boeckx 2005; see also Chapter 3). Their mobility would then be the result of the fact that once parallel dimensions are collapsed at the interfaces, adjoined material can be integrated into the X-bar schema at three juncture levels: $X^0$, $X'$, and $X''$, as schematized in (4).

Setting adjunction aside, what seems constant across the various developments of the X-bar schema is the conclusion that there are two significant levels of projection, in addition to the minimal state
of an element: X’ and X” in addition to X°. To put things differently, an element X can occupy the following three states (“projection levels”): (i) minimal, (ii) intermediate, and (iii) maximal, or unstretched, stretched, and maximally stretched.

This, as far as we know, is valid for all syntactically relevant categories, lexical and functional alike. It gives rise to the beautiful symmetry achieved in the mid-1980s, as represented in (5), and already discussed in Chapter 1.

\[
\begin{array}{ccccccc}
V’’ & N’’ & A’’ & P’’ & I’’ & C’’ \\
V’ & N’ & A’ & P’ & I’ & C \\
V & N & A & P & I & C \\
\end{array}
\]

Verb  Noun  Adjective  Preposition  Inflection  Complementizer

Grimshaw (1991) extended this symmetry by claiming that in addition to categories being able to “stretch” themselves (call this Intrinsic Projection), lexical categories can also be functionally extended (call this, as Grimshaw did, Extended Projection; see also van Riemsdijk’s (1998) concept of “macro-projection”). Remarkably, Grimshaw discovered that Extended Projections organize themselves the same way Intrinsic Projections do. They consist of a minimal (functionally unstretched) layer, a first level of functional stretching (an intermediate extended layer), and a final, maximally extended functional layer. According to Grimshaw, C(omplementizer) and I(nflection) are extended projections of V(erb); P(reposition) and D(eterminer) are extended projections of N(oun), as represented in (6).\(^1\)

\[
\begin{array}{cc}
C & P \\
I & D \\
V & N \\
\end{array}
\]

\(^1\) But see Hiraiwa (2005) for instances of criss-crossings of extended projections, which he grounds in the kind of symmetry considerations congenial to those at the heart of this study.
In recent years, lexical categories have come under attack, and have been gradually replaced by a-categorial roots, which get categorically specified contextually, by combining with “light” functional heads (also called semi-functional heads), the best known of which is “little v” (Chomsky 1995; Kratzer 1996). But Grimshaw’s insight has been preserved under this redefinition of lexical categories, with “little” categories now forming the first functional layer of Extended Projections. The representations in (7)–(10) illustrate this readjustment within Extended Projections, adopting node labels that are frequently used in the relevant literature. The specific node-labels used in (7)–(10) don’t matter much. What is significant is that the tripartite hierarchical organization observed at the phrase level (intrinsic projection) is retained at the level of functional extension. Even headedness is retained (albeit in an “extended,” more abstract sense), as van Riemsdijk (1998) points out (specifically, features shared among functional layers constitute the type of the whole projection).

\[(7)\]
\[
\begin{align*}
\text{V} = & \text{IP} \\
& \text{vP} \quad \text{(aka “VoiceP”)}
\end{align*}
\]

\[(8)\]
\[
\begin{align*}
\text{N} = & \text{NumberP} \\
& \text{nP} \quad \text{(aka “ClassifierP”)}
\end{align*}
\]

\[(9)\]
\[
\begin{align*}
\text{P} = & \text{PlaceP} \\
& \text{pP} \quad \text{(aka “AxialP”)}
\end{align*}
\]

2 The three-way division seen in (7)–(10) has been argued to define three checking domains across categories—an Operator domain, a Modifier domain, and a Thematic domain—which some (see di Sciullo 2005; Schweikert 2005; Grohmann 2003; Platzack 2001) have argued is a significant fact about how traditional units like the “clause” are built, and how derivations proceed—a claim I return to in Chapter 5.
The reader should understand that throughout this chapter I will be unable to reproduce the empirical evidence brought to bear on the existence of the rich array of projections under discussion. In the literature, projection series (“cartographies”) are often used to analyze the size of traditional lexical categories, and their internal composition, where a rigid ordering of morphemes invariably holds. The composition of “extended projections” can rarely be detected on the basis of simple examples. Very often, the cartography must be inferred from logical relations among minimal pairs such that if in minimal pair 1, A dominates/precedes B, and in minimal pair 2, B dominates/precedes C, then the structure A > B > C is said to underlie both sets of examples, even if no single example containing A, B, and C may occur. (For the most comprehensive application of this line of reasoning, see Cinque 1999.)

To help the reader navigate through the many representations discussed in this chapter, I will often try to reproduce a key example or two to simply flesh out the projections under discussion, and make them somewhat more concrete. But I stress that such examples are used for illustrative purposes only. They are not meant to convince the reader that the representations they exemplify are the correct ones, nor are they good substitutes for the complex, and typologically rich reasoning that characterizes cartographic studies. I therefore urge the reader to consult some of the sources mentioned in the context of each representation to find out more about them. My aim here is simply to find a way to organize the results coming from the cartographic literature, and, if successful, provide a guide for subsequent analyses.

With this caveat in mind, let me illustrate the structures in (7)–(10) by means of the examples in (11)–(14). (For a sample of studies on verbs, see Cinque 1999 and Julien 2002. On adpositions, see van Riemsdijk 1978, Koopman 1997, den Dikken 2003, Ramchand and
The existence of extended projections for the basic lexical categories N, V, A, and P is relatively well-established in the literature. But I would like to claim that the pattern just discussed does not stop at the level of Extended Projections. Specifically, I intend to argue that the tripartite organization that defines the X-bar schema is much more pervasive than previously thought (although I stress that the evidence to show this already exists in the literature; it just needs to be pulled together, which I try to do here).

4.3 X-bar everywhere

I suspect that each extended projection layer in (7)–(10) organizes itself along X-bar lines. Because we don’t yet know enough about some of these layers, I cannot show this to be true in all cases, but my suspicion is based on the fact that whenever a layer is submitted to closer scrutiny, the very same pattern emerges, and is clearly reminiscent of the X-bar schema.

Take CP, the topmost layer of the verbal Extended Projection in (7). As already mentioned at various stages in this study, Rizzi (1997) has famously argued that CP should be regarded as an abbreviation for a rich functional domain consisting of at least the four projections in (15). (See also Bhatt and Yoon 1991.)

(15) \( CP = \text{[ForceP [TopicP [FocusP [FinitenessP]]]]} \) (Rizzi 1997)

Evidence for two complementizer nodes comes from “complementizer” repetitions of the type seen in the Galician example in (16),
which also reveals the presence of a Topic node sandwiched in between.³

(16) dixeron que a este home que non o maltratemos [Galician]
said.3.pl that to this man that not him badly.treat.1.pl
‘they said that this man that we should not treat him badly’
(Raposo and Uriagereka 2005: 646)

Evidence for a Topic-Focus articulation comes from data like (17).

(17) Un libro di poesie, A GIANNI, lo regalerete [Italian]
A book of poems, to Gianni, it you.will.give
‘You will give a book of poems to Gianni’
(Beninca’ and Poletto 2004: 54)

What is particularly interesting for the claim I want to make in this chapter is that Rizzi (1997: 288) explicitly proposes the projections in (15) don’t all have the same status. According to him, ForceP and FinitenessP form the core of the “left periphery” (i.e., CP-zone). They are always present whenever we thought CP was present. By contrast, TopicP and FocusP are projected only when needed to host topically or focused material. Accordingly, if an embedded sentence introduced by a complementizer does not contain a topic or a focus (e.g., John knows [that Mary left]), only two of the projections in (15) will be present: ForceP and FinitenessP. FocusP and TopicP are, in some sense, optional. Accordingly, (15) could be represented as (18).

(18) [ForceP [(TopicP-FocusP) [FinitenessP]]]

Furthermore, like adjoined material, the TopicP/FocusP space accommodates multiple elements of the same kind, in the same way that adjoined material can stack.

(19) A Gianni, domain, QUESTO gli dovremmo dire! [Italian]
To Gianni, tomorrow, this to-him we.should tell
‘We should tell this tomorrow to Gianni’
(Beninca’ and Poletto 2004: 55)

³ For examples of complementizer reduplication in English see Franks (2005). Co-occurring complementizers are also found in languages like Polish (Szczegielniak 1999) and Dutch (Zwart 1996).
(18) already looks a lot closer to the original X-bar schema in Chomsky (1970), with optional material sandwiched at the X’-level (cf. (2)):

(20) \[ x'' \text{Spec-relation} [x' (Adj.) [x' \text{Complement-relation}]] \].

Subsequent research on the left periphery (in particular, Beninca’ (2001) and Beninca’ and Poletto (2004)) has brought (18) even closer to the X-bar schema, by arguing that the optional unit (TopicP-FocusP) can be “iterated.” Hanging topics have been argued to be projected above ForceP (21) (see Beninca’ 2001), and aboutness phrases (Rizzi’s 2006 SubjP) and lower topic/focus material have been argued to sit below FinP (see already Lasnik and Saito 1992; see also Bošković 1997; Maki, Kaiser, and Ochi 1999; Endo 2006) (22).

(21) (sono certa) questo libro, che non ne abbia mai parlato nessuno

‘I am certain that nobody has ever spoken about this book’

[Italian] (Beninca’ and Poletto 2004: 65)

(22) Kimi-wa John-ga nani-de-nara sono kuruma-o

You-Top John-Nom what-with-Top the car-Acc

naoseru-to omowa-nai-no?

can.fix-Comp think-Neg-Q

‘In which place don’t you think that John can fix the car?’

(Endo 2005: 28)

Putting all these results together, we arrive at a representation like (23).4

(23) [(Top/FocP) [ForceP [(Top/FocP) [FinP [(Top/FocP)]]]]]

The mobility of optional material witnessed in the left periphery essentially replicates what we reviewed above in the context of

---

4 I am well aware that subtle interpretive differences have been argued to exist between, say, TopicP above ForceP, and TopicP below FinP. (For a very interesting study revealing such subtle differences, see Beninca’ and Poletto 2004.) But in my opinion such subtle contrasts arrive at SEM. Structurally, TopicP above or below ForceP is the same. Referring to them by means of different, semantically richer node labels can be useful at some level of analysis, but it would obscure the patterns I am trying to reveal in this chapter. (For a similar point on AGRs and AGRo being “mnemonics” for what is essentially the same AGR-node, see Chomsky 1991.)
adjunction and the integration of optional material into the X-bar schema. (23) is essentially the same abstract representation as (4).

Similar results obtain in the thematic layer of the clause, the vP-domain, the lower level of the V-extended projection. The vP-domain is now standardly seen as consisting of at least two layers, vP and VP, hosting the two core arguments: the external and internal arguments, respectively (see Larson’s (1988) “One-argument-per-phrase” restriction). Optional, “oblique” arguments (often called extra- or applied arguments) have been integrated since Marantz (1993) via an extra projection (call it ApplP), sandwiched between vP and VP, as seen in (24). (25) illustrates this structure. (See Jeong (2007) for examples of stacked applicatives.)

\[(24)\] \[vP = [(vP [(ApplP)[VP]])]\]

\[(25)\] Mavuto a - na - umb - ir - a mpeni mtsuko [Chichewa]
Mavuto SP-PST- mold-APPL-ASP knife waterpot
‘Mavuto molded the waterpot with a knife’ (Baker 1988: 230)

But recent research into the nature of applicatives have made a convincing case for multiple insertion points for ApplP, arguing for the existence of low applicative heads (below VP) (26), high applicative heads (above VP and below vP) (27), and even super-high applicatives (above vP) (28). (On high/low applicatives, see Pylkkänen 2002, McGinnis 2003, Jeong 2007; on super-high applicatives, see Buell 2003, Ramchand forthcoming, Rivero 2004, Sedighi 2005).\(^6\)

\[(26)\] Umugóre a-rá-som-er-a umuhuúngu igitabo [Kinyarwanda]
woman SP-PR-read-APPL-ASP boy book
‘The woman is reading a book for the boy’ (Jeong 2007: 7)

\[(27)\] John gave Mary a book

\[(28)\] I-sikole si-fundela a-bantwana [Zulu]
7-7.school 7.sbj.study.appl.fv. 2-2.child
‘The children study at school’ (Buell 2003: 109)

\(^5\) Huang (1997) calls the applied object the inner object to distinguish it from the outer argument (external argument) and innermost object (internal argument).

\(^6\) Here too there exist interpretive differences among the various Appl-nodes, but, as Pylkkänen (2002) argues, these can be understood as the result of compositional analyses. They need not force us to posit different lexical items “High-Appl” and “Low-Appl.”
The net result of this research is a maximally articulated thematic domain like (29).

(29) [(ApplP) [vP [(ApplP) [VP [(ApplP)]]]]]

Again, the similarity between this structure (two core relations and a third, more mobile relation introducing extra material) and the X-bar schema is quite clear, and cannot be purely accidental. Indeed as soon as we examine the cartographic literature from the present perspective, we find the X-bar organization literally everywhere. Virtually each and every syntactic entity employed at one point in the syntax literature has been, or can be, decomposed into what amounts to a hierarchically organized three-member set of heads or features. (In cases where it hasn’t been done, I would argue that it should be done. Part of the message of this chapter is that the X-bar schema provides a research guide for cartographic analyses.)

Let me discuss a few more examples. Those discussed so far (CP and vP) provided examples of projection series with two clear anchor points, and projections hosting optional material shifting around them. But the cartographic literature also contains very clear examples of projection series very much like Grimshaw’s original notion of extended projection (albeit at a finer-grained level of analysis), with a clear progression from a minimal to an intermediate to a maximal state, where the intermediate state is not “optional” (like TopicP or ApplP), but “halfway” between the beginning point and the end point. (Crucially, in this case, unlike what we saw with TopicP/ ApplP, the intermediate point does not iterate. Each projection level is unique, in accordance with the considerations in Chapter 2.)

Ramchand (forthcoming) has proposed that the event structure (Aktionsart) established at the “VP” level can be decomposed into three layers of embedding, beginning with an Initiating Event phrase (InitP), going through a Process Event phrase (ProcP), and terminating in a Result Event phrase (ResP).

(30) [InitP [ProcP [ResP]]]

(31) John danced his way out of the room
In the same vein, Inflection ("IP") is now often decomposed into a Mood layer, containing a Tense layer, itself containing an Aspectual layer, as shown in (32). (Chomsky 1957; Cinque 1999).

\[(32) \, IP = [\text{MoodP} \, [\text{TenseP} \, [\text{AspectP}]]]\]

Similarly, agreement ("AGRP") (the "other" projection of IP, cf. Pollock’s (1989) TP-AGRP structure) standardly decomposes into a Person layer, Number layer, and Gender (or Class) layer. (See Ritter 1991, 1993; Sigurðsson 2006; Shlonsky 1989, 1997.)

\[(34) \, AGRP = [\text{PersonP} \, [\text{NumberP} \, [\text{GenderP}]]]\]

Gradable Adjectives (DegreeP) likewise decompose into a tripartite structure.

\[(36) \, [\text{SuperlativeP} \, [\text{ComparativeP} \, [\text{NeutralP}] auspke)]\]

The representation in (36) won’t surprise anyone, any more than Ramchand’s event structure in (30) would. Both are remarkably iconic. Just like an X-bar schema has a beginning, a transition point, and an end, so too do events. And just like an X-bar schema contains a minimal, intermediate, and a maximal degree of projection, so too do gradable adjectives. In this sense, the X-bar schema reflects SEM, and the former may be bounded by the conceptual resources of the latter (there is no higher degree than the superlative, so nothing projects above SuperlativeP). This is, in a sense, what I argued in Chapter 3: Merge itself is not upward bounded, but the resources of the external systems help define X-bar strictures on a “raw” or “bare” phrase structure. But as Bobaljik (2007) argues, part of the significance of (36) (and, I’d argue, (30)) as a syntactic representation is that its nesting structure becomes available to PHON, and
can therefore be used (as Bobaljik does) to explain morphological generalizations that would have a much more stipulative character if (36) were dismissed as mere conceptual structure. I will come back to such considerations at various points in this study.

Other tripartite structures that I have been able to identify have been argued for in the realm of quantifier expressions (38) (Heim 1982; Herburger 2000; di Sciullo 2005), pronouns (39) (Cardinaletti and Starke 1999; Dechaine and Wiltschko 2002), (non-verbal) stative predicates (40) (Roy 2006, 2007), and temporal structures (41) (Hornstein 1990; Demirdache and Uribe-Etxebarria 2000; Giorgi and Pianesi 1997).

(38) a. [Operator [Restriction [Variable]]]
   b. WhOp+[Rest atE] bookRest

(39) a. [Strong Pronoun [Weak Pronoun [Clitic]]]
   b. lui; le; l’ [French]
   ‘HIM, him, ’m’

(40) a. [Defining Predicate [Characterizing Predicate [situation-descriptive predicate]]]
   b. Paul est absent/acteur/un acteur [French]
   ‘Paul is away/actor/an actor’

(41) a. [Speech Time [Assertion Time [Event Time]]]
   b. John had been building a house

Reflexes of the X-bar schema may also be found in the following categories (although, as far as I know, explicit nested structures in these domains haven’t been explicitly posited):

(42) Nominal categories corresponding to the three semantic types in Chierchia (1984): [ <<e, t>, t> N [ <<e, t> N [ <<e> N] ] ]

(43) [independent [dependent [anaphoric]]] Tense (Landau 2004)

(44) Three types of reflexives, of the type found in Dutch (Koster 1985):
    [pronoun-zelf [zichzelf [zich]]]

(45) [R-expression [Pronoun [Anaphor]]] (Chomsky 1981)

Roy (2006) also characterizes the three classes of (non-verbal) stative predicates as giving rise to dense (situation-descriptive), non-dense (characterizing), and maximal (defining) readings. Roy argues that adjectives can only have a dense reading. Bare nouns can have a non-dense reading, but must be used with an (indefinite) article to be able to express a maximal reading. Here too, then, morphosyntactic complexity, visible at PHON, mirrors semantic complexity (SEM)—a clear case of mapping transparency (in the sense of Uriagereka forthcoming).
In the latter three cases, the hierarchical structure of elements may correspond to their ever-larger licensing domains. It may even be deducible from the structure of their licensors.

### 4.4 Extension by licensing

Cases of what I’d call “extension by licensing” arguably characterize a variety of units such as Negation, which decomposes into \{metalinguistic negation, \{sentential negation, \{constituent-negation\}\}\} (Cormack and Smith 2002), corresponding to the articulation of the clause [CP [IP [VP]]]. The same clausal architecture could be used to account for:

- the existence of three types of null pronominals: a thematic null pronoun (control PRO), a null subject of the Italian null-subject type (pro), and a null topic (the Chinese-type pro discussed in Huang 1984), which obviously correspond to \{VP, IP, CP\}
- the existence of three types of subjects (McCloskey 1997): a post-verbal subject (also called “low” or “in-situ” subject), a prototypical A-raised subject, and an A-bar, dislocated subject of the type argued for in Romance-style pro-drop languages, corresponding again to \{VP, IP, CP\}
- the existence of three types of resumptive elements: self-anaphors at the VP-domain (Hornstein 2001; Grohmann 2003), (weak) pronouns at the IP-domain (Boeckx 2003a), and (strong) pronouns/epithets at the CP-domain (Aoun and Choueiri 2000; Boeckx 2003a)
- the existence of three kinds of prepositions: prepositions that take arguments which they Case-check, prepositions that Case-check traditional Case assigners, and prepositions that function as Case-markers and do not undergo Case-checking (Bošković 2006)
- the existence of three types of PP-modifiers (manner < locative < temporal PP), corresponding to VP, AspectP, TP (Schweikert 2005)
- the existence of three kinds of classifiers, corresponding to the \n level (Bantu classes), the individuation level, NumberP (Borer 2005), and the determiner level, DP (Cheng and Sybesma 1999)

All these elements come in triplets of increasing size because they “connect” to three licensing regions that stand in an embedding relation.
One should not be surprised, from the present perspective, to find that Binding-Theoretic principles amount to three (Principle A, B, and C^8), since they regulate DPs that come into three kinds (anaphors, pronouns, and R-expressions). Likewise, it should be no surprise to find that the orientation of reflexives revolves around three notions: co-argument, subject, and point of view (see the papers in Cole, Hermon, and Huang 2001), corresponding pretty closely to VP, IP, and CP.

We could, of course, treat all these layers and regions of grammatical information as constituting distinct modules of the grammar (maybe even distinct modules at SEM), but, as Baker (2001b) has argued in the domain of thematic hierarchies, saying that all these “semantic” hierarchies or series reduce to a common phrase-structural representation (the X-bar schema) makes the substantive claim that relational prominence translates into embedding prominence, with consequences pertaining to morphosyntactic behavior that would be purely accidental (or, at least, much harder to explain) if the hierarchies at issue didn’t share the same representational format (cf. Bobaljik’s (2007) point concerning gradable adjectives mentioned above).

### 4.5 Iterated patterns

It may in fact have been the case that originally unstructured, “generic” concepts “acquired” structure or got diversified by virtue of being grafted onto independently generated structure. That is, once “thrown” into the syntax engine, a concept can be iterated by Merge, and find itself mapped onto a variety of hierarchies at SEM (and PHON). This may have been how “new” concepts originated (more appropriately, how concepts acquired new modes of presentation). (Hinzen 2007 would characterize this state of affairs as language having been ontologically innovative.)

^8 Alternatively, Principle A, B, and D (see Lasnik 1989).
For instance, consider the fact that the very same “concept” (or conceptual realm) may acquire different “flavors” depending on where it is expressed in the structure. Take the concept “Person.” Say the latter gets lexicalized (/digitized) as the abstract feature \( \pi \) with its three possible values \{+, −, \( \varnothing \)\}. \( \pi \) clearly manifests itself in all three portions of the clause (V’s extended projection), with an orientation that is characteristic of Merge. Thus, at the thematic/\( vP \) level, \( \pi \) becomes argument structure: Agent [point of origin], Patient [end point], and “Other” [“adjoined”]. At the Inflectional level, \( \pi \) becomes 1st person [point of origin], 2nd person [target], and 3rd person [“adjoined”]. Finally, at the Complementizer level, \( \pi \) encodes Speech roles (Speaker [origin], Hearer [target], Logophor\(^9\) [other]), of the type studied by Speas and Tenny (2003) to characterize the encoding of evidentiality, morphologically expressed in many languages (on speech roles, see also Sigurðsson 2004). The claim I just made about \( \pi \) is very similar to Hornstein and Uriagereka’s (2002) claim (cf. Chapter 2) that the structure of quantifiers, with their scope and restriction, may be nothing more than argument structure (a transitive verb) at a higher level of structure (see also Pietroski 2005, Herburger 2000, and Larson 2005).

As a matter of fact, several authors have already observed that various units such as Topic, Focus, and Agreement appear to be expressed at various clausal junctures (C-level, Inf-level, and V-level) (see Hallman 2000; Sportiche 1998; Koopman 2005; Belletti 2004a, 2005; Jayaseelan 2001; Manzini and Savoia 2004; and Poletto 2006). For example, what was once AGRP (Pollock 1989) was soon decomposed into AGRsP (for subject agreement), AGRoP (for object agreement), and “AGRcP” (for complementizer agreement (Shlonsky 1994; Cardinaletti and Roberts 1991, 2002)). This amounts to treating AGRP as an iterated category, along the clausal skeleton (extended projection of V), in a way very reminiscent of the distribution of adverbs in the basic X-bar schema (cf. (4)).

\(^9\) Interestingly, the category “logophor” again decomposes into \{Source, \{Self, \{Pivot\]\]\} (based on Speas and Tenny 2003).
Perhaps the most dramatic instance of such an iterative pattern is the one identified by Butler (forthcoming) in the context of the Cinque hierarchy.

Cinque (1999) put forth the hierarchy in (47) to capture co-occurrence restrictions among adverbs and inflectional heads—a hierarchy that Cinque stresses is only partial and temporary.\(^{10}\) (I have tried to illustrate each functional head by associating it with an adverb exemplifying it.)

\[
(46) \quad \text{CP} \leftarrow \text{AGRcP} \\
| \\
\text{IP} \leftarrow \text{AGRsP} \\
| \\
\nuP \leftarrow \text{AGRoP}
\]

At first the hierarchy in (47) appears to have little to do with the tripartite divisions discussed so far in this study, but as Butler (forthcoming) notes, Cinque’s labels hide the iterative pattern which Butler characterizes as in (48).\(^{11}\)

\[\text{[frankly Mood_speech act [fortunately Mood_evaluative [allegedly Mood_evidential [probably Mood_epistemic} [\text{once } T_{\text{past}} [\text{then } T_{\text{future}} [\text{perhaps } \text{Mood_irrealis [necessarily } \text{Mod_necessity } [\text{possibly } \text{Mod POSSIBILITY [usually } \text{Asp_habitual [again Asp_repetitive(I)} [\text{often } \text{Asp_frequentative(I)} [\text{intentionally } \text{Mod_volitional [quickly Asp_celerative(I)} [\text{already } T_{\text{anterior}} [\text{no longer } \text{Asp_terminative} [\text{still Asp_continuative} [\text{always Asp_perfect(?)} [\text{just Asp_retrospective} [\text{soon Asp_proximative} [\text{briefly Asp_durative} [\text{characteristically(?) Asp_generic/progressive} [\text{almost Asp_prospective} [\text{completely Asp_Sg_Completive(I)} [\text{tut} (\text{Italian)) Asp_Sg_Completive [\text{well Voice [fast/early Asp_celerative(II)} [\text{again Asp_repetitive(II)} [\text{often Asp_frequentative(II)} [\text{completely Asp_Sg_Completive(II)} [\text{tutto (Italian))] )))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))\]
\]
\]

\[\text{Cinque 1999: 106}\]

\(^{10}\) For an attempt to capture these restrictions from basic semantic principles, see Fortuny (2006).

\(^{11}\) The claim that Cinque’s hierarchy may be an iteration of a tripartite division was made by Enç (2004) on the basis of Turkish data. Thanks to Meltem Kelepir for bringing this work to my attention.
From Butler’s perspective, the now familiar X-bar schema reappears. Cinque’s hierarchy boils down to an instance of the Mood/Tense/Aspect triplet, used iteratively, with the Aspectual layer “thicker” at the bottom (closer to the VP-domain, traditionally associated with Aspect; see Borer 2005 and Ramchand forthcoming, among many others), and the Modal layer “thicker” at the top (closest to CP, traditionally associated with Mood; see Rizzi 1997, among many others).12

12 Adjectives inside nominals could be handled in a similar fashion. Scott (2002) proposes a very rich hierarchy of adjectival modifiers inside the noun phrase (i).

(i) Ordinal > Cardinal > Subject comment > Evidential > Size > Length > Height > Speed > Depth > Width > Temperature > Wetness > Age > Shape > Color > Nationality/Origin > Material

Laenzlinger (2005) shows on the basis of French data that Scott’s hierarchy (with some refinements) can be cut down to roughly five classes, reproduced in (ii).

(ii) [Quantif] Ordinal > Cardinal > [Speaker-Orient, Subject comment] > Evidential > [Scalar Phys. Prop. Size] > Length > Height > Speed > Depth > Width > [Measure Weight] >
A similar phenomenon of iteration may be at work in the context of negation. The most comprehensive cartographic study of negation I am aware of is Zanuttini (1997). Zanuttini examines the locus of negative expressions in Romance languages, and finds quite a bit of variation, ultimately leading her to postulate four distinct Negation heads in the clausal structure. Specifically, taking the inflected verb (in Infl‘) as her reference point, Zanuttini posits a pre-verbal NEGP and three post-verbal NEGPs, as schematized in (49). (All examples are from Zanuttini 1997.)

(49) [NEGP-1 [TP [NEGP-2 . . . [NEGP-3 . . . [NEGP-4]]]]]

NEGP-1 will host pre-verbal markers like standard Italian non (50).

(50) Gianni non le mangia [Standard Italian]
Gianni not them eats
‘Gianni doesn’t eat them’

NEGP-2 corresponds to the presuppositional negative markers like Piedmontese pa, which precede adverbs like ‘already’ (51).

The members of the fifth dimension (ordinals and numerals) in Laenzlinger’s analysis also arguably find their place in the nominal domain, as instances of NumberP.

Let me close this note with two observations. First note how the mapping in (iii) mirrors the mapping found in the sentential domain, where CP expresses discourse relation and speaker orientation, IP “inflected” elements, similar to degrees, and vP/VP more inherent notions.

Second, it is worth mentioning that just like Laenzlinger reduced Scott’s rich hierarchy of adjectives into five classes, Tenny (2000) also claims that Cinque’s rich hierarchy of adverbs falls into five classes. This may be yet another reflex of X-bar-like organization, with adjoined material (adverbs/adjectives) fitting into the X-bar schema at five loci: below X‘, above X”, and along the three levels of projection (X”, X‘, X”).
NEGPs are projections that host negative particles, many of which are invariant (e.g., Russian ne), though some are sensitive to mood, tense, or aspect (Hungarian ne/nem; Arabic lam/la). There are two main strategies for negating a sentence. Languages may negate the whole clause using a marker that functions like a verb taking a sentential complement, as seen in Tongan.

Other languages use negative markers that have the characteristics of a finite auxiliary. This is the case in Evenki (55).

As Zanuttini (2001: 513) observes, the NEGPs illustrated above are projections hosting negative particles, many of which are invariant (e.g., Russian ne), though some are sensitive to mood, tense, or aspect (Hungarian ne/nem; Arabic lam/la). In addition to negative particles, languages resort to two other strategies to negate a sentence. Polynesian languages, for example, negate the whole clause by means of a marker that has the characteristics of a verb taking a sentential complement. Consider (54), from Tongan.

Other languages use negative markers that have the characteristics of a finite auxiliary. This is the case in Evenki (55).
Putting all this information together, we obtain the following picture. Negative markers of the embedding sort (54/55) can occur either at the CP-juncture (54), or at the vP-juncture (55). In addition, negative particles can adjoin to the finite verb in Infl\(^*\). Alternatively, they can occur below Infl\(^*\). In the latter case, the negative particle can be associated with presupposition (closest to Infl\(^*\)), or be discourse-neutral, or be associated with focus (farthest from Infl\(^*\)).

The picture that emerges could be represented schematically in (56).

\[
\begin{array}{c}
\text{CP} \\
| \leftarrow \text{NEGP} \\
\end{array}
\]

\[
\begin{array}{c}
\text{IP} \\
| \leftarrow \text{NEGP} \\
\end{array}
\]

\[
\begin{array}{c}
\text{\textit{vP}} \\
| \leftarrow \text{NEGP} \\
\end{array}
\]

\[
\begin{array}{c}
\text{\textit{vP}} \\
| \leftarrow \text{NEGP} \\
\end{array}
\]

\[
\begin{array}{c}
\text{\textit{vP}} \\
| \leftarrow \text{NEGP} \\
\text{NEGP \ (presupposition)} \\
\text{NEGP \ (discourse-neutral)} \\
\text{NEGP \ (focus)}
\end{array}
\]

Viewed from this perspective the mobility of NEGP is very reminiscent of what one finds with AGR, which iterates around the core sentential units (V’s extended projection), and can in turn be decomposed into an ordered triplet ([Person [Number [Gender]]]) the same way the lower NEGP appears to be (in Romance).

Perhaps the decompositionality of the lower NEGP can be understood as indicating the fact that NEG\(^*\) is like French \textit{ne} (often phonetically null), which is supported by a particle like \textit{pas}, which in turn could be hierarchically organized along a standard X-bar schema.\(^{14}\)

\(^{13}\) On auxiliaries as \(v\), see den Dikken (1994) and Bošković (2001).

\(^{14}\) Negative markers like \textit{ne} and \textit{pas} are repeatedly used to illustrate “Jespersen’s Cycle.” The latter describes the historical development of negative forms in a variety of languages such as French, where the original negative form (\textit{ne}) underwent two stages of weakening, the first one requiring support of another element (\textit{pas}) to express negation, and the second stage rendering the original negative form entirely optional. It may not be too much of a stretch to see in this the workings of X-bar syntax: From a full-blown element (XP), \textit{ne} became a head (X\(^*\)) requiring a complement, and then became optional material (X\(^*\)).
At any rate, the similarity between NEGP and AGRP is well worth exploring further. Both appear to act like adjoined categories, grafted onto the clausal skeleton, as shown in (57):

\[
\text{(57) } \begin{array}{c}
\text{CP} \\
\text{IP} \\
\text{rP} \\
\end{array} \xrightarrow{\text{NEGP/AGRP}} \end{array}
\]

Interestingly, in the cartographic literature, agreement and negation have been grouped together because of their problematic character. As both Cinque (1999) and Julien (2002) have observed, the beautiful hierarchy that emerges from the extraordinary consistency found across languages for both affix ordering and adverb placement (extended to restructuring predicates in Cinque 2006) breaks down as soon as agreement and negation are factored in. Summarizing results from the cartographic literature, Svenonius (2007a: 265) notes that “[o]rders of T-Asp-V, T-Caus-V, and so on were demonstrated . . . to show great cross-linguistic regularities. Negation [and agreement] provide a startlingly different picture.”

The framework developed here, which is nothing more than exploiting the X-bar schema in the context of groupings of projections, can make sense of the odd behavior of AGRP and NEGP.\footnote{One could take the problematic nature of agreement and negation as an argument in favor of Chomsky’s (1995) proposal that agreement is not a legitimate projection. Chomsky defends his claim by pointing to the fact that AGRP consists of uninterpretable features only. While this may be true of AGRP, it is hard to see how this could extend to sentential negation (which is interpretable/interpreted). Accordingly, Chomsky’s suggestion would not enable us to solve the hierarchical mapping problem at hand in a sufficiently general fashion.}

Both are “outside” the clausal skeleton, on a parallel plane (like adjoined material typically is). Their somewhat erratic linearization pattern comes from the fact that they can be “collapsed” onto the clausal skeleton at various junctures.

In this sense they do not pattern with the adjuncts discussed in Cinque (1999) because these, as we saw above, express notions
directly related to the core clausal skeleton. That is, Cinque’s adjuncts are not on the same plane as NEGP and AGRP. The main reason why I think Cinque fails to find unity among adjuncts, agreement markers, and negative markers is because he tried to map them all onto a single syntactic projection line. The perspective offered in this chapter is that the organization of adverbs, verbal affixes, agreement markers and negative markers do converge, but in parallel dimensions. All of these dimensions ultimately collapse onto a single “line” when syntax gets mapped onto the interfaces, but at that point, we expect variation, since the different dimensions can be integrated into one another at different juncture points.

4.6 Capturing typological restrictions

Having shown how the X-bar schema matches fairly closely the results of cartographic projects, let me now turn to other typological restrictions that may be the result of the X-bar organization imposed by syntax.

Before proceeding, I should note that the pervasiveness of the X-bar schema that I am arguing for here should not be understood as a claim to the effect that this is the sole source of hierarchy. Clearly hierarchies exist outside of syntax, and in some cases, they may be isomorphic to the X-bar schema (as in the case of gradable adjectives), not necessarily the result of “syntacticization.” Only comparative studies across mental modules and across species will reveal which hierarchies are independently generated.

Similarly, I wouldn’t like to give the reader the impression that any triad one finds in language is to be analyzed as being organized along the lines of the X-bar schema. Although I think that triple manifestations of a given dimension are suggestive, there exist a few three-member sets for which I have not been able to find any shred of evidence for an internal hierarchy (nesting). Take, for example, the existence of three major lexical categories (\{A, N, V\}), if Baker
(2003) is correct in viewing adpositions as functional items, of the three possible clause types (\{Interrogative, Declarative, Exclamative\} values for ForceP/$\Sigma P$ (Laka 1990)), or the three values for GenderP (\{Masculine, Feminine, Neuter\}), or the three shifters/discourse variables studied by Schlenker (2005) (\{Mood, Tense, Person\}).

All these may be regarded as flat triplets. Such triplets may be the result of the fact that concepts, once digitized, can take on any of the three possible values: [+,-,Ø], which will form the basis for three parallel projections, not three nested projections. Alternatively, some of them may be the result of the fact, well established in the cognitive science literature, that humans and many other species are able to chunk conceptual spaces into small groups (of three, and in some cases four), a cognitive operation technically known as subitizing (see Dehaene 1997, among others; see also Pinker 1997 on the pervasiveness of three-way divisions).

Accordingly, three-member sets in language may have multiple sources. What is special about the hierarchies examined in this chapter is not so much the fact that they involve three members, but the fact that these members organize themselves in a nested pattern. What is lacking in the context of subitization or lexicalization is the internal hierarchy imposed by, and perhaps unique to, narrow syntax (Hauser, Chomsky, and Fitch 2002).

Let me end this brief note of caution by saying that although imposing an internal hierarchy on some of the flat triplets one finds may not look very promising at first, one must leave open the possibility for there to be an internal structure inside the triads just listed. One of the great lessons that emerged from work in syntax over the past thirty years is that very often closer empirical scrutiny reveals subtle, but robust asymmetries where at first symmetries are expected.

This may be the case for Person. While it is fairly standard to treat 3rd person as the odd one out, it is not clear whether 1st person and 2nd person should be hierarchically ordered, X-bar like. Most studies on phi-feature geometry (see, e.g., Harley and Ritter 2002) assign the same hierarchical status to both 1st and 2nd person ([+author; −participant] for 1st person, [−author; +participant] for 2nd person). However, Quinn (2006) provides evidence for a 1st person >
2nd person hierarchy on the basis of Algonquian data. (Quinn’s evidence is particularly relevant as it comes from the complex Inverse system in Algonquian, a system traditionally thought to require a 2nd person > 1st person hierarchy (see McGinnis 2005).)

Likewise, the apparently symmetric values of Tense (Future, Present, Past) may turn out to be nested, if Baker and Travis (1997) are right in treating the values of Tense as similar to the dimensions found in the nominal domain (definite, indefinite nominals), which show signs of nesting.

Finally, consider Number. It is known that the latter standardly takes on the two values {singular}, and {plural}, but it is also known that number can take on additional values like {dual}, {trial}, and {paucal}. As far as we know, number cannot take on any other value (see Corbett 2000). This could signal a cognitive limit, or it could be the reflex of an X-bar-like schema of the sort given in (58).

If (58) is tenable, the X-bar schema could be used to make sense of typological findings, by imposing limits on variation.

For example, the X-bar schema may allow us to explain why macroparametric effects are concentrated in a way that Baker (1996) captured well in this quote:

It is not clear that there are other natural language representation systems that do not fit somewhere within the triangle defined by head-first/head-last/adjoined, and the related triangle defined by Case-marked/agreement-marked/uninflected. (Baker 1996: 506)

The categories head-first, head-last, and adjoined are precisely the three defining X-bar relations, once “head-first” and “head-last” are

\[ \text{(58)} \]

\[
\begin{array}{ccc}
X'' \text{ (plural)} & \text{X'' (paucal)} \\
\mid & \mid \\
X' & X' \text{ (trial)} \\
\mid & \mid \\
X'' \text{ (singular)} & X'' \text{ (dual)}
\end{array}
\]

\[ I \text{ thank Eytan Zweig (p.c.) for encouraging me to look at Corbett’s survey.} \]
understood along the lines of Kayne’s (1994) Universal Base hypothesis, which rigidly orders heads before complements (head-first), specifiers before heads (head-last).

The X-bar schema may also prove useful in the domain of verb clusters recently studied by Svenonius (2007a). Svenonius examines the known cross-linguistic orderings of sets of morphemes, corresponding to three hierarchically organized elements (symbolized as 1-2-3), for example, C-T-V. Svenonius arrives at the following topological variations. (I reproduce Svenonius’s terms for some of these patterns right next to the sequences the terms refer to.)

(59) a. 1-2-3 “Straight”
   b. 1-3-2 “Curl”
   c. 3-2-1 “Roll-Up”/“Snowball”
   d. 3-1-2 Long-distance extraction
   e. 2-3-1 Constituent Fronting
   f. 2-1-3 Short-extraction

Pattern (59a) corresponds to the situation one finds with English main verbs: an undisrupted C-T-V sequence (60a). (59b) is found in languages that raise V past T, like French, or German (C-V-T) (60b). (59c) corresponds to robustly head-final languages like Japanese (Verb-Tense suffix-C suffix) (60c). (59d) corresponds to long-distance dependencies like Wh-movement (from within VP to SpecCP), or perhaps long-distance participle fronting in Slavic (the long head-movement cases analyzed in Rivero (1991) and Roberts (1993)), or V-fronting (60d). (59e) captures instances of VP-fronting (60e), while (59f) captures instances of movement like T-to-C raising in English (60f).

(60) a. that (it) will rain
   b. dass (es) regnen wird [German]
      ‘that it will rain’
   c. (ame-ga) fu-ru to rain-nom fall-nonpast Comp
      ‘that rain will fall’
   d. Rain, it certainly will
   e. [Kiss [John]], Mary will
   f. Will it rain?
Of particular interest in the present context is Svenonius’s claim that the patterns in (60a–f) exhaust those that are needed to capture the range of word-order permutations attested in natural language. Considering the range of options offered by permutations of two elements only would not be enough to capture the cross-linguistic variation we find (only two possible patterns would be expected). More crucially, taking four elements into account would give rise to too many cross-linguistically unattested patterns (sixteen patterns). Svenonius concludes that the possible reorderings of exactly three elements suffice to capture the space of possibilities. Svenonius’s result follows immediately if the space of possibility is defined by the X-bar schema, which makes room for three, and only three levels of projections, occupied by elements that can be reordered by movement.

4.7 How cartographies emerge, and why

So far the message I have wanted to convey is that cartographies are not a different type of syntactic object, nor are they random collections of X-bar projections. Instead, following Grimshaw’s lead, I have been at pains to show that cartographies are “macro-” projections, with the very same characteristics that define regular projections.

To make my case I have made use of standard X-bar terminology, but I think we are now ready to address the minimalist why-question: Why should cartographies exist? Can we get to the bottom of extended projections, and see these projection(-level)s emerge from the same principles that guided our investigation of basic Merge in Chapter 3?

I believe we can, although I will only be able to demonstrate this for a few cartographies, due to the fact that the syntactic behavior of only a few of the members of the cartographies discussed in this chapter is sufficiently well understood for me to attempt to derive it. My starting point will be the observation I made in Chapter 3 to the effect that the labeling algorithm in terms of Probe-Goal
relations argued for there depicts natural language syntax as a computational engine that ensures that only one relevant thing happens at any given time. If more than one thing happened at once, SEM and PHON would receive ambiguous instructions, which must be prevented at all costs. Accordingly, a major task of the syntactic component will be to space out the various Probe-Goal relations that must be established in such a way that only one Probe-Goal relation takes place per Spell-Out unit.\(^{17}\)

One consequence of this “No-Ambiguity” restriction imposed by the external systems that I discussed at the end of Chapter 3 amounts to a Distinctness Condition imposed on relations taking place within a phrase. In order to avoid sending contradictory instructions at the interfaces, syntax must ensure that the Probe-Goal relation underlying the first instance of Merge must be distinct from the Probe-Goal relation underlying the second instance of Merge.\(^ {18}\) As I pointed out in Chapter 3, this distinctness condition immediately results in a ban on phrase-internal movement (Anti-locality). It also forces the members of a given phrase to be sufficiently different (featurally speaking), giving rise to Identity-Avoidance effects of the type discussed by van Riemsdijk (forthcoming) and Richards (2006). Here I would like to propose that the distinctness condition also has the effect of imposing a ban on multiple arguments per phrase (cf. Larson 1988). Put simply, I would like to suggest that if the same kind of Probe-Goal relation is established between a head and its complement, on the one hand, and between a head and its specifier on the other, the instructions sent to the external system in the form of labeled set will not be unique/unambiguous enough, and will be filtered out.

Accordingly, if multiple theta-roles have to be assigned, multiple VP-like projections will have to be erected if the event structure of a sentence is to be articulated in a way that can be parsed by SEM (and

\(^{17}\) By Spell-Out unit, I have in mind a model like the one proposed in Epstein and Seely (2002, 2006), where each product of Merge is mapped onto SEM and PHON.

\(^{18}\) Let me stress that when I say “syntax must ensure . . .,” I do not mean that syntax can only generate legible outputs. No look-ahead is involved; Merge proceeds the way it does, “unrestricted,” with “wrong” derivational options filtered out at the interfaces.
PHON). Similarly for multiple A-bar relations. This will give rise to an extended VP-domain and an extended left periphery.

But multiple projections alone won’t be enough to achieve the right results. The heads of the projections within a given domain will have to be distinct, in two ways: first, they will have to be distinct in order to embed (establish a Probe-Goal relation), and second, they will have to be distinct even after they have entered into a Probe-Goal relation so as to make sure that the relations they will establish will not give rise to ambiguous instructions. And yet, if the heads are distinct, how can they be part of the same domain? I think that part of the answer to these questions comes from Fortuny’s (2006) discussion of what he calls “discontinuous” categories. As Fortuny notes, once a concept C is lexicalized/digitized, it gives rise to two possible heads: a head specified for +C, another specified for −C (I return below to the third option: a head not specified for C (φC)). I would like to claim that the creation of distinct values during the lexicalization pattern is enough to make heads of the same domain sufficiently similar and at the same time sufficiently distinct to enter into Probe-Goal relations.

Furthermore, once we realize that the highest head within a given domain will have to enter into a Probe-Goal relation with the next higher domain, it becomes possible to keep the heads within a given domain distinct by relying on the presence of an additional feature on the highest head of the domain required to connect with the next domain up the tree.

Let me be more concrete by considering the VP/vP domain. Given that all “thematic” heads express Event structure/Aktionsart (see Ramchand forthcoming), let us suppose all of them are lexically endowed with a lexical feature α. The specific identity of this feature is not very important. I chose α because it evokes A(ktionsart), and because the α-domain thus formed constitutes the first domain of interpretation from a bottom-up perspective.

Following Fortuny, I take it to mean that there can be a lexical item/head specified for +α, and another for −α. We now have two heads that can take part in event articulation. The −α head will probe for a +α head, yielding a pair <−α, +α>. Because the higher head (“−α”) will be selected by T, it will have to be specified for the appropriate
feature (say, +T), which can now combine with T (i.e., −T), forming the pair <−T, +T>.

The same reasoning works for the CP-domain as well. The latter is standardly assumed to express force. Since CP typically marks the end of the derivation, let us suppose that all C-heads are endowed with a lexical feature ω (a term that also evokes O(perator)). Applying Fortuny’s logic, we obtain a [+ω]-head and a [−ω]-head. These will merge, forming the pair <−ω, +ω>. Since the C-domain has to be connected with the lower part of the tree (specifically, T), the lower head of the C-domain (“+ω”) will have to be endowed with a T-feature; specifically, −T. The latter will be able to probe the T-head, which is now +T, having received its value in virtue of merging with a +T-head (the highest head in the thematic domain).

Notice that because the heads of a given domain are distinct by virtue of their feature value (+/−), or featural content, nothing prevents them from establishing identical Probe-Goal relations with the elements they independently combine with. For instance, both α-heads can probe for, say, φ-features when they combine with the elements that will function as their arguments.

A typical derivation therefore works on distinct featural planes, as represented in (61). (The ± on the linking T-head is meant to indicate that it turned from a [−T]-head into a [+T]-head after combining with the higher α-head.)

\[(61)\]

\[
\begin{array}{c}
\text{argument structure} \\
+\phi H^o & +\phi H^o \\
\hline
-\phi & -\phi
\end{array}
\]

\[
\begin{array}{c}
\text{clausal skeleton} \\
H^o & H^o & H^o & H^o & H^o \\
-\omega & +\omega & -\alpha & +\alpha & \\
-\phi & -\phi & \\
\end{array}
\]

\[
\omega\text{-domain} \quad \alpha\text{-domain}
\]

In more traditional tree format:

\[^{19}\text{Grohmann (2003) uses the term \(\omega\)-domain in much the same sense used here.}\]
Using more traditional labels, we obtain the following structure:

(63) Force P
    /     /
   Forceº FinP
    /     /
   Finº TP
    /     /
   Tº vP
    /     /
   DP vº VP
    /     /
   Dº Vº DP
    /     /
   Vº Dº
(64) provides a stage-by-stage view on the derivation. The reader can verify that only one Probe-Goal relation is established at each derivational stage, and no two stages relate items with exactly the same featural composition. In this way, the entire derivation is QED-compliant (in the sense discussed in Chapter 3).

(64)  
   a. V-DP: φ
   b. V-v: α
   c. v-DP: φ (but V ≠ v; ⊤, stage (c) ≠ stage (a))
   d. v-T: T
   e. T-Fin: T (but T is the Goal in (e), and the Probe in (d))
   f. Fin-Force: ω

The derivation just discussed has a few interesting aspects that are worth highlighting. Perhaps the most striking aspect of (64) is that among the members of the core clausal skeleton, T stands out in various ways. First, it relates two domains (α and ω). Second, it does not contain any feature defining either of the two domains (α or ω) (one could say that it is [øα, øω]). Third, by the end of the derivation, it does not contain the highest occurrence of any feature: +T is on Fin°. In this way, T is very much like the intermediate level of projection in the standard X-bar schema, or in its bare version: it is neither maximal nor minimal, but it connects the two domains (specifier domain and complement domain). If we assume, as is standard, that only the highest occurrence of a feature is interpreted at SEM, T receives no interpretation, as does X’, which is standardly claimed to be “invisible.”

The similarity between T and X’ can be made more conspicuous by rearranging the symbols in (61) as in (65b) (ignoring the argument DPs).

(65)  
   a. X”
   b. +T
      WP X’
      X° AP
      ω +T
      α +T
Just like X’ emerges from the binary branching requirement needed to define unambiguous paths in Kayne’s (1984) system, so does T emerge here to clearly demarcate the beginning and end point of both the α- and ω-domains. In other words, the existence of T may be understood as the solution to keep the α-domain and the ω-domain maximally distinct, while allowing them to relate (indirectly) to one another and form a (macro) phrase (extended projection). In other words, the presence of T allows for the formation (and connection) of four distinct cartographies: a thematic or VP domain (α-domain), a left periphery (ω-domain), an extended T-projection (cf. (32) above) (T on an α-element, aka Aspect; T on T, aka Tense; and T on an ω-element, aka Mood),20 and an entire clause (V-T-C). All of these form domains with clear beginnings and clear ends. They are all QED-compliant.

Thinking of Chomsky (2004: 110), one could say that the way the derivation proceeds in (64) guarantees the creation of a set of instructions with unambiguous “dual” semantics information: a predicative domain (α-domain) and a propositional domain (ω-domain); alternatively, an A- and an A-bar domain.21

It should be obvious that the derivation in (64) could be expanded by adding φ-features to T or α-projections, creating the possibilities of Internal Merge. I will not dwell on this possibility here, as Internal Merge (chain formation) is the topic of the next chapter, but let me point out a few things related to this possibility here.

First, the presence of φ-features and T-features allow an α-element to expand in two directions (/dimensions) (in addition to being able to expand “on its own,” via +/−α): φ-features allow α-elements to connect to DPs (arguments), and T-features ultimately allow the α-domain to be connected to the ω-domain. We could represent this derivational horizon as in (65).

---

20 Alternatively, Event time-reference/assertion time, and speech time (cf. (41) above).
21 Like Chomsky, I posit the existence of a duality of semantics as an axiom. But it would not surprise me if the latter found its origin in non-linguistic domains. Perhaps the fact that our primate brain appears to split the flow of visual information into two streams (a “what” stream and a “where” stream) may be relevant here.
The existence of three derivational dimensions represented in (65) make clear that attempts like Cinque (1999) or Julien (2002) to find a common projection line on which to map all morphemes cannot succeed without leaving out one dimension (either the adjunct dimension, or the agreement (/negation-focus) dimension. It is a well-established fact in the mathematical literature that 3D-objects cannot be mapped directly onto 1D-objects. 1D-representations can only be directly projected from 2D-representations.

If we follow Watanabe 2004 in taking negation to partake in feature-checking via a feature focus, and if in turn we treat focus as another \( \phi \)-feature (see Miyagawa 2007), the Topic-focus articulation that “floats” around core projections may simply be a reflex of the addition of \( \phi \)-features on core functional projections. We thereby readily capture the iterative aspects of syntax discussed in Section 4.5.5 above. We can also explain why topics fail to intervene in matters of selection. Recall from Chapter 1 that Shlonsky (2006) notes that in some languages like Hebrew topics can intervene linearly in between the force indicator (a wh-phrase) and a selecting verb. The relevant example is repeated in (66).

\[
\text{(66) Sa’alta oti et ha sefer le mi le haxzir [Hebrew]}
\]
\[
\text{asked.2sg me ACC the book to whom to return}
\]
\[
\text{‘You asked me to whom to return the book’}
\]

The presence of a topic in (66) is no longer problematic under the present approach as the \( \phi \)-relation involved can be said to take place on a distinct plane from the \( \omega \)-dimension that is selected for. This is very clear if we represent topics like we represent arguments in (61). This is shown in (67).
In other words, the selection problem is only apparent. It arises only if we view elements organized in linear terms.

One other interesting aspect of the derivation in (61)–(64) is the necessary presence of a T-feature on the higher α-member. If we interpret T-features as “providing” Case, as in Pesetsky and Torrego (2001), we can view the T-feature on the higher α-member as a way of deriving Burzio’s generalization, which expresses the idea that the head introducing the external argument is responsible for assigning Case to the object. Note that because the mapping to SEM and PHON demands that there only be one Probe-Goal relation per Merge, it follows that the Case of the argument introduced by the highest α-head (ν for the external argument in transitives; V for unaccusatives) will have to be matched/checked by the T-feature on T.

Finally, let me note that the derivation in (61)–(64) shares obvious features with Chomsky’s phase-system (Chomsky 2000 and subsequent works). As in the present approach, Chomsky’s system recognizes the existence of two core domains: the ν-phase and the C-phase, and assigns a special, non-phase status to T. Like the present system, where the highest member of the α-domain shares a feature with the higher domain, the edge of a phase in Chomsky’s system is said to belong to the next higher phase. But the parallelism between the present system and Chomsky’s phase-system should not obscure the differences. For Chomsky, V is never a phase, but in the present system, V can be the maximal +α-element if only one argument is introduced into the derivation. Also, unlike Chomsky’s system, the

---

22 For an accessible introduction to phase-based derivations, see Hornstein, Nunes, and Grohmann (2006: ch. 10).
The present approach sees no need to introduce a Phase-Impenetrability Condition in syntax, which renders some portions of the clause inaccessible as the derivation proceeds. Finally, the present system does not claim that T inherits features from C (compare Chomsky 2007, forthcoming; Richards 2007). In the present system, T shares features with both the “C”-domain and the v/V-domain. Chomsky’s (2007, to appear) claim that Tense is interpreted on C follows from the fact that the highest T-feature resides in the ω-domain.

Let me close this very brief discussion of Chomsky’s phase-system by pointing out that the parallelism between phases and the derivational domains in (61)–(64) is not accidental. Just like I have argued that the domains in (61) are “macro projections,” so are phases.

For Chomsky, the clause consists of a lower phase-head (v for Chomsky), and a higher phase-head (C), both of which are surrounded by non-phases (V, I, matrix V). This could be represented as in (68), which highlights the projection nature of phases.

(68) ( ) (non-phase; matrix V)
    | XP higher phase (C)/“max” phase
    | (X') (non-phase; I)
    | X lower phase (v)/“min” phase
    | ( ) non-phase (V)

Considering the representation in (68), one can say that perhaps the main difference between phase-based theories and the present approach is that the α-domain and the ω-domain are kept separate and distinct in

---

23 See Boeckx (2007a: ch. 3) and Boeckx and Grohmann (2007) for empirical and conceptual arguments that divorcing syntactic locality from phasehood is desirable. See also Jeong (2007), Bošković (2007), and Fox and Pesetsky (2005). Divorcing Spell-Out units (“phases”) from locality allows me to adopt Epstein and Seely’s (2002, 2006) proposal that each product of Merge gets mapped onto PHON and SEM, without any unwanted syntactic consequences.
the present approach, whereas for Chomsky the edge of the lower phase (i.e., the lower-phase head and its specifier) are part of the next higher phase). In the present framework, T provides the point of transition, “phases” (/domains) do not intersect directly.

4.8 Conclusion: The fractal nature of syntax

The present chapter has made one simple claim: the traditional X-bar phrase can be regarded as the common denominator across the various hierarchies in natural language syntax that cartographic projects have helped reveal.

It could in fact be said that the present chapter has revealed the fractal nature of narrow syntax. For every X one examines, X turns out to project a path (an X-bar-like projection), and each path-member can in turn be expanded in an X-bar fashion. Syntactic patterns boil down to something like (69).

\[ \text{(69)} \]
Fractals are characterized by properties like dimensionality, infinity, recursivity, phase transition, and self-similarity (Mandelbrodt 1982)—key properties of natural languages, independently identified by various syntacticians, whose arguments are thus strengthened by the present approach.

I hope to have shown that there is much systematicity to what at first sight looks like a disparate group of findings (extended projection, Cinque-hierarchy, shell-structures, CP-peripheries, functional zones, phases, etc.). It seems quite clear to me that the vast majority of the hierarchies independently argued for in syntax organize themselves along the traditional X-bar schema, to such a degree that it is hard for me to think of this as an accident.

Natural language syntax appears to be a realm of which one can say that it’s all phrase, phrase, phrase. Taking my lead from Chapter 3, I have argued that the reason for this pervasive X-bar organization is that it provides a way of sending maximally unambiguous instructions to the interfaces. The latter imposes a certain computational limit, a window of opportunity for natural language syntax, but within that window, infinite self-similar structures can form. So syntax loses nothing of its expressive power.

I should point out that although I have often talked about hierarchies in X-bar terms, there is a clear sense in which the fractal nature of syntax argued for here could not be formulated if we...

---

24 Phase transition refers to situations where an element is halfway between two dimensions, as it were. In syntax I’d argue that this captures situations where the maximal point of one X-bar schema corresponds fairly closely to the minimal point of the next higher domain/dimension, due to the feature-sharing relation that must be established. We saw this in Section 4.7 when we discussed the role of T as a point of transition. The T feature on the higher α-member makes the latter the highest member of the α-projection and the lowest member of the T-projection. The same situation obtains in other syntactic domains. For instance, in the domain of adpositions (“PPs”), members of the lowest dimension, referred to as pP in (9) above, are very close in nature to the light prepositional relator of that has often been characterized as the highest, complementizer-like domain of noun phrases (think also of the ambiguous status of Case-markers in languages like Japanese). Similarly for Mood-Tense (irrealis mood as future tense) and Tense-Aspect (Past tense as terminated aspect), with several authors going back and forth between two adjacent domains in their own writings (see, e.g., Belletti (2004a) on the “low IP area” and Belletti (2005) on the “left edge of VP”).
adhered to the traditional understanding of projection in X-bar theory. In the original X-bar schema, it makes little sense to call a head a maximal unit. But as soon as projection levels are understood in more relative terms, as in Chomsky’s bare phrase structure, terms like minimal, intermediate, and maximal can be used at different levels of representation/abstraction. In this sense, I think that it is fair to say that a more minimalist way of looking at things may help reveal an underlying unity that would be hard to express in more traditional terms.

As I pointed out in various sections, the main reason behind my pushing hard to reveal the X-bar nature of cartographies was to maintain the simplest notion of locality in the domain of selection. If cartographies are phrases, then selection can be said to always take place between adjacent phrases.

In addition to this explanatory advantage, the present approach provides a theoretical frame on which to hang facts and guide future cartographic studies. Currently, cartographic approaches have few principles to guide them, a fact that has led to the critique that such projects don’t go much beyond stamp-collecting (see Cinque 1999: 224 n. 11). I hope to have convinced the reader with this chapter that this is unfair criticism, and I hope that the search for missing elements that would reveal X-bar-like structures will be on the agenda of cartographic projects in the future.

I began this chapter with a few basic questions (cf. Section 4.1). Let me conclude by summarizing the answers that previous sections have provided. The first question pertained to the status of projections: Are they all created equal? The short answer is no, some are minimal, others intermediate, and yet other maximal elements of X-bar schemas. In part following den Dikken (2006), one could say that some elements like T are linkers (intermediate projections), others are relators (minimal projections), and yet others could be called culminators (maximal projections). Baker (2003) was right in recognizing three basic projection types. He called them lexical, transparent (/extending), and opaque (category-changing). I have called them minimal, intermediate, and maximal. Although this degree of convergence between the present study and previous
investigations is interesting to note, I think the present approach provides clearer answers to questions like “How does one go about characterizing a core projection?” (For me, this could be a cover term for domains like the $\omega$-domain), or “What is the difference between TenseP and FinitenessP?” (for me, both projections share a T-feature, and are needed to keep the $\omega$-domain and the $\alpha$-domain separate).

Finally, and perhaps most importantly, the present chapter makes clear that cartographies do not constitute a distinct module of the grammar, and do not require independent principles of organization. They are simply reflexes of Merge and the interface conditions imposed on it.
Islands and the locality of chains

The goal of this chapter is to examine the structure and internal composition of chains through the same lenses we have used to study the most basic instances of Merge (Chapter 3) and the formation of cartographies (Chapter 4). As stated in Chapter 2, I intend to show that the very specific locality effects we find in the context of long-distance dependencies (chains) are there for the same reason we find endocentric projections (unique labels) and extended projections (cartographies): all three “facts” are imposed on Merge to render the products of Merge (“syntactic objects”) maximally compliant (unambiguous; i.e., easily searchable) at the interfaces.

5.1 How to approach the issue

To ensure efficient mapping, syntax must provide unambiguous instructions. The specific way in which I have proposed to understand “unambiguous instructions” is in terms of “edges”: the extremities of syntactic objects must be clearly demarcated, and must be quickly reachable, by which I mean that once the beginning of a syntactic object is identified, its end should be nearby (i.e., the search for the end point must be kept to a minimum). This is what I have called the “Quick Edge Detection” requirement (QED). Concretely, QED-compliance means that the products of Merge must have one easily identifiable head and a non-head close by. I argued in Chapter 3 that the optimal way to achieve this is to impose a binary branching requirement on Merge, and an algorithm that takes the head/label of Merge to be the unique Probe relating the two Merge partners. That
way, the product of Merge consists of an unambiguous head and an adjacent (unambiguous) non-head, and nothing else. I argued in Chapter 4 that cartographies, just like X-bar projections, are nothing but side-effects of binary branching and unique probing. If multiple relations must be established among various elements, they cannot be established all at once, they must be spaced out to keep each relation maximally distinct. I should maybe stress again that such conditions as binary branching or unique probing need not be seen as “part of” Merge. Merge may be said to apply freely, its output being subject to QED-compliance. That is, binary branching and unique probing are added onto Merge to ensure interface legibility. In McCawley’s (1968) terms, binary branching and unique probing are node-admissibility conditions.

The net result of these conditions is the formation of highly asymmetric, maximally unambiguous syntactic objects. But this is not to say that asymmetry rules in syntax. In this chapter we will see that the fundamentally symmetric character of Merge is restored in the form of chains.

Recall that even if the product of merging $\alpha$ and $\beta$ is asymmetric ($\langle \alpha, \beta \rangle$), Merge itself applies symmetrically. When we say that $\alpha$ and $\beta$ merge, we don’t mean that one is merged to the other irreversibly. Merge is order-independent: $\text{Merge}(\alpha, \beta) = \text{Merge}(\beta, \alpha)$. Although this symmetry is immediately broken in the guise of labeling (either $\alpha$ or $\beta$, whichever one probes the other, projects), I would like to explore the possibility that the symmetry of Merge continues to exist, in a subtle form, in the guise of chain formation: although it is (say) $\alpha$ that projects, $\beta$ projects too, in parallel as it were. If tenable, this claim has important consequences for the form of chains (and for the locality of long-distance dependencies): chains should be the same as projections. As stated in Chapter 2, products of Merge should be equivalent to products of Move. Thus, as in the case of Merge, we expect to find reflexes of binary branching and unique probing in the composition of chains.

I suggested in Chapter 2 that we do find such reflexes in chains, although I used traditional X-bar terminology to state my case there. Here I will focus on how X-bar junctures emerge in chains. As was the case in Chapter 4 for cartographies, the argument will necessitate a
detailed examination of the types of features involved in chain formation. Because I realize that the perspective advocated here may be unusual, let me point out that the argument presented in this chapter will end up agreeing with two independently established characterizations of chains and their locality. The first characterization is Rizzi’s (2006) claim that chains start at an s-selectional position and end at a “criterial” position. According to Rizzi, there is no chain link lower than the s-selection position, nor is there any position higher than the criterial position. Any intermediate position is neither s-selectional nor criterial. The second characterization is Richards’s (2001) and Boeckx’s (2003a) condition on chains that are “too strong”—what Boeckx (2003a) called the Principle of Unambiguous Chain (PUC). The PUC says that a chain can contain at most one strong position (/occurrence), where a strong position (/occurrence) is defined in terms of checking of a feature associated with an EPP-property.

There is a fair amount of overlap between Rizzi’s and Boeckx’s/Richards’s characterizations. Although they differ in detail, I suspect that both would end up ruling out roughly the same set of bad sentences. Rizzi’s study focuses more on the effects of his characterization for LF/SEM, while Boeckx and Richards concentrate more on the effects of their characterization for PF/PHON. I will argue that both perspectives are correct, and necessary if one is to derive the characterizations just discussed from more basic principles (something that neither Rizzi nor Boeckx or Richards try to do).

5.2 From Last Resort to Bounding

The proposals made by Rizzi, Boeckx, and Richards code the fact that movement is (upward) bounded. Of course, we have known

---

1 The results of Mueller and Sternefeld’s (1993) Principle of Unambiguous Binding, as well as Bošković’s (2005b, forthcoming) treatment of freezing effect will also be encompassed by the analysis developed in this chapter. Thanks to Sam Epstein and Terje Lohndal for pointing this out to me.

2 Richards (2001), unlike Boeckx (2003a), allows for chains that are too strong to be formed in the syntax, provided they are repaired at PHON. Though not insignificant, I ignore this difference in what follows.
This since Ross unambiguously established the fact that there are domains out of which movement cannot take place.\(^3\) This is the very idea behind islands. But to this Rizzi, Boeckx, and Richards added the fact that movement is subject to Last Resort. It is not just that an element may end up being “stuck” in virtue of being in a certain domain (say, an adjunct clause), an element may also end up being “stuck” once it has reached a certain checking site (a strong position/occurrence for Boeckx/Richards; a criterial site for Rizzi).

When Chomsky introduced the notion of Last Resort in Chomsky (1986a), he did so to rule out instances of movement from a Case-checking position to another Case-checking position, as in (1).

(1) \[ {^{*} \text{John seems} \ [t' \text{is} \ [t \text{happy}]]} \]
\[
\begin{array}{c}
\text{Case} \\
\uparrow \\
\text{Case}
\end{array}
\]

But the idea of Last Resort generalized to virtually all features, to rule out instances of multiple checking of the same feature by the same element (ban on iterated wh-checking, ban on iterated topicalization, ban on multiple theta-checking, etc.) (I return to a few controversial cases below). Thus we can state (2).

(2) Last Resort condition on movement
An element E can only check a feature F of type T once

The literature on Relativized Minimality has independently examined the issue of feature type. At present, it seems relatively uncontroversial to say that there are three types of features: “Theta,” “Case,” and “A-bar”—corresponding roughly to the three layers of the clause: V, T, and C. But, as we will see in a moment, the distribution of these features is quite complex, and will require close scrutiny.

For now let me reiterate an observation I made in Boeckx (2003a) (see also Gallego 2007). As soon as we combine the notion of upper bound of movement (island) with Last Resort (feature checking), the possibility arises to characterize islands in featural terms, as

\[^3\text{For now, I ignore the fact that for Ross only some instances of (movement) rules are subject to islands. I’ll return to this important caveat below.}\]
opposed to the more traditional geometrical/configurational terms. If we take seriously the idea that (some) checking sites constitute an upper bound for movement, characterizations of island effects like “movement out of left-branches is banned” won’t suffice. They will have to be replaced with statements that focus more on relations than on configurations, like “movement out of Case-checking sites is banned.” The fact that all Case-checking sites are left-branches may thus be a mere side-effect. The geometry of the tree alone may not constrain movement. This in turn gives rise to the possibility (exploited in Boeckx 2003a) that no domain is an absolute island, since all islands will have to be relativized to the type of features in chain formation. Ross’s idea that islands only constrain some movement rules can thus be entertained again, in a fresh, minimalist light.

The possibility of revival of the featural view on islands may be good news, given what we have learned about islands since Ross’s original study. Although island effects are found in all languages, there is some variation in the patterns of extraction that may be hard to capture on a purely configurational view of locality. The latter appears too rigid and universal to allow for variation. By contrast, given that features are standardly considered to be the locus of variation, we may expect that different feature combinations will result in different extraction patterns.

Needless to say, these are just a priori considerations in favor of a featural view of islands. As always, the proof of the pudding remains in the eating, but a priori considerations may help us choose which pudding to start eating first.

In this chapter, I would like to establish (in part building on Boeckx 2003a) the correctness of the statement in (3).4

(3) An element can only move to a single feature-checking site

Like (2), I intend the condition in (3) to act as a boundary condition on movement. The feature-checking site to which the element

---

4 Bošković (forthcoming) makes a similar claim, although his interpretation of the ban on multiple feature-checking under movement and his implementation of it differs from mine in many respects.
moves is intended to mark the upper bound of that element’s chain, the edge defining the chain’s maximal extension. Since the central contention of this work is that syntactic objects can only have two edges, (3) would follow from QED. Taking the site of External Merge to mark one edge of the chain, the feature-checking site occupied by Internal Merge would count as the other edge. Perhaps the most challenging situation for (3) is the alleged existence of mixed A+A’-chains (movement for Case checking followed by A-bar movement, as in subject questions like *who does John think t’ may have been arrested t?). Mixed chains appear to contain two checking sites. I will thus devote special attention to situations where they may arise below.

To even begin to explore the path defined by the considerations above, it is important to stress what the hypothesis in (3) does not say. Here two remarks are in order. First, the hypothesis in (3) does not say that an element can only move once. So long as no feature checking is involved, the number of movement steps is potentially infinite. This is the same point I made in previous chapters concerning Merge and adjunction. QED-considerations impose a single complement and single specifier requirement, but they do not impose any upper bound on adjunction (Merge situations not involving featural transactions).

Building on my previous work (Boeckx 2003a) as well as on Takahashi (1994) and Bošković (2002a, 2007), I have argued at length in Boeckx (2007a) that intermediate landing sites of movement do not involve feature checking. Repeating the evidence here would take me too far afield. I will simply state that the correctness of the present analysis crucially depends on the claim that the formation of intermediate chain links is not motivated by immediate feature-checking considerations. Specifically, I will follow Boeckx (2007a) and assume that an element is free to move/adjoin so long as it contains an unvalued feature.

The claim that movement can proceed solely in virtue of the fact that some feature remains unchecked appears in a variety of recent minimalist frameworks; see López (2002, 2007); Preminger (2007); Putnam (2007); van Craenenbroeck (2006), Stroik (1999, forthcoming); and Hornstein (forthcoming). Thanks to Angel Gallego for bringing some of these works to my attention.
Second, the hypothesis in (3) does not say that an element can only participate in one feature-checking relation. It only says that an element cannot be internally merged in more than one checking site. If all instances of feature checking required movement, as was assumed in early minimalism (cf. Chomsky’s (1993) generalized spec-head relation), (3) would not be empirically viable. But as soon as the existence of Agree is recognized, it can be used in service of (3) to handle potentially problematic instances of multiple feature checking. As a matter of fact, once Agree is assumed, we face a different problem: if Agree exists as a mode of feature checking, doesn’t it render (3) vacuous? If every time an extra checking relation is established Agree can be appealed to, (3) will always be satisfied. We could even say that no checking relation is established under movement. The issue is this: Since the existence of Agree essentially dissociates movement from feature checking, what counts as a checking site for Internal Merge? The key then is to define the connection between feature checking and movement, to which I now turn my attention.

5.3 Checking and Movement

The connection between checking and movement has been called various names over the years. It is called strength in Chomsky (1993). It is called the EPP-property in Chomsky (2001). It is called a criterion in Rizzi (1996, 2006). The precise nature of the connection is still a matter of controversy. But if I want (3) to be a genuine constraint on chain formation, I will have to take a stance on this issue.

At this point it is useful to recall what (3) is in aid of. (3) is put forth to ensure that a chain be QED-compliant, which means that it

---

Over the years, Chomsky has used the EPP to refer to the subject requirement in Chomsky (1981) (“finite clauses must have subjects”) (call this the narrow EPP), or to the driving force behind any movement, including intermediate movements in a chain (call this the broadest EPP) (Chomsky 2000). By EPP in the text I intend something broader than the narrow EPP, but not as broad as the broadest EPP. For me, EPP means movement driven by any feature (A- or A-bar).
should have at most two distinct edges. For projections, edges were defined in terms of Probe-Goal relations under Merge. For larger units like cartographies, edges were defined in terms of feature domains.

Recall the clausal skeleton argued for at the end of Chapter 4. I argued there that a clause consists of two domains (a thematic/predicative domain and a discourse-oriented/propositional domain), which are linked by a series of T-features. ((4) reproduces this clausal skeleton, along with the features involved.)

\[
\begin{array}{c}
\text{Force}^\circ \\
\text{Fin}^\circ \\
\text{T}^\circ \\
\end{array}
\]

\[
\begin{array}{cccc}
\omega & +\omega & -\alpha & +\alpha \\
-T & \pm T & +T \\
\end{array}
\]

On the basis of (4) we could define a domain as all the heads containing an occurrence of a given feature (see Fortuny 2006). Strictly speaking, there are thus three domains in (4), an \(\alpha\)-domain (\(v\)-\(v\)), a T-domain (\(\text{Fin}\)-\(T\)-\(v\)), and an \(\omega\)-domain (\(\text{Force}\)-\(\text{Fin}\)). But the object in (4) is QED-compliant, for there are only two domains whose intersection is empty: the \(\alpha\)- and the \(\omega\)-domain. (4) is really like a traditional X-bar object, of the type in (5). (I use triangles to represent domains.)

\[
\begin{array}{c}
X'' \\
W \\
X' \\
X'' \\
Z
\end{array}
\]

I would like to understand (3) in the same manner. More precisely, I would like to understand the term checking site in (3) in terms of checking domain, and claim that a chain can only encompass, or span, at most one checking domain other than the checking domain it is merged in. In other words, a moving element can only land in one checking domain.
By checking domain I mean a set containing two heads containing F, \{-F\} and \{+F\}. The idea pursued here can be expressed in (6).

(6) A chain can be defined by two domains, one associated with the External Merge position, the other with the Internal Merge position.

In the case of argument-chains, which I am focusing on here (see below for discussion of adjuncts), this means that a chain extending beyond the α-domain can either span a T-domain \{-T, +T\}, or the ω-domain \{-ω, +ω\}. More concretely still, understanding (3) in terms of checking domain means that an element can only be remerged once in a domain that looks like (7).

\[
\text{checking domain}
\]

(7) \[ [H_1^\circ [-F] \ [\_ \ H_2^\circ [+F]]] \]

Checking site

I submit that the configuration in (7) is a very important one for the interfaces, because it counts as an unambiguous checking site, a checking domain. Although the existence of Agree (feature checking at a distance), and the possibility of non-feature-driven movement (adjunction, as in successive cyclic movement), make it hard to identify checking relations that are tied to movement, (7) provides exactly that. Unless an element E is externally merged in the position marked by “\_” in (7), movement to “\_” is required for E to establish a Probe-Goal relation with the higher occurrence of F, H_\circ/C1. As (8) makes clear, if E does not move, H_\circ/C1 will block the relation (relativized minimalism), because H_\circ/C1, H_\circ, and E stand in an asymmetric c-command relation, and share a feature.

(8) \[ [H_1^\circ [-F] \ [H_2^\circ [+F]] \ldots [E]] \]

\[
\text{checking site}
\]

\_

\[\text{DIAGRAM}\]

The term “non-feature-driven movement” may be misleading. Strictly speaking, all instances of movement are feature-driven, but instances of intermediate movements are only indirectly feature-driven. They are made possible by the presence of an unchecked feature on the moving element, but intermediate landing sites are not created for immediate checking purposes.
The only way to void the minimality effect is for E to move above $H_2^\circ$, that is, fill the “__” slot in (7). This, I claim, counts as the establishment as an unambiguous checking site—only one of which can be established per chain, lest the chain run afoul of QED. Any other movement relation could be non-feature-related, hence ignored by the interfaces for purposes of edge identification.

Let me now focus on concrete derivations involving A- and A-bar chain formations. Of all the features involved in chain formation, Case is the one that has been subjected to the most intense scrutiny, ever since Vergnaud’s (1977) letter. By Case, generative linguists typically mean structural Case (nominative or accusative), not tied to any specific interpretive effect (unlike inherent Case, closely associated with specific thematic values). To avoid postulating a feature that is never interpreted on either the Probe or the Goal, I will adopt Pesetsky and Torrego’s (2001) proposal that Case-features on nominals match T-features on the Probes.

Standard introductions to Case have it that accusative Case is closely associated with V, and nominative with T. But the picture is more complex.

Since Burzio (1986) it is standard to regard assignment of accusative as dependent on the presence of an external argument, i.e., v, not just V, as schematized in (9).

\begin{equation}
(v^\prime_{[+ext, \theta]} [__ V^\circ_{[-\phi]}]): \text{accusative Case}
\end{equation}

Likewise, it has been argued recently that assignment of nominative Case is dependent on the presence of $[+\text{Finite}] C^\circ$ (see Chomsky 2007, forthcoming, and references therein), as schematized in (10).

\begin{equation}
[C^\circ_{[+\text{fin}]} [__ T^\circ_{[-\phi]}]): \text{nominative Case}
\end{equation}

Using features we introduced in Chapter 4, we could reformulate (9) and (10) as in (11). (The $\phi$-features on functional heads are there to indicate the presence of a Probe-Goal relation with a DP.)

\begin{align}
(11) & \quad a. \ [\text{Fin}^\prime_{[+T]} [__ T^\circ_{[-\phi]}]): \text{nominative Case} \\
& \quad b. \ [v^\prime_{[+T]} [__ V^\circ_{[-\phi]}]): \text{accusative Case}
\end{align}
Although perfectly symmetrical, the two schemas in (11) conceal an important asymmetry. The element bearing accusative Case comes to occupy the “__” slot in (11b) by External Merge, whereas the element bearing nominative Case comes to occupy the “__” slot in (11a) by Internal Merge (necessarily, as we discussed in the context of (7)). Accordingly, (11a) counts as an unambiguous checking site (tied to movement), but (11b) doesn’t. This, in turn, means that elements fitting in the schema in (11a) will not be able to extend their chains past that point, whereas elements fitting in the schema in (11b) will be able to do so.

This correctly captures the well-known generalization that elements in SpecTP are subject to more severe locality conditions than elements within VP (witness the lack of that-trace effects with objects). But let me stress right away that the special status of SpecTP is not dependent on T alone, it crucially depends on T(P) being dominated by Fin\[^{o}_{\{+T\}}\]. It is the conjunction of Fin\[^{o}\] and T\[^{o}\] that “freezes” the element in SpecTP. I would like to claim that this difference between SpecTP alone and SpecTP dominated by Fin\[^{o}_{\{+T\}}\] provides a way to understand why the presence of a complementizer prevents the element in SpecTP from undergoing A-bar movement (that-trace effect), while the absence of a complementizer readily allows the subject chain to be extended. (Note that the complementizer need not be literally absent in (12a); lack of relevant featural specification would suffice for present purposes.)

(12) a. *Who do you think that t’ is t happy  
b. Who do you think t’ is t happy

I am aware of the fact that it may be risky to rely on the that-trace effect to make a very general theoretical point, given the notoriously variable character of the effect. I will examine this variation below. For now, I intend (12) to serve as an illustration of the claim I am making, not as the motivation for it.

Let me now turn to A-bar chain formation.

By A-bar chains I mean chains used to establish relations like Topicalization, Focalization, and Wh-movement. I take it that \{Focus\}, \{Topic\}, and \{Wh\} are the values that the force feature on
nominals can have. (Consider the close correspondence between Focus, Topic, and Wh on the one hand and the values of ForceP/\Sigma P (Emphatic, Declarative, Interrogative) on the other.)

I would like to use the same reasoning I did in the context of Case chains and say that A-bar chains necessarily terminate as soon as they reach a checking site that is defined by a pair of \( \omega \)-occurrences, Force\(^{\circ} \) and Fin\(^{\circ} \) (corresponding to Fin\(^{\circ} \) and T\(^{\circ} \) for Case chains), as represented in (13).\(^8 \) (Again, the \( \phi \)-features on the host indicate the establishment of a Probe-Goal relation with a DP.)

\[
(13) \quad [\text{Force}^{\circ}_{[-\omega]} \ [\_ \text{Fin}^{\circ}_{[-\phi]}]]
\]
\[\uparrow\]
A-bar checking site

With the understanding of checking site we have reached, we can now reformulate (3) as in (14).

\[
(14) \quad \text{An element can only move to a single Tense checking site or Force checking site}
\]

Given that both the Tense checking site and the Force checking site involve a projection in the C-domain, we could regard (14) as a generalized C-trace effect.

We can think of (14) as the way the external systems scan chains. They can easily identify where chains begin (the point at which the element was introduced into the derivation), and they stop scanning a chain as soon as they hit upon a Force-Fin combination (\( \omega \)-chains), or upon a Fin-T combination (T-chains). In other words, movement itself is blind to the accumulation of featural relations. It is only the interfaces that stop scanning chains once the first unambiguous checking site is reached.

I should stress that (14) allows for the formation of featurally mixed chains, or more accurately for mixed checking (Tense + Force). The only thing it disallows is the presence of two checking sites as part of a single chain.

---

\(^8\) Rizzi (1997, 2004a) would characterize my generic FinP\(_{[-\omega]}\) as TopicP, FocusP, or IntP, but the logic of my argument would be unaffected by this.
Graphically speaking, we can think of trivial chains as in (15a), maximal A-chains as in (15b), maximal A-bar chains as (15c), and illicit chains as in (15d). (Solid lines represent chain links/Merge sites; dotted lines represent (possible) Agree relations, not factored into chain composition.)

The graphs in (15) make clear how (15d) violates the ban on binary branching. Rotating the graphs in (15) makes it even clearer. (Only solid lines/movement relations are represented.)

The representations in (16) are Brody-style mirror-theoretic projections (see Brody 2000, 2003). They easily translate into the standard projections in (17).

The graphs in (17) show first how chains can be understood as projections, and how too much checking leads to a departure of the X-bar schema (17d), hence to a violation of QED.
Returning to (14), and comparing it to our starting point (the Last Resort condition in (2)), we can say that although both (2) and (14) tell us when chains of a certain type have to stop, (14) is less of a Last Resort condition than (2). (14) is geared towards the satisfaction of QED. Last Resort is to be understood as a condition on Agree: once valued, a feature no longer partakes in valuation. Last Resort is in some sense independent from chain formation, but as I tried to make clear in the above discussion, Last Resort considerations paved the way towards the formulation of the chain condition in (14).

The next question, which will occupy us in the remainder of this chapter, is whether (14) is the right condition to capture all the locality effects in long-distance dependencies. What follows is a long argument that it is.

5.4 Generalized C-trace effect

The first conclusion we can draw from the above discussion is that elements that are forced to form Case chains are more at risk than those elements that can check Case in situ. In particular, they run the risk of being sandwiched between Fin\(^\circ\)\([+T]\) and T\(^\circ\)\([+T]\). All else equal, we predict that raised subjects (or, more generally, arguments having undergone A-movement) won’t be able to form A-bar chains.

As I already demonstrated in Boeckx (2003a), this is the right prediction (see also Richards 2001, and Rizzi and Shlonsky 2007). A great number of languages show that raised subjects resist A-bar movement. To the extent that raised subjects move out of their Case position, they require the use of special strategies.

5.5 Avoiding freezing

Languages indeed resort to special strategies to extract displaced subjects: they use a special complementizer (witness the well-known complementizer alternations known as the *that*-trace effect in Eng-
lish already alluded to above) (18), or they wh-raise the clause that
contains the subject (witness the cases of clausal pied-piping in
Basque and Quechua) (19), or they insert a “resumptive” pronoun
in the original subject position (as in Vata, Yoruba, and Edo) (20a),
or an expletive (as in Dutch and Vallader) (20b). Alternatively, they
suppress subject-verb agreement with wh-extracted subjects (wit-
ess the anti-agreement phenomenon in Turkish, Berber, and
Welsh) (21). (Richards 2001: 147–78 contains a wealth of illustrations
and relevant discussion of the various strategies at issue.)
Complementizer alternation:

(18) a. *Who did John say that t left
b. Who did John say ø t left

Clausal pied-piping:

(19) [Nor joango d ela] esan du Jonek [Basque]
who go aux Q say aux Jon
‘Who did John say will go’

Resumption/expletive:

(20) a. Dè òmwán nè ó dé èbé [Edo]
Q person that he buy book
‘Who bought a book’
b. Wie denk je dat er komt [Dutch]
who think you that expl comes
‘Who do you think is coming?’

---

9 I put resumption here in quotes because the resumption pattern in Vata is quite peculiar. Typically, resumptive pronouns in highest subject positions are banned (the so-called highest subject restriction; see Borer 1984b; Aoun and Li 1990; McCloskey 1990; Boeckx 2003a). Vata seems to behave the opposite way, as it limits resumptive elements to subject positions. I think that Vata falls into place if we treat the so-called resumptive pronoun in subject position as an agreeing expletive (see Rezac 2004 for extensive discussion of this option on the basis of Colloquial Czech); that is, if we reduce the Vata case to the pattern found in Dutch. (For independent arguments from Yoruba that subject-resumptives are expletives, see Adesola 2005.)

10 Boeckx (2003a: 43) shows that similar strategies are used to allow extraction of possessors (highest subjects within NPs).
Anti-agreement:

(21) a. Y dynion a ddaeth
    the men that is-come
    ‘The men that came’
b. *Y dynion a ddaethant
    the men that are-come
c. Chwi a ddaeth
    You(pl) that is-come
    ‘You who came’
d. *Chwi a ddaethoch
    You that are-come

All the strategies just discussed are ways of avoiding further movement of displaced subjects, either literally (via clausal pied-piping, or via filling the displaced subject slot by a resumptive/expletive pronoun), or by modifying the status of the extraction site (i.e., avoid the creation of a full-blown Case-checking site in the sense above).

Based on the definition of checking site put forth above, viz. a pair of heads bearing the relevant feature, I would like to suggest that modifying the featural make-up of either head suffices to avoid the formation of a checking site. For subjects, this means either modifying $T^\circ$ or $\text{Fin}^\circ$. To be of any relevance to Case chains, modification could take either of two forms: either complete absence of $T^\circ$ or $\text{Fin}^\circ$, or removal of their ability to value the $T$-feature on DPs (i.e., $[+T]$).

I will explore the latter possibility here, leaving open the possibility that the entire head may go missing.\footnote{My reason for doing so is that $T$ and $\text{Fin}$ play cartographic roles, and accommodate elements that would lack a proper host if $T$ and $\text{Fin}$ went completely missing.}

Consider again the configuration in (11a), repeated here as (22), which defines a Case-checking domain requiring movement.

(22) $[\text{Fin}^\circ [+T] [\_ T^\circ [\_ [\_ \phi] ]]]$

The most minimal adjustment to break the lethal Case-checking conspiracy of (22) is either remove the value of the $T$-feature on $\text{Fin}^\circ$, or remove the ability of $T^\circ$ to establish a Spec-Head relation with a DP ($[[\_ \phi]]$). $T$ can keep its $T$-feature value to license/value Case on DP, but if it cannot host the DP in its specifier (via the
establishment of a Probe-Goal relation through $\phi$; cf. Chapter 3), then the DP cannot be sandwiched between $\text{Fin}^\circ$ and $T^\circ$. In turn, if the $T$-feature on $\text{Fin}^\circ$ is deactivated (if it lacks a value), then it cannot value the DP’s Case feature. In other words, what we need to avoid “Case-freezing” is either an uninflected $T^\circ$, or an inactive $T$-feature on $\text{Fin}^\circ$. In the former case, SpecTP cannot be defined, in the latter case, $\text{Fin}^\circ$ and $T^\circ$ will not establish any featural transaction, meaning that $\text{Fin}^\circ$ will not project. The product of Merge will be a case of adjunction. $\text{Fin}^\circ$ will be present semantically and maybe even morphologically, but syntactically, the only labels at the edge of the clause will be ForceP and TP.

In sum, the strategies required to extend the chain of a subject that would otherwise freeze due to Case are anti-agreement (23a) or complementizer alteration (23b).

(23)  
  a. $[\text{Fin}^\circ [+T] [\_ T^\circ]]$
  b. $[\text{Fin}^\circ[T] [\_ T^\circ[-\phi]]]

I will now turn my attention to situations illustrating either of these two strategies.

5.5.1 Anti-agreement

Evidently, the use of one or the other strategy to avoid freezing will be more clearly manifested in some languages than in others. For example, the role of inflection in extraction will be clearer in languages where inflection is rich and overtly manifested. It is there that changes in inflection are most dramatic. Languages like English, where agreement morphology is very poor, will therefore be of little use here. Nevertheless, the fact that chain extension is typically easiest out of non-finite, uninflected clauses is telling. One thinks here of raising or control (on control as raising, see Hornstein 1999, 2003, and Boeckx and Hornstein 2003, 2004, 2006, and Boeckx, Hornstein, and Nunes, in prep.)

(24)  
  a. John is likely [t to be home]
  b. John is eager [t to go home]
Unlike English, quite a few languages also allow for subject movement out of finite clauses, in both raising and control contexts. However, there are severe restrictions on when such movement can take place. Among these is the role of inflection.\textsuperscript{12} Recently, Landau (2004) has gathered an extensive survey of so-called “finite” control phenomena, constructions where movement takes place out of finite clauses.\textsuperscript{13} (See also Rodrigues 2004 and Fujii 2005, 2006.)\textsuperscript{14}

\begin{enumerate}
\item (25) Hem kivu se yelxu ha-bayta mukdam [Hebrew]
\begin{itemize}
\item they hope that will-go,3pl home early
\end{itemize}
\begin{itemize}
\item ‘They hoped to go home early’
\end{itemize}
\item (26) I Maria prospathise na divasi [Modern Greek]\textsuperscript{15}
\begin{itemize}
\item the Maria tried,3sg C-subjunctive read,3sg
\end{itemize}
\begin{itemize}
\item ‘Mary tried to read’
\end{itemize}
\end{enumerate}

Landau documents numerous cases where extraction is allowed out of subjunctive clauses or is limited to 3rd person agreeing subjects. In other words, extraction can take place if the finite clause out of which extraction takes place is temporally deficient (subjunctive) or $\phi$-deficient (limited to 3rd person agreement, hence, as I will assume

\begin{enumerate}
\item Boeckx, Hornstein, and Nunes (in prep.) discuss others. In particular, Boeckx, Hornstein, and Nunes focus on the fact that A-movement out of an argument clause appears to violate the A-over-A, and show that raising/control out of finite clauses is only possible when the clause itself does not count as a proper target for movement.
\item It should be noted that Landau does not assume that actual raising takes place in such cases. For arguments that it does, see Hornstein (2003), Boeckx and Hornstein (2006), and Boeckx, Hornstein, and Nunes (in prep.).
\item Rodrigues ties the possibility of movement from subject to subject position (restricted to 3rd person elements) in (i) to the loss of “strong” (Italian-like) inflection in Brazilian Portuguese, which fits perfectly with the story developed here.
\item (i) a. O Joao disse que comprou um carro novo [Brazilian Portuguese]
\begin{itemize}
\item the Joao said that bought a car new
\end{itemize}
\begin{itemize}
\item ‘Joan said he bought a new car’
\end{itemize}
\item b. O Joao disse [que <Joao> comprou um carro novo]
\item Similar facts hold for standard raising cases (see Alexiadou and Anagnostopoulou 1998).
\item (i) Ta pedhia dhen fenonte na doulevoun [Modern Greek]
\begin{itemize}
\item the children not seem,3pl C-subjunctive work,pres
\end{itemize}
\begin{itemize}
\item ‘The children do not seem to work’
\end{itemize}
\item For raising out of subjunctive clauses, see Uchibori (2000) for Japanese and Uriagereka (2006) for Romance.
\end{enumerate}
here, the result of default morphological assignment, or the result of a covert expletive). As Landau observes, deficiency is not exclusively expressed on T. Several languages (e.g., Modern Greek) do not resort to special morphology for subjunctive clauses, as distinguished from the morphology used for declarative clauses. Instead, they use a special complementizer. So deficiency is to be understood in reference to the C-T complex, not T or C alone, as one would expect if A-chains are defined in terms of Fin-T, as argued above.

The overall deficiency of finite clauses out of which A-movement can take place is summarized in the following table (adapted from Landau 2004):

<table>
<thead>
<tr>
<th>movement</th>
<th>no movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>infinitive</td>
<td>C-subjunctive</td>
</tr>
<tr>
<td>T</td>
<td>$-T, -\phi$</td>
</tr>
<tr>
<td>C</td>
<td>$-T$</td>
</tr>
</tbody>
</table>

The overall message in (27) is that A-chain extension for subjects is licit only if intermediate valuing sites (C-T complex) are defective, tense-wise\(^{17}\) or $\phi$-feature-wise.\(^{18}\) This is exactly what the present approach expects.

\(^{16}\) It is interesting to note that numerous languages require tense/mood-shift from realis to irrealis (“subjunctive”) when wh-movement takes place (for Hausa, see Tuller 1986; for Palauan, see Georgopoulos 1985).

\(^{17}\) On T-defectiveness in subject-raising/control out of “finite” clauses, see Fujii (2006) for a particularly strong case made on the basis of Japanese.

\(^{18}\) In the text, I focus on displaced subjects, but it is interesting to note that anti-agreement also applies to displaced objects triggering strong agreement, as in Karitiana (i) and Yimas (ii). (Data from Richards 2001.)

(i) Mora-mom y- 'it ti-/*i- oky-t [Karitiana]
   what-Abs my father OT/3,Obj kill non-fut
   ‘What did my father kill?’

(ii) a. Wara ipa- na -am-n [Yimas]
    what pl.Abs Def eat Pres
    ‘What are we going to eat?’
   b. *Wara na- kay am-n
    what 3sg.Abs pl,Nom eat Pres
    ‘What are we going to eat?’
Turning to A-bar chain extension of displaced subjects, we find exactly the same situation as the one just discussed for A-chain extension. Either the complementizer is changed, or the inflection is.

As Richards (1997, 2001) documents, languages vary as to how “anti-agreement” is expressed. (Unless otherwise noted, the data on anti-agreement come from Richards’s work.)

Some languages limit the amount of subject agreement to 3rd person, instead of using the fully agreeing form typically associated with subjects. This is the case of Fiorentino and Trentino.

(28) Quante ragazze gli e/*le sono venuto con te [Fiorentino]
how-many girls it is/they are come with you
‘How many girls came with you?’

(29) Quante putele e vengnu/*le sono vegnude con ti [Trentino]
how-many girls is come/they are come.pl with you
‘How many girls came with you?’

Other languages, like Selayarese, Chamorro, and Kinande, eliminate the agreement marker entirely (sometimes replacing it with

Similar facts obtain in the context of object extraction in Bantu, as mentioned in Baker and Collins (2006), who note that object extraction takes place only if the object does not trigger agreement on linkers.

The ban on extraction out of fully agreeing/displaced objects is very clear in Basque (see Boeckx 2003a for data and discussion).

Such considerations suggest that shifted objects “culminate” in a configuration like (iii), which is on a par with the Fin-T combination characterizing raised subjects.

(iii) [T^c[+T] [v^c[+c]]]

Phillips (1993) argues that anti-agreement forces the verb to stay low. I find this very interesting in the context of the present work, which relates projection (head-chains), and movement (XP-chains).

Richards (2001: 154–5) observes that anti-agreement is suspended in several languages, including Chamorro, in case the extracted subject binds an anaphor or pronominal variable. Witness (i).

(i) ?Hayi ha- li’i’ gui’ [Chamorro]
who 3sg.erg see him
‘Who saw himself?’

For some remarks on Binding, see Section 5.8.5.
a “wh-agreement” marker, as in Chamorro\(^21\) and Kinande), or replace the finite agreeing verb form with a participle (Berber, Turkish).\(^22\)

\(\begin{align*}
(30) & \text{Inai ng erang(*-i) loka} & \text{[Selayarese]} \\
& \text{who intrans brought.3abs banana} \\
& \text{‘Who brought bananas?’} \\
(31) & \text{Hayi fuma’-gasi/*ha-fa’-gasi i kareta} & \text{[Chamorro]} \\
& \text{who WH.washed/3sg.washed the car} \\
& \text{‘Who washed the car?’} \\
(32) & \text{IyOndI y’ u-ka-langIra/*a-ka-langIra Marya} & \text{[Kinande]} \\
& \text{who-clt C-clt wh-pres-see/3sg-pres-see Marya} \\
& \text{‘Who sees Marya?’}
\end{align*}\)

\(^{21}\) The -um- marker in Chamorro is not only a wh-extraction marker, but is also used in the context of control (A-movement) (i).

(i) Malagu’ gui’ bumisita si Rita \[\text{[Chamorro]}\]  \\
\text{want he AT-visit Unm Rita} \\
\text{‘He wants to visit Rita’}

This may be taken as yet further evidence that A- and A-bar extractions are indeed on a par, as argued here, since the same marker is used for both A- and A-bar extension.

\(^{22}\) Apart from impoverishing subject agreement, several languages resort to the anti-passive strategy to extract subjects. As its name suggests, anti-passivization is, in essence, the opposite of passivization. Whereas passivization can be said to turn the logical subject into an oblique and promote the object, anti-passivization leaves the subject intact, and turns the object into an oblique/adjunct. More specifically, under anti-passivization, a transitive subject “exchanges” its ergative agreement morphology for absolutive agreement (agreement typically associated with complements, and typically null).

\(\begin{align*}
(i) & \text{a. x-ø-s-watx’e naj hun-ti’} & \text{[Jacaltec]} \\
& \text{Asp-3Abs-3erg-make he one-this} \\
& \text{‘He made this’} \\
& \text{b. x-ø-w-il naj x-ø-watx’e-n hun-ti’} \\
& \text{Asp-3Abs-1 Erg-see him Asp-3.Abs-make-AC one-this} \\
& \text{‘I saw the man who made this’}
\end{align*}\)

Nakamura (1996) notes that anti-passivization, when available in a given language, is obligatory in the case of extraction of transitive subjects, as is the case in Dyirbal, Chukchi, and Inuit.

As Richards (2001: 158–9) notes, the effect of anti-passivization is only available in transitive clauses. Interestingly, Chamorro restricts anti-agreement to transitive contexts.

\(\begin{align*}
(ii) & \text{I famagu’on man-nango/*mu-nango gi tasi} & \text{[Chamorro]} \\
& \text{The children PL-swim/AT-swim LOC sea} \\
& \text{‘The children are swimming in the sea’}
\end{align*}\)
Interestingly, in many languages, the impoverished agreement morphology used in the context of subject extraction is identical to the morphology that is associated with “in-situ” subjects, that is, those subjects that fail to raise to the canonical subject position, and remain within VP. (A similar fact also obtains in the context of subextraction, as I discuss below.)

This immediately links anti-agreement to the presence of an expletive/resumptive-like element in several languages (Vata, Vallader, Dutch, Yiddish), and to Rizzi’s (1982) insight that in pro-drop

\[\text{(i)}\]

(a) *Who do you suppose that ___ will leave early?  
(b) %Who do you suppose that’ll leave early?  
(c) *Who do you suppose that ___ is leaving?  
(d) %Who do you suppose that’s leaving?  
(e) *Who did you think that ___ would leave?  
(f) %Who did you think that’d leave?
languages like Italian subject extraction proceeds from a VP-internal position, not through SpecIP (a point that is explored in great detail, and strengthened in Gallego 2007; see also below).

(37) Ver hot er moyre az es yet kumen [Yiddish]
who has he fear that it will come
‘Who does he fear that will come?’

Taraldsen (2001) argues on the basis of comparative evidence from Vallader that the well-known complementizer alternation in the context of subject extraction in French should be analyzed on a par with Yiddish and Dutch. That is, French \( \text{qui} \) is a reduced form of \( \text{que} + \text{il} \text{expletive} \).24

(38) Qui Jean a-t-il dit qui/*que est venu [French]
who Jean has-he said that is come
‘Who did Jean say came?’

In such circumstances, the extracted element fails to agree because some other element takes care of agreement, not because agreement itself is impoverished. But the logic of the situation is the same: the element heading the A-bar chain does not occupy a full-blown Case-checking site.

5.5.2. Complementizer-manipulations

Let me now turn to instances of so-called complementizer alternation in the context of subject extraction, which I will argue is but another manifestation of languages manipulating the C-T complex (in this case, the C-part of the complex) to allow chains to extend beyond what would otherwise be their culmination point.25

Perhaps this effect can be assimilated to instances of subject extraction in the presence of an expletive/resumptive (in this case, verbal) element.

24 Rizzi and Shlonsky (2007) provide several arguments against equating –\( i \) in \( \text{qui} \) with expletive \( \text{il} \). However, the logic of Taraldsen’s argument can be preserved if we take –\( i \) to be a C-type expletive distinct from \( \text{il} \), and more akin to the agreeing subject expletive/resumptive found in Vata, Yoruba, and other languages (cf. n. 9).

25 The role of “complementizers” in extraction may also be reflected in the nominal domain, as Gavruseva (2000), Boeckx (2003a), and Bošković (2005a) have argued (these authors implicate the nature of determiners in their accounts). I refrain from going into a discussion of extraction from nominals, as Pereltsvaig (2007) shows that many, perhaps most, cases of extraction are better analyzed as instances of partial copy pronunciation, as opposed to instances of extraction.
A well-known manifestation of the complementizer alternation at issue is the *that*-trace effect in English, first observed in Perlmutter (1971).

(39) a. *Who did John say that t left
    b. Who did John say ø t left

But English is certainly not alone in manipulating its complementizer system to stretch subject chains. Fassi-Fehri (2005) offers what looks like similar data from Arabic, and Gracanin-Yuksek and Takahashi (forthcoming) do the same for Haitian Creole. Kandybowicz (2006) shows that Nupe behaves exactly like English, not only in displaying *that*-trace effects in the same circumstances, but also in the range of strategies it uses to obviate such effects.

The well-known complementizer alternation in Irish arguably falls into the same category. Like English, Irish uses a special complementizer for extraction.

(40) Dúirt sé gur bhual tú é \[Modern Irish\]
    ‘He said that you struck him’

(41) An t-ainm a hinsiodh dúinn a bhí ar an áit
    ‘The name that we were told was on the place’

Creoles like Gullah also provide similar evidence. Here the complementizer used for extraction is morphologically distinct from the corresponding declarative complementizer, which corresponds to a verb of saying/quoting (as in many African languages such as Ewe or Edo and many South-East Asian languages such as White Hmong). (Data from Mufwene 2001.)

(42) This da young man ø/*say come yah yesiday \[Gullah\]
    ‘This is the young man that came here yesterday’

(43) Faye answer say/*ø Robert coming
    Faye answered COMP Robert coming
    ‘Faye answered that Robert was coming’

So-called wh-agreement languages (Chamorro, Selayarese, etc.) also offer evidence of the relevant sort, once properly interpreted. As I
argued in Boeckx (2003a, 2007a), wh-agreement is a misnomer. It is not an instance of $\phi$-feature on complementizers valued by the moving wh-phrase. Instead, wh-agreement refers to a morphological change affecting the verbs/complementizers lying along the path of wh-movement. The change concerns the Case/agreement relation that typically holds between the subordinating verb and the embedded complementizer. When wh-movement happens, the Case/agreement relation is affected (in many cases, such as Selayarese, the case/agreement relation disappears). Consider, for example, the following sentences from Selayarese.

(44)  
La-$\tau$alle-i doe?-iňjo i Baso?  
3-take-3 money-the h Baso?  
'Baso? took the money'

(45)  
Ku-isse?-(i) *(kuko) la-$\tau$alle-i doe?-iňjo i Baso?  
1s-know-3 COMP 3-take-3 money-the h Baso?  
'I know that Baso? took the money'

(46)  
Apa mu-isse? la-$\tau$alle i Baso?  
what 2fam-know 3-take h Baso?  
'What do you know that Baso? took?'

(44) shows a basic sentence from Selayarese, a VOS language that indicates subject agreement as a prefix on the verb, and object agreement on a suffix. (45) shows that the complementizer is obligatory in embedded declaratives. So is the agreement between the matrix verb and the CP argument (see Finer 1997 for details). (46) shows that all (object) agreement suffixes and the complementizer must be absent in the case of extraction. I take it that all these changes one way or another reflect the deactivation of the T-feature on Fin$^\circ$.

I cannot discuss the topic of complementizer alternation without saying anything about the fact that some dialects of English readily allow (39a) (see Chomsky and Lasnik 1977; Sobin 1987). For such dialects I have to say that the relevant change at the C-level is not morphologically reflected, or that that in (39a) lexicalizes Force for such speakers.26

26 See Lohndal (2007b) for an extension of this proposal to the variation of that-trace effects in Scandinavian languages.
The latter possibility may actually be what underlies “anti-*that*-trace” effects even in dialects where *that*-trace effects otherwise obtain.\textsuperscript{27}

By far the best-known case of “anti-*that*-trace” effect is the so-called “adverb effect,” illustrated in (47).

(47)  This is the man who I said that, next year, *t* will buy your house

Since Bresnan (1977) it has been known that inserting adverbial material between the complementizer *that* and the subject wh-trace obviates a *that*-trace effect (see also Culicover 1993). Interestingly, fronted thematic material in what appears to be the same position does not have this rescuing effect. Witness (48).

(48)  \textit{*This is the man who I said that, your house, *t* will buy \langle your house \rangle next year}

More surprisingly still, Haegeman (2003) showed that adverbial material, once fronted long-distance, loses its rescuing effect (49), a fact which has not yet received an adequate explanation in the literature.

(49)  \textit{*This is the man who I said that, next year, *t* expects that John will buy your house \langle next year \rangle}

The approach developed here can make sense of the data just given once we make the following claims, compatible with the treatment of adjuncts discussed in Chapter 3. What is needed is the following set of assumptions:

(i) Short-distance (i.e., clause-internal) fronted adverbials are base-generated in their surface positions. In the case of (47), I take the adverbial to be located in SpecTopicP, sandwiched between FinP and ForceP (see Rizzi 1997).

(ii) Long-distance fronted adverbials have no other choice but to move to their surface position, on a par with fronted arguments. Unlike base-generated adverbials, they must involve a Probe-Goal relation.

\textsuperscript{27} Kandybowicz (2006) reports that in many dialects where *that*-trace effects otherwise obtain, unstressed or phonetically reduced complementizers allow subject-chain extension. His data are reproduced in (i).

(i) a. \textit{*Who do you think that ___ wrote Barriers?}
   b. \textit{\sqrt{}/?Who do you think th’t ___ wrote Barriers?}
   c. \textit{*Who do you hope for ___ to win?}
   d. \textit{\sqrt{}/?Who do you hope f’r ___ to win?}

I would interpret this fact as indicating that reduced/unstressed complementizers occupy Force‘.
Those fronted elements that participate in a Probe-Goal relation lie on the same plane as the subject to be moved, hence they trigger a minimality effect. Base-generated adverbials are on a separate plane, and therefore do not interfere.

Going back to (47), the linear order makes clear that that is above the position occupied by the adverbial. If I am right in taking the latter to be in SpecTopicP, that means that that lexicalized Force\(^c\), not Fin\(^c\). Being a morphological exponent of Force\(^c\), it does not enter into the computation of the Case chain of the subject that is extended in (47).

A similar explanation can be offered for the lack of that-trace effect in (50).

(50) This is the man who I said that under no circumstances t would want to buy your house

Here it is SpecFocusP that is filled by an adverbial, and which indicates that that is an instance of ForceP.\(^{28}\)

The fact that that-trace effects are notoriously variable should not come as a surprise. It is just another case of lexical variation. The fact that its effects are found in domains that go beyond simple, unembedded clauses (Lightfoot’s 1989 Degree-0) makes it even harder for the learner to fix the right parametric value. The task is made even more difficult by the presence of two complementizer-heads (Force\(^c\) and Fin\(^c\)), both of which appear to be lexicalized as that in various circumstances.

Let me also stress that complementizer manipulation is but one way of stretching subject chains. Since Case-checking sites are defined in terms of both Fin and T, languages may choose to manipulate the other member of the Fin-T unit, in the form of anti-agreement (agreement weakening, resumption, etc.).

\(^{28}\) Interestingly, Drury (1999) and Kandybowicz (2006) note that stressing the verbal element obviates that-trace effects too.

(i) a. *Who do you think that __ wrote Barriers?
   b. \(\vee/p\)Who do you think that __ WROTE Barriers
   c. *Who do you THINK that __ wrote Barriers
   d. A: I didn’t think that John would survive.
   B: \(\vee/P\)Well then, who did you think that __ WOULD?

This fact can be captured under the present system by taking stressed verbal elements to activate the CP-field (FocusP, in particular), and forcing that to occupy Force\(^c\).
Accordingly, the fact that that-trace effects do not obtain in all languages/dialects is thus no ground to relegate such effects to components other than narrow syntax (PF in particular), as Kandybowicz (2006) has recently suggested (building on others, see the references in his paper). A non-narrow syntactic treatment of that-trace effects would be hard pressed to explain the adjunct-argument or short-distance/long-distance adjunct asymmetries in anti-that-trace effects. These appear to require the resources of syntax. Be that as it may, I hope to have established in the preceding pages that displaced subjects are hard to extract.

Languages that lack the possibility to manipulate either Fin° or T° will have no choice but to move the entire clause containing the subject, if they are to place the latter in the right scope position. This is arguably the reason behind the pied-piping strategy used in Quechua and Basque (among others) to extract subjects:

(51) a. *pi-taj Maria-ka [t chayamu-shka-ta] kri-n [Imbabura Quechua] who-wh Maria-top arrive-NL-Acc believe-Agr
   ‘Who does Maria believe has arrived?’

(52) Iza no manasa ny lamba amin’ ny savony [Malagasy]
    who prt pres-at-wash the clothes with the soap
    ‘Who washes the clothes with soap?’
Because the morphology varies according to which element is extracted, it is tempting to treat it as agreement (Rackowski (2002), and Rackowski and Richards (2005) treat it as Case-agreement morphology). For these languages, then, it appears that only agreeing elements can extract. Clearly, this generalization is unexpected under the approach developed here. Fortunately for the present system, it is quite clear that the “subject-only” restriction is not to be interpreted as “only agreeing elements (‘subjects’) can extract.”

The languages imposing a subject-only restriction on wh-movement have several properties that are worth pointing out. First, they are all V-initial languages. Second, wh-questions in such languages take the form of clefts. Third, what is meant by “subject” in the “subject-only” restriction is not quite the typical meaning of subject: it should read “topic,” or, more formally, the element that triggers topic morphology (glossed ‘at’ in (52), ‘tt’ in (53), and ‘ct’ in (54)) on the verb. Both Richards (2001) (for Tagalog) and Pearson (2005)
(for Malagasy) have argued extensively that “subject” is a bad term for the restriction under discussion. Both argue that the restriction is about “topics,” a restriction related to the fact that clefts are used to form questions in the relevant languages.

Fourth, such topic morphology does not encode phi-features, but thematic information (if we insist on it being Case morphology, it must be seen as inherent Case morphology, a reflex of thematic information): ‘at’ stands for “agent topic”, the morphology that is used when agents are extracted. Theme topic (‘tt’) morphology is used when themes are extracted; and Circumstantial topic (‘ct’) morphology is used when instrumentals are extracted.

So the morphology is sensitive to thematic information only. As such, it poses no threat to the claims made in this section.

Note, finally, that the fact that only “topics” can be extracted does not mean that extraction has to take place from a high topic position. Since we are dealing with clefts, I would like to suggest that the instances of extraction in the languages under discussion take the form of (55), with the XP “extractee” base-generated outside the clause and related to its thematic position via an empty operator moved from a clause-internal position to the Force-domain.

(55) XPᵢ [CP Opᵢ C [IP...<OP>]]

Extraction for wh-movement would then take place from a VP-internal position. This would render Malagasy and Tagalog well-behaved under the present set of hypotheses. (Potsdam (2006) has provided independent arguments from Malagasy in favor of a derivation like (55).)

I conclude that the “subject-only” restriction does not constitute a counterexample to the approach developed here.

The last example of subject extraction I would like to discuss in this section is *who* left? The literature goes back and forth as to whether *who* is pronounced in its Case position or its A-bar position; in our terms, in SpecTP or SpecFinP. As far as I am concerned, if *who* has moved to its A-bar position, then the Case site must have been manipulated (either by modifying Tᵢ or Finᵢ). Furthermore, the only way in which *who* would be able to stay in SpecTP is if the CP-domain is not split into ForceP and FinP. As soon as these two
projections are split, movement is required, as Force will not be able to reach who in SpecTP across another element bearing an [\omega]-feature (Fin\(^o\)). (This is the same reasoning that led me to claim that movement of the subject DP to SpecTP is forced if Fin\(^o\)[\_\_+T] is to agree with the subject.)

So the present account is relatively neutral on the position of who. It offers both options: a manipulated Case domain if A-bar movement is to take place, or an unsplit C-domain if movement of the subject stops in SpecTP.

Interestingly, den Dikken (2006) provides evidence for both options, and notes that subjects cannot remain in their Case positions if the C-domain is independently activated, which I take to mean that the SpecTP option is not available if there is independent material requiring the presence of an articulated C-domain, exactly as predicted by the present approach.

### 5.6 Subextraction, CED, and QED

Up to now we have discussed situations where moved elements are at a disadvantage. Because they form non-trivial chains, such elements typically remerge with elements that close them off, and exhaust their derivational potential. Even if they have the right feature to relate to a higher functional head, the chains they will form if they move beyond the first checking site they landed into will be deemed illegitimate at the interfaces, on QED-grounds.

In this section I would like to discuss another way in which displacement significantly restricts chain-formation potential. This time the focus will not be on the moved element itself, but on the elements that it contains.

Previous research, beginning with Cattell (1976) and Huang (1982), has established that extraction out of displaced constituents is generally banned (I return to alleged exceptions in Section 5.7), as is extraction out of adjoined constituents. This is the well-known Condition on Extraction Domain (CED) generalization, which says that extraction out of “non-governed”/thematically related domains is banned.
The only way the present framework could capture the CED without any extra assumption is if the chains formed by movement out of displaced or adjoined constituents end up violating the QED-requirement imposed at the interfaces. I hinted at how this was the case in Chapter 2. Here I want to be as specific as I can about how such QED-violations occur.

The three cases to compare are (56) [licit subextraction], (57) [illicit extraction out of a displaced constituent], and (58) [illicit extraction out of an adjoined constituent].

(56) Who$_i$ did John see [pictures of $t_i$]

(57) *Who$_k$ have [pictures of $t_k$], been $t$, annoying Bill

(58) *Who$_i$ did John arrive [after Bill saw $t_i$]

The key to capture the contrast between (56) and (57)/(58) is the fact that chain links are defined in terms of their sisters. For a chain to be licit, its members must form an unambiguous path—essentially, a straight line from beginning to end, with no interfering point in between that could be taken as an end point. This is what graphs like (15) are intended to capture.

Now consider (56). Assume first, as is standard in the literature, that Case is checked in situ via Agree inside nominals, at least in the case of complements of of-phrases (this is another way to say that such complements do not check the EPP, unlike pre-nominal possessors). This will allow us to set aside the possibility that the chain formed by subextraction has already reached its end point internal to the extraction site. Accordingly, the chain formed by who in (56) is defined by its thematic relation to the head noun picture on the one hand, and its A-bar checking site (final landing site), on the other. The chains formed by movement of who in (57) and (58) are comparable. In (57), the chain of who is also a path between the two defined by its thematic relation to the head noun picture on the one hand, and its A-bar checking site (final landing site), on the other. And in (58), the

---

30 Whether that thematic relation to the head noun is mediated by a functional head like Pred$, as Baker (2003) (among many others) argues, is immaterial to the case at hand.
chain of *who* is defined by its thematic relation to *see* and its A-bar checking site. So the A-bar chains are equivalent, and they are not interrupted by the presence of any checking site occupied by the moving element along the way. They can all be represented as in (59). (I ignore the Case-Agree relation, as this is irrelevant for chain-computation purposes.)

\[
\begin{array}{c}
\text{Force}^*_{[+\omega]} \\
\downarrow \\
\text{Fin}^*_{[+\omega]} \\
\downarrow \\
x \quad \alpha
\end{array}
\]

Since the chains themselves are equivalent, they cannot be the source of the contrast at hand. Instead, the cause must lie in the way the chain links relate to one another. Since we are talking about subextraction, the point of origin must be crucial.

The point of origin of the chain in (56), the $\alpha$-link, is the unmoved *pictures of _*, which forms the trivial chain in (60), connected to its theta-assigner *see* ($\alpha$).

\[(60) \quad x \rightarrow \alpha\]

The point of origin of the chain in (57) is the displaced *pictures of _*, which forms the complete Case-chain in (61).\(^{31}\)

\[
\begin{array}{c}
\text{Fin}^*_{[+\tau]} \\
\downarrow \\
\text{T}^*_{[+\tau]} \\
\downarrow \\
x \quad \alpha
\end{array}
\]

The point of origin of the chain in (58) is the whole adjunct clause *after Bill saw _*, which projects a full clausal skeleton (a projection/chain), featurally unconnected to the main line of the tree.

\(^{31}\) Subextraction is also impossible if it takes place from an A-bar-moved element, whose chain would be defined by ($\alpha$, Fin$^*_{[+\omega]}$-Force$^*_{[+\omega]}$), parallel to (61).
From this description of the point of origin of the chains formed by subextraction in (56)–(58), it is clear that the difference lies in the fact that in the illicit cases, the point of origin is a complete chain/projection that cannot be extended further. In other words, the α-link of the chain formed by subextraction in (57)–(58) cannot be linked to form a path with the other extremity of the chain. Put differently, the links of the subextraction chain cannot form a connected object, and the symmetry between chain and projection breaks down. The situation is similar to one where we would try to relate a thematic domain and a C-domain that are part of a completely disconnected tree (which would correspond to (58)), or having a tree with two root nodes (which would correspond to (57)).
To sum up, the $\alpha$-links in (57)–(58) already have a fully defined path. Forcing them to be part of another path inevitably results in the formation of non-well-defined, hence, ambiguous paths, which are ruled out at the interfaces, where non-ambiguity is required (by QED). By contrast, the $\alpha$-link in (56) is free to be a member of a chain because it itself hasn’t formed a chain/path yet.

The situation we face is one familiar to topologists. Topology is a field that explores how an object can be stretched in various ways. Stretching is allowed so long as no tearing (disconnection) happens. It is the nature of adjoined constituents and displaced constituents that they inevitably lead to a tear.

It is true that a projection can be established by relating two disconnected elements (a specifier and a complement), while a chain cannot. But the reason a chain cannot do so is because a chain is the product of Internal Merge. It is defined over an already pre-existing object, which must be preserved (in a way not unlike Emonds’s Structure-Preservation hypothesis).

In more traditional terms, the upshot of our discussion is that if subextraction is to be successful, the extraction domain must be “open” to extension (a conclusion I reached in Boeckx 2003a; see also Rackowski and Richards 2005 and Gallego 2007 for a similar generalization concerning the ban on extraction out of displaced constituents).

Note that although the present approach captures CED-effects, it does not treat displaced and adjoined constituents completely alike. Although both types of constituents lead to ambiguous chains if subextraction takes place out of them, the cause of ambiguity is somewhat different: ambiguity results in the case of adjoined structures because these do not relate featurally to the domain targeted by subextraction, whereas in the case of displaced constituents, ambiguity is the result of the displaced element having exhausted its chain extension potential.

Accordingly, asymmetries between the so-called subject condition and the adjunct condition are not completely unexpected from the present perspective. The approach to locality developed here is based
on the idea that instructions to the interfaces cannot be ambiguous. It does not say that the source of ambiguity must be the same in all cases of illegitimate objects.

5.7 On the robustness of the CED-generalization

Before turning my attention to another aspect of locality, I should dispel a few doubts cast in the recent literature on the correctness of the CED-generalization. In particular, Chomsky (forthcoming) has argued that some instances of extraction out of Case chains are licit. Likewise, Rizzi (2006), in part building on original observations by Esther Torrego, has argued that extraction out of A-bar chains is sometimes licit. If Chomsky and Rizzi are correct, it is not clear how the present approach would be able to handle this sort of data. For this reason, I want to examine both types of situations, and show that the counterexamples to the original CED are only apparent. There will thus be no need to modify the account of subextraction possibilities developed above.

Let us consider Chomsky’s (forthcoming) cases first. Chomsky focuses on the paradigm in (65). (The judgments are Chomsky’s. CAPS indicate focus. Many speakers consulted regard (65b-b’) as degraded.)

\[(65) \]
\[a. \text{It was the CAR (not the TRUCK) of which they found \}[\text{the driver t}]\.} \]
\[a’. \text{Of which car did they find \}[\text{the driver t}]\?} \]
\[b. \text{It was the CAR (and not the TRUCK) of which \}[\text{the driver t}]\ was found. \]
\[b’. \text{Of which car was \}[\text{the driver t}]\ awarded a prize? \]
\[c. \text{*It was the CAR (and not the TRUCK) of which \}[\text{the driver t}]\ caused a scandal.} \]
\[c’. \text{*Of which car did \}[\text{the driver t}]\ cause a scandal?} \]

Chomsky takes the facts in (65), especially (65b-b’), to suggest that extraction out of derived internal arguments is possible, and goes on...
to develop an intricate phase-based account of the contrast between (65b-b’) and (65c-c’) in terms of minimal search, simultaneous probing, and feature inheritance. I will not summarize his account here (see Broekhuis 2006 for an excellent summary), since it is framed in terms that I have not discussed here. What is more, there are several arguments against Chomsky’s empirical generalization, and therefore against his technical analysis.

Broekhuis (2006) in particular has provided several arguments from Dutch that argue against Chomsky’s view that extraction out of derived complements is possible. Let me repeat Broekhuis’s arguments here.

First, as Broekhuis notes, the Dutch counterparts of the constructions in (65) are of a very restricted sort in the sense that the allegedly extracted PP can only be headed by a limited set of prepositions. In Dutch this set is exhausted by van ‘of’ and over ‘about’; PPs headed by, e.g., clearly locational prepositions are never extracted from DP. This is illustrated in (66) for extraction from object; (66b) is acceptable when the PP is construed as a locational adverbial phrase but not on its intended reading as a modifier of the noun huis ‘house’. It is not at all clear how (66b) is to be excluded under Chomsky’s account.

(66) a. Jan heeft het huis op de hoek gekocht [Dutch]
    Jan has the house on the corner bought
    ‘Jan has bought the house on the corner’

    b. *[Op de hoek], heeft Jan [het huis t] gekocht

Second, it is not so clear whether the van-PP in (67a) (which corresponds closely to Chomsky’s crucial example) is really extracted from the object. Under the right contextual and pragmatic conditions the object de eigenaar ‘owner’ can be replaced by a pronoun. Since pronouns normally resist modification, this suggests that the preposed van-PP does not function as a complement or a modifier of the noun, but rather as an independent adverbial phrase (as suggested in Broekhuis, Keizer, and den Dikken 2003: 258). Note in passing that the preposed van-PP in (67) triggers a contrastive reading, and that apparently the same holds for Chomsky’s examples in (65).
Another piece of evidence in favor of the claim that the preposed van-PP is an independent adverbial phrase is that the preposed van-PP can be modified by a focus particle, whereas a post-nominal van-PP cannot (see Broekhuis, Keizer, and den Dikken 2003: 257); if the preposed van-PP in (68a) originates from within the object DP, the unacceptability of (68b) would be very surprising.

As a final argument against Chomsky’s conclusion, Broekhuis notes that even if we were to maintain that the extracted PP really did originate in the displaced subject, the Dutch facts in (69) show that Chomsky’s contrast between subjects that were internal arguments and subjects that were external arguments does not always obtain.
It seems to me that Broekhuis’s arguments show that something special, other than direct extraction out of displaced subjects, is going on in (65).

As a matter of fact, the claim that sentences like (65b) require a special treatment goes back to Kuno (1973).

Discussing examples similar to (65) (already reported in Ross 1967), Kuno notes that the judgments of such sentences greatly vary among speakers. He goes on to note that for all speakers, subextraction with pied-piping of the preposition is considerably better than subextraction without pied-piping. Contrast (65b–b') with (70a–a'). (Chomsky forthcoming: 13 n. 34 also notes this fact.)

(70) a. *It was the CAR (and not the TRUCK) which [the driver of <which>] was found
   a'. *Which car was [the driver of <which car>] awarded a prize?

Kayne (1984: 189) reports a similar contrast, reproduced in (71).

(71) a. ?Of which words is learning [the spellings <of which words>] difficult
   b. *Which words is learning [the spellings of <which words>] difficult

I take the role of pied-piping to point to the correctness of Broekhuis’s claim that the extracted PPs are modifying adjuncts (which are known to resist P-stranding more than arguments, see, e.g., Hornstein and Weinberg 1981).33

Note that one need not claim that such adjuncts are base-generated in their surface positions, they may well move from within the subject. Since adjuncts, unlike arguments, are not integrated into subject chains, the fact that subjects culminate in SpecTP won’t affect them as much as it does in the case of argument subextraction.34 (We will see later in this study that other types of adjunction, as in the context of resumption, render otherwise impossible extraction possible.)

I conclude from these remarks that the ban on extraction out of displaced subjects is robust.

---

33 On some prepositions that look DP-internal, but aren’t, see Kayne (2005), who explicitly discussed cases involving of.
34 I do not have an explanation for the internal/external argument contrast reported by Chomsky (65b vs 65c). For an attempt to explain the English contrast in a way that could be implemented here, see Gallego (2007).
This does not entail that extraction out of subjects is impossible. The system developed here predicts that such extraction will be possible if subjects haven’t built chains.

This is directly related, of course, to Rizzi’s (1982) seminal work on post-verbal subjects in pro-drop languages. In-situ subjects are known to behave on a par with complements in terms of extraction. The same is true for subextraction. This is easily established on the basis of various languages.

Take English. As Lasnik and Park (2003) point out, extraction from subjects in existential constructions is possible.

(72) which candidate were there [posters of t] all over the town?

The contrast between (72) and (73) follows if (as is standard) we take subjects in existential constructions not to have culminated.

(73) *which candidate were [posters of t] all over the town

A similar contrast obtains in Dutch (data from Broekhuis 2006).

(74) a. Wat hebben er [t voor mensen] je moeder bezocht? [Dutch] What have-3PL there for people your mother visited ‘What sort of people have visited your mother?’

b. *Wat hebben [t voor mensen] je moeder bezocht? What have-3PL for people your mother visited ‘What sort of people have visited your mother?’

Finally, the same pre-verbal/post-verbal asymmetry can be seen in Spanish. (The data are taken from Uriagereka 1988, but I should point out that the database was greatly expanded in Gallego 2007, who has shown in detail that subextraction is only possible from non-displaced subjects).35

(75) De qué conferenciantes te parece que… [Spanish] Of what speakers you seem-3SG that…

a. …(?)me van a impresionar [las propuestas t]? me go-3PL to impress-INF the proposals ‘Which speakers does it seem to you that the proposals by will impress me?’

b. …*[las propuestas t] me van a impresionar? the proposals me go-3PL to impress-INF ‘Which speakers does it seem to you that the proposals by will impress me?’

35 Gallego (2007) also shows that a similar asymmetry between displaced (opaque) and non-displaced (transparent) arguments obtains in the realm of objects, as the current system predicts (cf. n. 18 above).
Let me now turn to alleged cases of extraction out of A-bar chains. Torrego (1985: 31) notes the acceptability of cases like (76) in Spanish.\footnote{36 The Torrego effects are one of the only two cases of acceptable extraction out of A-bar chains that I know of. The other instance, which, unlike the Torrego effects, gives rise to fairly robust acceptability across speakers, comes from the literature on Japanese scrambling.}

\begin{equation}
(76) \ \text{[De qué autora] no sabes [qué traducciones \textit{t}] han won \textit{t} of what author not know-\textit{2SG} what translations have-\textit{3PL} ganado premios internacionales? won awards international}
\end{equation}

‘Which author don’t you know what translated books have won international awards?’

Rizzi (2006: 114) reports similar facts for Italian (77), as do Chomsky (1986b) and Lasnik and Saito (1992) for English.\footnote{37 The examples discussed by Chomsky and Lasnik and Saito involve stranding of the preposition \textit{of}, as in (i).} (The status of these examples varies greatly from speaker to speaker.)

Rizzi (2006: 114) reports similar facts for Italian (77), as do Chomsky (1986b) and Lasnik and Saito (1992) for English.\footnote{37 The examples discussed by Chomsky and Lasnik and Saito involve stranding of the preposition \textit{of}, as in (i).}

(77) \[\text{¿De qué autora no sabes qué traducciones han ganado premios internacionales?}\]

Takahashi (1994) observes that it is possible to scramble out of scrambled phrases, as shown in (i).

\begin{equation}
(\text{i}) \ [\text{Sono hon-o}]_{1} \ \text{John-ga} [\text{Mary-ga t}]_{1} \ \text{katta to}]_{1} \ \text{Bill-ga itta to omotteiru [Japanese]}
\end{equation}

‘That book, John thinks that Bill said [that Mary bought]’

Bošković and Takahashi (1998) explain away this exception of the ban on extraction out of displaced constituents by analyzing scrambled phrases as being base-generated in their surface positions. But there are rather strong reasons to doubt the base-generation analysis of scrambled phrases (see Boeckx 2003c; Kato 2006; Miyagawa 2006). Instead, I proposed in Boeckx (2003c) that at least some instances of Japanese-type scrambling amount to resumption (with a covert resumptive pronoun) (see Lee (2006) for supporting evidence from Korean). If this is correct, the violation of the ban on extraction out of displaced constituents in (i) is on a par with the well-attested fact that resumptive structures “save” island violations. Boeckx (2003a) develops a theory that accounts for why resumption has this effect. I return to resumption and other instances of island repair below. For a different analysis of (i), compatible with the present analysis, see Boeckx (2007d).
Such facts led Rizzi to claim that only the head of a displaced constituent is frozen in place. Rizzi’s characterization goes against our treatment of subextraction above. However, as Gallego (2007) and Gallego and Uriagereka (2007b) show in detail, Rizzi’s characterization is not correct. Specifically, they argue that the subextracted element in (76) is best analyzed as an aboutness phrase base-generated outside the A-bar chain headed by qué traducciones. In other words, subextraction does not take place in (76) (and, by extension, in (77)–(78)).

Gallego and Uriagereka provide several empirical arguments in favor of treating the “subextracted” element in (76) as an aboutness phrase. First, they note that the verb used by Torrego (saber ‘know’) readily makes room for such an aboutness phrase.

They also point out that the addition of an aboutness phrase in examples like (76) renders the example unacceptable. Witness (80).

(i) ??Which author do you wonder [how many books of <which author>] Bill read <how many books of which authors>

Though marginal, (i) is reported to be better than (ii).

(ii) *Which author do you believe that many books of <which author> will be read by Bill

Most speakers I consulted judge (i) and (ii) as equally unacceptable. I do not have an explanation for why some speakers find (i) somewhat better.
Gallego and Uriagereka argue that the presence of an additional aboutness phrase forces subextraction. Since this leads to unacceptability, it seems safe to conclude that subextraction in (76) is only apparent.

In sum, subextraction possibilities appear to be limited to cases of extraction out of non-displaced (theta-marked) elements, as the present approach leads us to expect.

5.8 Island “repair”

So far I have focused on how QED-considerations give rise to well-known island effects. In this section I turn my attention to instances of island circumvention, and show how these can be not only accommodated, but in fact predicted under the approach to locality developed here. The term “island circumvention,” or more often “island repair,” is often used to talk about how certain instances of resumption or ellipsis lead to acceptable results in contexts where we would otherwise expect island effects.

Pied-piping is also often mentioned in this context. But here I would like to expand the scope of island circumvention phenomena by including cases like wh-in-situ and rules of construal like binding and control. I will show how all these phenomena can be unified under the present approach.

5.8.1 Preliminary remarks

Let me start by pointing out that a term like island repair is not a very good one to use in a framework like the present one, which does not take islands to be primitives of the theory. For me island is a useful descriptive notion. It refers to an effect that can surface in
the data, not something that certain operations can cover up, or repair.

If island effects don’t arise, it is because syntax has found ways to abide by the QED-requirement imposed at the interfaces. In this sense my approach will be very different from the one taken in recent years by Lasnik (2001, 2005), Merchant (2001), and Richards (2001), for whom operations taking place at the interfaces (e.g., ellipsis) can hide things that syntax failed to execute properly.

Needless to say, the repair phenomena, like virtually all other phenomena dealt with in this study, have been subject to extensive, extremely detailed investigation, and I cannot hope to provide a comprehensive survey of all the facts discussed in the literature. I can only hope to show how the major generalizations can be accommodated under the present framework, in some cases by substantially revising them.

As I have often done in this study, I will first state the basic intuition that I will pursue in the following pages, before addressing the details of implementation, so that the reader can keep the bigger picture in mind when looking at specific examples. I submit that the overall strategy behind island circumvention consists in avoiding the formation of an ambiguous chain—a chain that would not have a unique beginning or a unique end. In all cases, island circumvention will be achieved by splitting a chain that would otherwise be ambiguous into two unambiguous chains. In all but one case (pied-piping), chain splitting will require the introduction of a resumptive element that starts its derivational life as a twin of the moving element (ultimately the antecedent of the resumptive pronoun), and performs part of the checking task of the moving element. I will argue that the chain splitting strategy is used not only in standard cases of resumption (as I originally proposed in Boeckx 2003a), but also, more controversially, in cases of lack of island effects involving wh-in-situ, ellipsis, and binding/control. In the case of pied-piping, the proxy element will not be a resumptive item, but will be the entire constituent containing the element that would otherwise move and form an ambiguous chain.
5.8.2 Resumption

I will begin my discussion of chain splitting with resumption, since in this case my task has been made easier by the fact that chain splitting has already been argued to be able to capture the relevant facts (see Boeckx 2003a; see also Henderson 2006; Belletti 2006; Kato 2005; Fujii 2005; Lipták and Vicente forthcoming; Taraldsen 2005; Boeckx and Grohmann 2004; Meral 2006; and Salem 2001.)

Broadly speaking, there are two traditional approaches to resumption. One takes resumption not to involve movement of any sort (see, e.g., Chomsky 1977; McCloskey 1979, 1990, and much subsequent work). The antecedent of the resumptive pronoun is base-generated in its surface position, and relates to the resumptive pronoun via a construal rule at SEM. Since resumption does not involve movement, the base-generation approach predicts the lack of island effects we typically find when resumption is involved. The other approach to resumption assumes that movement is involved, and takes the resumptive pronoun to be a lexicalized trace (see, e.g., Perlmutter 1972; Pesetsky 1998) at PHON. Such an approach is committed to the idea that island effects are in some sense (rarely made precise)\(^{38}\) PF-phenomena.

In Boeckx (2003a) I proposed a chain-splitting analysis of resumptive patterns that combine insights from the two traditional approaches.

Together with the base-generation analysis I argued that resumptive pronouns are not added at PHON to repair an instance of movement that would cause a PF-crash. Resumptives are there from the start of the derivation. But unlike the base-generation analysis, I maintained that the antecedent of the resumptive pronoun moves from within the island. In fact, I argued that the antecedent moves from a position that is very close to, but not identical to the position it would move from, were the resumptive pronoun absent. Specifically, I argued that the resumptive pronoun and its antecedent are merged as a big nominal unit, and later get

\(^{38}\) But see Hornstein, Lasnik, and Uriagereka (2007) for an attempt.
split by movement of the antecedent, as illustrated in the derivation in (81). (English examples are used here for convenience. For data from languages with more robust resumptive patterns, see Boeckx 2003a.)

(81) [Which man] did Sue say that Mary met the girl who claimed that [he [t]] kissed Sally?

The derivation in (81) is akin to that of preposition-stranding or the one proposed for quantifier-floating in Sportiche (1988).

I argued in Boeckx (2003a) that the reason movement from within an island domain is possible is because I took islands pretty much like I have analyzed them in this chapter, not as absolute, opaque domains, but as the results of “too much” checking (resulting in ambiguous chains). The proposal I put forth in Boeckx (2003a) is this: the presence of a resumptive pronoun inside the island essentially allows the moving element to not be involved in any checking relation other than the one relating it to its final landing site. In other words, the resumptive pronoun acts as a proxy checker for the moving element. It is the resumptive pronoun that forms a chain internal to the island. Since the resumptive pronoun does not move out of the island, it does not form an ambiguous chain, and since the element that moves from within the island does not do any checking inside the island, it too does not form an ambiguous chain.

The checking done inside the island by the resumptive pronoun is of the “A”-type (Case/agreement), hence the frequent characterization of resumptive pronouns as minimal DPs. In terms of the present chapter, we can say that the resumptive pronoun takes care of the Fin-T checking site, allowing its antecedent (the moving element) to land in another checking site (Force-Fin site).

Thus, resumptive elements are there as surrogates, to one of the checking sites that would make the antecedent’s chain ambiguous.

Crucially, for the chain-splitting account to work, it is imperative that the element moving out of the island not be involved in any featural transaction inside the island. For if it were, its chain would be no different from the chain formed by movement out of an
adjunct, or a displaced element. Recall that in these cases, subextraction is ruled out because the moving element partakes in a featural transaction with an element that itself forms a full-fledged chain—ultimately giving rise to a QED-violation.

Accordingly, I follow Boeckx (2003a) in claiming that the antecedent of the resumptive pronoun merges with the latter as an instance of adjunction (no feature checking).

Evidence from the absence of any featural exchange between the resumptive pronoun and its antecedent comes from the anti-agreement effects I documented extensively in Boeckx (2003a). As I showed there, there exist numerous cases where a resumptive pronoun and its antecedent fail to match (i.e., agree) in Case or $\phi$-features. The following examples provide a sample.

Anti-Person Agreement

(82) A Alec, tusa a bhfuil an Béarla aige [Irish]
'Hey Alec you aN is the English at-him
'Hey Alec you that know(s) English'

Anti-Number Agreement

(83) Na daoine a chuirfeadh isteach ar an phost sin [Irish]
'the men C put-cond-3sg n for the job that
'The men that would apply for that job'

Anti-Gender Agreement

(84) Dè a’mhàileid a chuir thu am peann ann [Sc. Gaelic]
'which the.bag-Fem C put you the pen in-3-Masc
'Which bag did you put the pen in?'

Anti-Case Agreement

(85) a. Bha thu a’geàrradh na craobh [Sc. Gaelic]
'be-pst you cutting the tree-Gen
'You were cutting the tree'
b. Dè a’chraobh a bha thu a’geàrradh
which tree.Nom C be-pst you cutting
'Which tree were you cutting?'
As Merchant (2001) has noted independently, antecedents of resumptive pronouns act in terms of Case and phi-features as if they were outside the A-system, unrelated to the gap position filled by the resumptive pronoun. And yet, as numerous studies have shown (see Guillot 2006; Salzmann 2006, in particular), reconstruction effects and other tests such as superiority (Boeckx 2003a; Boeckx and Hornstein forthcoming) strongly suggest that the antecedent of the resumptive pronoun originated from a site that is marked by the resumptive pronoun.

A chain-splitting derivation captures both aspects of this paradoxical situation by taking the antecedent of the resumptive pronoun out of the A-system in which it originates, letting the resumptive pronoun (a quintessential A-element, made up of phi-features only) take care of the A-system, and turning the antecedent into an adjunct.

The resumptive pronoun has the effect of dissociating the path of its antecedent from the domain it originated from. (I argued in Boeckx (2003a) that in those cases where the resumptive pronoun and its antecedent appear to match in phi-features, it is not a case of matching under Agree, but simply a case of accidental matching.)

To put it differently, the resumptive pronoun puts its antecedent on a parallel track, allowing it to project a chain of its own. It is one rare case where being an adjunct gives an element more freedom and flexibility of movement (typically, adjuncts are subject to more severe locality conditions, for reasons I return to at the end of this chapter). Once detached featurally from its extraction domain, the antecedent of the resumptive pronoun essentially forms a trivial chain. Although it moves, its chain merely consists of the one checking site, corresponding to its final landing site.

The chain-splitting analysis of resumptive patterns advocated here is another illustration of two key ideas in this work, and in minimalism more generally: first, the idea that movement and feature-checking are dissociated (the idea behind “Agree”), and second, the idea that Merge is source-independent: Internal Merge = External Merge. The chain-splitting analysis uses both ideas to capture
the effects of base-generation without abandoning the idea that movement is involved in resumption.

The present analysis also accommodates the findings of works that take the antecedent of the resumptive pronoun to be a proleptic element, adjoined to the edge of the island and moving from there to its surface position (see Iatridou 1995, and especially Salzmann 2006).

Treating resumptive patterns in terms of chain-splitting has a number of virtues, which I would like to briefly mention now.

First, by treating the antecedent of the resumptive pronoun as an adjunct, it immediately explains why some languages use the very same complementizer for A-bar dependencies with pure adjuncts like why and with resumptive structures.

Thus, as is well known, Irish uses a distinct complementizer form in the context of resumption (aN (86)) as opposed to aL, used for standard extraction (87)).\(^{39}\) Irish also uses the same aN complementizer for wh-adjuncts (88).

\[(86) \text{An fear aL bhuail tú} \quad \text{[Irish]}\]
\[
\text{the man aL struck you} \\
\text{‘The man that you struck’}
\]

\[(87) \text{An fear aN bhuail tú é} \text{ the man aN struck you him} \text{[Irish]}\]
\[
\text{‘The man that you struck (him)’}
\]

\[(88) \text{Cén fháth a-r dhúirt*/a dúirt tú sin} \quad \text{what reason aN said aL said you that} \]
\[
\text{‘Why did you say that?’}
\]

Second, since the function of resumptive pronouns is to take elements outside of the Theta/Case-domains, it follows automatically that there will be no resumptive strategies with “pure” adjuncts, an old observation in the resumptive literature that is now derived as a theorem under the present system. Since pure adjuncts do not partake in featural transactions, they do not require a proxy element to alleviate their checking burden.

\(^{39}\) I here focus on the basic distribution of complementizers in Irish. For more complex patterns, see McCloskey (2002) and Boeckx (2003a).
Third, by merging the resumptive pronoun and its antecedent, I am claiming that they share semantic properties (like adjuncts and their hosts do). I can thereby explain why wh-phrases linked to resumptives have semantic features typical of pronouns (discourse-linked, more referential, etc.), in a more straightforward manner than is available if resumptive pronouns are introduced at PF.

For example, it is known that when a trace in a relative clause is c-commanded by a quantified expression, the sentence is ambiguous between a “single-individual” and a “multiple-individual” interpretation. Consider (89), from Hebrew.

(89) Ha-iSa Se kol gever hizmin hodeta 
The-woman that every man invited thanked
   a. ‘the woman every man invited thanked him’
   b. ‘for every man x, the woman that x invited thanked x’

As Sharvit (1999) points out, building on Doron (1982), if a resumptive pronoun is used in the same environment, the sentence ceases to be ambiguous (at least, as Sharvit notes, when the sentence is uttered out of the blue). The only available reading is the single-individual reading, characteristic of more referential elements.

(90) Ha-iSa Se kol gever hizmin ota hodeta lo
The-woman that every man invited her thanked to-him
   a. ‘the woman every man invited thanked him’
   not: b. ‘for every man x, the woman that x invited thanked x’

Fourth, since the present account does not treat islands in terms of rigid, geometric domains, but in terms of checking sites involved in a chain, it correctly predicts that resumptive structures can be found outside traditional island contexts like adjunct clauses, etc. From the present perspective, nothing prevents the use of resumptive pronouns in main clauses (instances of what Aoun, Choueiri, and Hornstein 2001 are forced to call apparent resumption).

(91) Miin i2ft-o 
Who saw.2sg-him
‘Who did you see?’
Let me repeat that my intention here is not to cover all the properties of resumptive structures, only their salient features. Boeckx (2003a) discusses various intricate patterns that I cannot go into here. All I hope to have done is give the reader the essential logic that the present framework makes to capture instances of resumption. Let me now turn to less obvious patterns of resumption.

5.8.3 Wh-in-situ

The first pattern I would like to discuss is the case of wh-in-situ and the absence of island effects that often obtains with “covert wh-movement” (Huang 1982). My reason for doing so here is because my analysis of resumption naturally extends to some cases of wh-in-situ once Watanabe’s (1992) analysis of the latter is adopted.

The phenomenon of wh-in-situ played a crucial role in the development of comprehensive theories of chain licensing in the GB framework (see especially Huang 1982 and Lasnik and Saito 1984, 1992). For example, the existence of island effects with wh-adjuncts provided the strongest possible argument for covert (phrasal) movement.

\[ (92) \quad \text{‘You believe Lisa why came DE claim’} \]

But the importance of wh-in-situ has declined over the years for, I think, two major reasons.

First, the phenomenon turned out to be much less uniform than what one may have suspected at first. “Wh-in-situ” is a cover term for many phenomena, with vastly diverging properties, and it is quite clear that no uniform analysis will do to capture them all. Thus, Chinese wh-in-situ is distinct from Japanese wh-in-situ, which in turn is distinct from Iraqi Arabic, French, or Hindi wh-in-situ, which is again distinct from Malay wh-in-situ (see Watanabe 2001 for a good survey; see also Simpson 2000 and Richards 2001). For instance, Chinese wh-in-situ only shows island effects with wh-adjuncts; Japanese wh-in-situ in addition shows wh-island effects; Iraqi Arabic, French, and Hindi wh-in-situ elements are (tensed)
clause-bounded; and Malay allows wh-arguments, but not wh-adjuncts like “why” to remain in situ. And this quick survey says nothing about wh-in-situ in multiple questions, which behave differently in English from the way they behave in Lebanese Arabic (see Aoun and Li 2003).

Second, it has never been clear why in-situ wh-arguments do not show island effects (93). Solutions often end up taking the form of a stipulation.

(93) Ni xiangxin [Lisi mai-le shenme de shuofa] [Chinese]
You believe Lisa bought what DE claim
‘You believe the claim that Lisa bought what?’

It is this very fact, the contrast between adjunct and argument wh-in-situ, that I would like to address here.

The approach I will rely on is the one put forth by Watanabe (1992, 2001). Watanabe argues for a change of perspective on wh-in-situ of the type found in Chinese and Japanese. Instead of arguing that the actual wh-phrases move covertly, he essentially splits the wh-word in situ into two parts, an indefinite part (that remains in situ) and a wh-part, which he claims is a null operator. Only the latter, according to him, undergoes movement.

Phrased in these terms, Watanabe’s analysis mimics my proposal for resumption: a wh-phrase (empty in the case of wh-in-situ) is merged with an element in a theta-position (in the case of wh-in-situ, an indefinite pronoun).40

If we assume that the operator is adjoined to the indefinite, “resumptive” element, we can immediately capture the lack of island effects with wh-arguments in situ. Islands don’t emerge with these for the very same reason they don’t in cases of resumption.41 There

40 Huang (2003) comes close to making this claim by analyzing wh-in-situ as another reflex of a discontinuous element of the type he takes to characterize Chinese-style “analytic” languages. Interestingly, Demirdache (1991) already argued for a connection between wh-in-situ and resumption, by treating the resumptive pronoun as an in-situ operator moving at LF. I argue for the same connection, but in the opposite direction.

41 The oft-noted contrast between Chinese and Japanese in terms of wh-islands (absent in Chinese, present in Japanese) may follow from the fact that Chinese wh-in-situ phrases are morphologically bigger, hence, I assume, featurally richer than Japanese
is nothing special about islands and covert wh-movement (“wh-in-situ”). Interestingly, if it is true that resumption is not available to pure adjuncts, a resumptive analysis will not be available to (pure) adjunct wh-in-situ. We therefore expect island effects with these. Perhaps then, wh-adjuncts in situ are better viewed as lower copies being pronounced (perhaps for the same reason that wh-operators are null in these languages, which Cheng (1991) argues is tied to the existence of overt clause-typing complementizer-heads.)

The above remarks clearly do not do justice to the very rich literature on wh-in-situ, but they point to the fact that resumptive strategies may be useful outside the domain where they are typically appealed to.

As a matter of fact, a similar argument has been made in Wang (2007). Wang proposes an extension of the resumptive pattern argued for here (chain splitting) to characterize the lack of island effects in certain ellipsis contexts. I’d like to capitalize on Wang’s suggestion here, as the standard account of island repair under ellipsis does not fit well with the approach to locality argued for in this work.

5.8.4 Ellipsis

Ross (1969) originally noted the fact that that sluicing (taken to be IP-ellipsis) appears to rescue island violations.

For instance, not pronouncing the portion of the clause containing an illicit subextraction domain, yields an acceptable result. Compare the CED-violation in (94) and its sluicing version in (95).

ones. As Watanabe (2001) notes, whereas Chinese uses the same word shenme to express ‘what,’ ‘everything,’ and ‘something,’ Japanese dare ‘who’ must be morphologically “strengthened” to mean ‘someone’ (dareka) or ‘everyone’ (daremo). Perhaps we could understand the richer character of Chinese wh-in-situ phrases as capable of taking on a more referential, D-linked reading, which we know independently to facilitate extraction out of wh-islands (see Cinque 1990; Rizzi 1990; Starke 2001; Boeckx and Jeong 2004).

42 From this perspective, it may not be surprising to find that some languages, such as Malay, lack in-situ wh-adjuncts (Cole and Hermon 1998). Under the present account, it means that these languages lack empty operators that can stand alone—clearly a morpholexical parameter.
As Merchant (2001) and Lasnik (2001, 2005) have shown, the island repair potential of sluicing is quite systematic.43

Building on original suggestions in Ross (1969), Merchant (2001) and Lasnik (2001, 2003, 2005) have taken data like (94)–(95) to point to the fact that it is not the case that movement out of certain domains is inherently prohibited. Taking sluicing to be IP-ellipsis at PF (for which there is independent evidence; see Merchant 2001), data like (95) naturally lead to the claim that islands are not constraints on movement, but output filters at PF. Under this view, ellipsis deletes the bad outputs before they get ruled out. In other words, deviant outputs created in narrow syntax are not immediately filtered out. The claim that movement may take place out of “island” domains fits nicely with the approach developed here, as does the claim that islands are interface phenomena. But the claim that ellipsis, the mere fact of not pronouncing portions of the sentence, can remove bad outputs does not fit well with the overall approach developed here.

I have claimed throughout that locality conditions result from ambiguous instructions provided at the point when syntax maps onto the external systems. So the source of locality violations is to be found in the shape of syntactic instructions to PHON and SEM. On this view, locality violations are detected even prior to the determination of how elements are pronounced (as silent, stressed, etc.). So, ellipsis should not have any rescuing effect. In other words, the

43 Although sluicing appears to alleviate most island effects, Merchant does not claim that movement can take place out of any domain. In particular, Merchant distinguishes between cases where no island is found in the ellipsis site (this is how he analyzes the adjunct condition, which he subsumes under what he calls “propositional islands”) and cases where movement out of classic islands such as subjects in SpecIP is allowed, provided the island is part of the ellipsis site. However, Lasnik (2001, 2005) provides compelling evidence that sluicing can repair any island. I adopt Lasnik’s conclusion here.
repair potential of ellipsis comes as a surprise when presented as Ross, Merchant, and Lasnik have done.

Fortunately, Wang (2007) has proposed an account of the island-repair effect of ellipsis that can readily be accommodated here. Wang proposes that the reason islands are repaired under ellipsis is because the ellipsis site contains a resumptive pronoun that is licensed by the correlate of the presence of an indefinite DP in the antecedent clause. Wang’s proposal is illustrated in (96).

(96) John made the claim that Fido bit someone, but I didn’t hear who

Wang’s analysis shares with the standard Rossian analysis the idea that the wh-phrase moves out of the ellipsis site, but does not require the absence of island effects to be tied to the lack of pronunciation.

Wang observes that his account readily explains why in the absence of an indefinite correlate in the antecedent clause, sluicing fails to repair islands. (Chung, Ladusaw and McCloskey (1995) call such cases “sprouting.”) Under Wang’s analysis, the island effect is due to the fact that resumption is not available.

(97) *Agnes wondered how John managed to cook, but it’s not clear what food

The example in (98) shows that the presence of an indefinite correlate repairs the island violation in (97). (99) shows that sprouting is perfectly licit if the ellipsis site does not contain an island.

(98) Agnes wondered how John managed to cook a certain food, but it’s not clear what food

(99) John managed to cook, but it’s not clear what food

Wang notes that a resumption analysis immediately makes sense of the fact that most island repair cases of sluicing in the literature tend to have a D-linked wh-phrase as a remnant. As we saw above, wh-phrases that are resumed tend to be associated with a definite (resumptive) pronoun, and therefore tend to be more referential.
Wang’s analysis is very appealing, as it dissociates island-repair from lack of pronunciation. It allows us to unify repair by ellipsis and repair by resumption. But before endorsing it, we must examine the arguments provided in Merchant (2001) against a resumption-based account of island repair in ellipsis contexts.

Although Merchant leaves the door open for a resumption strategy in some sluicing cases, he claims that the “collective force” of the arguments we are about to discuss “put[s] a nail in the coffin of any hope that sluicing could be reduced to a resumptivity strategy in any sufficiently general way” (2001: 145).

Merchant provides four arguments against a generalized-resumption-based account. We will see that the arguments are not overwhelming, particularly under the analysis of resumption pursued here (and adopted by Wang 2007).

The first argument Merchant provides comes from case morphology. He notes, as we did above, that wh-antecedents of resumptive pronouns tend to not be case-marked. If they are case-marked, they tend to bear default case morphology, not the morphology we would expect them to bear if they were extracted like regular wh-phrases. The example in (100) illustrates this.

(100) Who/*whose did the police say that finding his car took all morning

Merchant in fact takes this lack-of-case generalization as evidence for a base-generation analysis of resumption. But, as I argued in Boeckx (2003a), this is too strong a conclusion. Absence of case

---

44 Wang’s analysis predicts that island repair under ellipsis will be unavailable to wh-adjunct remnants, since such adjuncts cannot be resumed. The prediction appears to be correct (see (i)), but, unfortunately, there is a confounding factor. Pure adjuncts like *why resist a long-distance reading under sluicing even in non-island contexts (as discussed in class lectures by Howard Lasnik at the LSA Summer Institute 2005).

(i) *He wants to interview someone who works at the soup kitchen for a certain reason, but he won’t reveal why [* under the long-distance construal of *why]*

(ii) *?John claimed that Mary left for some reason, but I don’t know exactly why

45 Wang examines Merchant’s arguments #1 and #4. His solution to argument #1 is not fleshed out enough, and appears to be different from mine. (Wang appears to suggest that since he adopts a movement analysis of resumption, case could be assigned to the wh-phrase prior to movement. But this goes against the motivation behind the chain-splitting account proposed in Boeckx 2003a and adopted here.)
morphology on the wh-phrase does not militate against a movement account. It simply militates against the claim that wh-antecedents of resumptive pronouns participate in case-checking. This is exactly what we argued in favor of above: Antecedents of resumptive pronouns do not partake in any checking relation prior to reaching their final landing site. Notice that all this entails is that wh-antecedents of resumptive pronouns do not check case. This still leaves the possibility of them bearing a matching case morphology acquired through non-syntactic means (default case morphology, for example, which I take to be assigned at PF). At any rate, lack of case on wh-antecedents constitutes the first part of Merchant’s argument against a resumptive strategy. The second part of his argument pertains to case morphology on wh-remnants in sluicing. One of the strongest arguments for treating the wh-remnant in sluicing contexts as having moved out of the ellipsis site comes from the fact that wh-remnants tend to bear the case morphology they would be assigned by their regular case-assigner.

 Merchant takes the contrast between (100) and (101) to be a strong argument against a resumption-based account of island repair under ellipsis. But this need not be so. We could, for example, say that the case morphology assigned to the wh-remnant at PF must be the same as the one borne by the relevant element in the antecedent clause for reasons of recovery. Since the element that would otherwise bear the case morphology (the resumptive pronoun) is null, recoverability of case information trumps default case morphology assignment (under something like the Elsewhere principle that demands that the most specific morphological exponent be selected whenever competition for vocabulary insertion arises; see Halle 1997). In other words, the presence of case morphology in (101) does not entail the absence of a resumptive pronoun in the ellipsis site. The contrast between (100) and (101) is the result of the fact that the resumptive pronoun is unable to bear case morphology in (101).
The second argument Merchant provides against a resumption-based account comes from the claim that some languages that have been argued to lack resumptive pronouns (Merchant (2001: 145) cites the case of West Flemish) nevertheless do allow sluicing (and island repair under sluicing). Again, the argument is not overwhelming. Although we do not fully understand why some languages make extensive use of resumption while other languages don’t, Boeckx (2003a) and McCloskey (2002) note that no language has a pronominal paradigm exclusively devoted to resumption. A resumptive pronoun tends to look like a regular pronoun. So, the claim that a language lacks resumptive pronouns is strange, since it comes close to meaning that such a language would lack pronouns. Furthermore, as Kennedy and Lidz (2001) have argued, ellipsis reveals the presence of elements for which a given language tends to lack morphological exponents. For example, Kennedy and Lidz argue that long-distance reflexives exist in English, but they are confined to ellipsis contexts because the language lacks the right morphology to spell them out. We could thus say that the resumptive use of some pronouns is confined to ellipsis contexts in some languages.

The third argument that Merchant provides against a resumption account is made in passing. Merchant notes that the wh-phrase in sluicing can have certain readings (e.g., a functional reading) that wh-antecedents of resumptive pronouns lack. This argument is not very strong either. The fact that wh-antecedents of resumptive pronouns lack certain readings may have to do with the type of resumptive pronouns they relate to. Though typically such pronouns are referential/definite, this need not be so. The pro posited in Wang’s account could (in some cases at least) be an indefinite, and therefore render some readings available that would not be available if the resumptive pronoun were truly definite. Wang’s account of island repair depends on there being a resumptive element, not on that element being like a definite pronoun. (Though, as Wang notes and as mentioned above, in many cases, it probably is a pronoun, given the D-linked nature of the wh-remnant.)

The fourth argument provided by Merchant against a resumption account comes from his Preposition-stranding generalization. The generalization is this: If a language lacks P-stranding in normal
contexts, it will also lack P-stranding in sluicing contexts. Thus, in languages such as English and Scandinavian languages, which all allow regular argument wh-phrases such as *who* to strand a preposition under wh-movement (the (b) sentences in examples (102)–(103)), we also find the possibility in sluicing of omitting a preposition that corresponds to a preposition marking the correlate of the wh-phrase in the antecedent to the deleted clause, as shown in the (a) sentences in (102)–(103). (Examples taken from Merchant 2001.)

(102)  English  
   a. Peter was talking with someone, but I don’t know (with) *who*  
   b. Who was he talking with?  

(103)  Swedish  
   a. Peter har talat med någon; jag vet inte (med) *vem*  
       Peter has talked with someone I know not with who  
   b. Vem har Peter talat med?  

   Norwegian  
   a. Per har snakket med noen, men jeg vet ikke (med) *hvem*  
       Per has talked with someone but I know not with who  
   b. Hvem har Per snakket med?  

In other Germanic languages, such as German, which generally do not allow preposition-stranding under wh-movement, retention of the preposition under sluicing is obligatory:

(104)  German  
   a. Anna hat mit jemandem gesprochen, aber ich weiß nicht, *mit* *wem*  
       Anna has with someone spoken but I know not  
       with who  
   b. *Wem hat sie mit gesprochen?  

Given that resumptive pronouns routinely “rescue” preposition-stranding violations (see Boeckx 2003a), Merchant’s P-stranding generalization comes as a surprise under Wang’s analysis. However, it should be pointed out that the generalization is not without problem. Though robust, it is subject to quite a bit of variation, in a way not expected by Merchant. Thus, as he notes
the ban against P-stranding is much less strong in sluicing contexts than in non-sluicing context in many languages. The very speakers who generally disallow P-stranding under wh-movement find it much more acceptable under sluicing. Furthermore, exceptions to the P-stranding generalizations have been reported in the literature (see Almeida and Yoshida (2007) for Portuguese, Szczegielniak (2005) for Polish, and Vicente (2006) for Spanish; Heidi Harley (p.c.) for Finnish). Although some of these exceptions may only be apparent, it is not clear that all of them are.

In addition, as Wang notes, the P-stranding ban is most robust with prepositions that are minimal, or bound morphemes. Preposition stranding is readily available under sluicing if a free/separable preposition is used. Witness the following examples from Hebrew (taken from Wang 2007).

\[
\begin{align*}
(105) \ a. & \text{ Dani katav le-mishehu, aval ani lo yode’a *(le-)mi [Hebrew]} \\
& \text{ Dani wrote to-someone, but I not know to-who} \\
& \text{ Dani katav le?} \\
& \text{ Who Dani wrote to} \\
& \text{ ‘Who did Dani write to?’} \\
& \text{ Yoshi diber al sefer mesuyam. Aval ani lo zoxer} \\
& \text{ Yoshi talks on book specific. But I not remember} \\
& \text{ (al) eize sefer} \\
& \text{ on which book} \\
& \text{ ‘Yoshi talked about a specific book, but I don’t remember which’}
\end{align*}
\]

It could be that those minimal, inseparable prepositions are case-markers, in which case the ban on preposition-standing under sluicing may be due to the sort of recoverability condition discussed with respect to Merchant’s first argument.

Be that as it may, it is clear that Merchant’s P-stranding argument is not foolproof, and therefore not as lethal as he claimed. I conclude from this discussion that Wang’s resumption analysis is tenable, and fits very nicely with the other instances of island repair discussed so far.

If true, Wang’s resumption analysis suggests that lack of island effects under sluicing should not be seen as an argument against a

\[46 \text{ Both Vicente and Szczegielniak independently argue that the contexts that allow P-stranding are instances of what Merchant calls “pseudo-sluicing” (cleft-structures).}\]
syntactic treatment of islands. Not pronouncing the extraction site, or, in the case of wh-in-situ, not pronouncing the moved operator, should not be regarded as the cause of the lack of island effect.\textsuperscript{47} Taking this conclusion together with my argument against a PF-treatment of resumptive pronoun, we can conclude that island effects (or lack thereof) are not directly related to matter of pronunciation (or lack thereof).

5.8.5 Construal

In this section I would like to make a few remarks on construal rules, which traditionally include relations that fall under the rubric of binding and control. Specifically, I would like to suggest that local binding and obligatory control fall under the rubric of resumption, required to avoid chains that would otherwise violate the QED-requirement imposed at the interfaces.

Since at least Chomsky (1973) it has been known that conditions on chains, specifically A-chains, mirror conditions on certain binding configurations (e.g., Condition A of the binding theory). The two types of dependency were unified by defining filler-gap relations as antecedent-trace(/dependent) relations, mirroring antecedent-referentially dependent element (e.g., anaphor)/binder-bindee relations, as represented in (106)–(107).

\begin{align*}
(106) & \text{John}_i \text{ got arrested } t_i \\
(107) & \text{John}_i \text{ likes himself}_i \\
\end{align*}

\textsuperscript{47} Fox and Lasnik’s (2003) study could also be interpreted as an argument against dissociating island repair and “silence.” Fox and Lasnik show that VP-ellipsis, unlike sluicing, fails to rescue sentences like (i).

(i) *A biography of one of the Marx brothers is going to be published next week, but I don’t know which she did \textsuperscript{(say that a biography of t is going to be published)}

Worse still, a sometimes legitimate output is rendered ill-formed by VP-ellipsis, as in (ii):

(ii) *They said they heard about a Balkan language, but I don’t know which Balkan language they did \textsuperscript{(say they heard about t)}

(cf. They said they heard about a Balkan language, but I don’t know which Balkan language they said they heard about t)

Fox and Lasnik conclude that the unacceptability of (i) and (ii) is the result of a violation of the parallelism condition on ellipsis applied to the chains formed by movement.
On grounds of Inclusiveness, Chomsky (1993) abandoned the idea that chains are antecedent-trace relations, and proposed instead that they are collections of identical elements (copies) (equivalently, series of occurrences). This led to a dissociation of the chain/binding connection until recent work reanalyzed binding relations in terms of movement. For instance, Hornstein (2001) suggests that anaphors (falling under the purview of Condition A) be regarded as morphologically altered copies, as in (108).

(108)  John likes $<$John$>$ $\rightarrow$ add $-$self$+$ turn John into him = himself

Such a view appears to receive rather strong support from languages where reflexives are morphologically identical to their antecedents, as in Thai, Hmong, and other languages (see Lee 2003; Boeckx, Hornstein, and Nunes 2007).

(109)  Pov yeej qhuas Pov [Hmong]
       Pao always praise Pao
       ‘Pao always praises himself’

In a similar spirit Kayne (2002) suggested we view pronominal binding as an instance of clitic doubling, as schematized in (110):

(110)  John said that [he [$$<$John$>$]] would be happy

I agree with these authors that it would be desirable to unify movement and binding, since they are subject to many similar requirements. In addition to the similarities noted by Chomsky (1973) and many others since, we could note that there is an intuitive connection between Binding Condition D (a less referential element cannot act as an antecedent for a more referential element; Lasnik 1989) and, say, Improper Movement (a more comprehensive, A-bar, domain $[T + \omega]$ cannot be “embedded” under a less comprehensive Case-domain $[T]$ in a chain). Likewise, there is an intuitive relation between some aspect of Condition C (excluding $*$John$_1$ likes John$_i$) and the ban on two maximal checking sites within a chain discussed in this study.

In a similar vein, it is tempting to relate control to movement on the basis of the fact that control environments match those
environments that are most transparent to movement (see Boeckx and Hornstein 2006). However, formulating these intuitions precisely, in terms of chains understood in the way developed in this book’s previous chapters, is not straightforward. Specifically, given the form of chains adopted in this book, binding and control dependencies cannot reduce to simple chains (see Rizzi 2006 for a similar observation). For example, if we were to treat anaphors as part of an A-chain (say, the lowest member of an A-chain), following the idea that Condition A (an anaphor must have an antecedent in domain D) has the same locality as A-chains, we would be faced with an illegitimate object, since such an anaphor-containing A-chain would contain two thematic sites, a QED-violation. For this reason I will not claim that binding relations and movement chains are identical at the level of narrow syntax. But this does not entail that we have to reject any link between construal and movement. In particular, the analysis of resumption adopted here allows us to get around QED-violations, without giving up the idea that antecedents relate to the pronouns they bind via movement. On the basis of this, I would like to suggest we treat anaphors and other bound elements as resumptive elements. This is not a completely novel idea. The intuition that bound elements are resumptives has been pursued by Kayne (2002), Zwart (2002), and Grohmann (2003).

Furthermore, certain facts about reflexivization appear to favor a split-chain approach to binding. As Zwart noted (building on work by Safir 1996), the morphology of reflexives cross-linguistically appears to fall consistently into two major classes: reflexives are either (semantically bleached) body-parts, or marked with focus morphology. Both morphological types fit well with a split chain account. When a reflexive is a body part, the relation to its antecedent could be seen as a predicative structure (as argued in Uriagereka 2002), pretty much like the one that obtains between a resumptive element and its antecedent, which I claim is merged adjacent to it, forming a big DP.

As for reflexives with focus morphology, it is a well-established fact that focus-marking is tightly connected to anti-agreement (see Ouali 2006; Simpson and Wu 2000; Watanabe 2004), which in turn...
is an important ingredient of the split-chain account of resumption adopted here. (Even cases of copy-reflexives of the type illustrated in (109) fit the general pattern once we take into account the fact that a common strategy of focus-marking cross-linguistically is in terms of doubling (reduplication). Accordingly, copy-reflexives need not be seen as pronounced “traces” of movement, but rather as resumptive elements “duplicated” to mark focus (an instance of “anti-agreement”).)

A resumptive analysis could also be posited to capture instances of local control (i.e., obligatory control), as suggested in Kayne (2002). In this case, the resumptive would be null (a null D° in Kayne’s 2002 analysis).

This, at first, appears to be a step backward in our understanding of control. As proponents of the movement approaches to control like Hornstein’s (1999) have stressed, a movement analysis immediately accounts for why “PRO” (the controlled element) is always phonetically null.48 Since traces/copies left by movement are not pronounced, it is no surprise to find that PRO is null, if PRO is actually a trace/copy left by movement. The null-ness of PRO is not as straightforward under a split-chain analysis like the one I am suggesting.

We could simply stipulate the null character of the resumptive in control structure. But maybe we can actually do a bit better. Perhaps, the null-ness of the resumptive element in control structures is related to the fact that it finds itself in environments that are morphologically least specified for phi-features (non-finite clauses). Perhaps some matching effect between the absence of φ-marking in non-finite domains and on PRO is at work. After all, proponents of a movement approach to control must stipulate that non-obligatory control PRO, which they assume to be pro, is null too. Perhaps both obligatory PRO and non-obligatory PRO are null for the same reason: they are found in non-finite environments.

48 At least in the typical cases, this is true of movement chains as well, where typically, the highest member of the chain is pronounced. For more complex cases such as backward control/raising and copy control/raising, see Polinsky and Potsdam (2006) and references therein.
But although obligatory PRO and non-obligatory PRO are morphologically alike, they have a different distribution and interpretation, which Hornstein’s movement account of obligatory control derives. For Hornstein, obligatory PRO is a trace/copy of movement, whereas non-obligatory PRO is like a resumptive pronoun, found in contexts where movement is typically thought to be impossible (traditional “islands”). How can we capture this difference if even obligatory PRO is treated as a resumptive element, as Kayne suggested?

A similar problem arises in the context of Kayne’s proposal to treat bound pronouns as resumptive elements separated from their antecedents via movement. If reflexives are treated in the same way, how can we capture the fact that reflexives must be bound in a much more local domain than bound pronouns?

I would like to suggest that the difference between obligatory PRO/reflexive and bound pronoun/non-obligatory PRO is simply the result of the fact that in the former, but not in the latter, the antecedent is remerged in a Theta/Case-domain that is directly connected to that occupied by the antecedent. Both the resumptive and the antecedent end up in the same extended projection. I suggest that this local connection enforces dependent reference interpretation on the reflexive. By contrast, bound pronouns and their antecedents wind up in distinct, unconnected Case-domains, which I take to mean that they need not share a referential index at SEM.

From the present perspective, the difference between a reflexive and a resumptive pronoun boils down to which checking site the antecedent is remerged into. Needless to say, the suggestions made in this section must be tested in many more domains of binding than I can afford to discuss in this chapter. My goal here is merely to sketch a way to reconcile the idea that “construal rules” involve movement with my overall claim that movement chains are subject to a QED-requirement.

I should point out that I expect that not all construal dependencies will turn out to be part of narrow syntax. Those dependencies that appear to be governed by the weakest kinds of locality conditions
(some Condition C effects, logophor licensing, licensing of what Giorgi 2007 aptly calls attitude-sensitive pronouns, also known as long-distance anaphors) seem to me more amenable to semantic/pragmatic treatments (see also Hornstein 2001, 2007, and Reuland 2001, 2006), and as such will not be directly affected by the proposals made in this book. But to the extent that construal rules mirror known syntactic phenomena, I think that a resumption analysis is worth entertaining.

5.8.6 Pied-piping

Let me now turn to the last important “island repair” phenomenon I want to discuss here, the phenomenon of pied-piping.

Like so many other island-related phenomena, the link between pied-piping and island obviation was established in Ross (1967), but thorough discussions of pied-piping, its range, its basic mechanism, and exactly how it obviates islandhood are relatively few. Perhaps the impression that pied-piping is heavily parametrized across languages has contributed to the relative lack of attention devoted to the phenomenon by generative grammarians. But as Heck (2004) shows, the impression of massive variation in this domain of grammar is, as has often turned out to be the case, illusory. Building upon previous findings, especially Webelhuth (1992), Heck has come up with five generalizations about pied-piping that appear to be constant across languages. I reproduce them in their schematic form here.

First, pied-piping can be recursive. Witness (111).

(111)  a. [[[Who]se friend] did you see?
   b. [[[Who]se friend]’s mother] did you see?
   c. [[[Who]se friend]’s mother]’s house] did you see?
   etc.

Second, pied-piping is accompanied by movement of the featurally relevant item to the edge of the pied-piped constituent. (Van Riemsdijk (1982) called this movement “internal” movement. Heck calls it “secondary” movement.)
Contrast (112) and (113).

(112)  [[who]-se brother] did you see?
(113)  ?*[the brother of who] did you see?

Third, pied-piping acts as a repair strategy in island contexts (the topic of this section).

(114)  a. *Whose did you see brother?
       b. [Whose brother] did you see?

Generalizations four and five, which I am about to mention, are, according to Heck, only true of what he calls massive pied-piping, i.e., pied-piping that occurs more freely than we would expect on the basis of the first three generalizations just listed. Heck claims that massive pied-piping is confined to non-embedded contexts (matrix clauses, parentheticals, appositive relatives, etc.). Generalizations four and five can be collapsed as follows: There cannot be any overtly filled specifier, or complementizer between the edge of the pied-piped constituent and the pied-piper (the featurally relevant element triggering pied-piping) even if the latter occurs in situ, inside the pied-piped constituent (contra generalization 2).

Relevant examples appear in (115)–(116).

(115)  a. The elegant parties, [to be admitted to one of which] was a privilege, had . . .
       b. *The elegant parties, [for us to be admitted to one of which] was a privilege, had . . .

(116)  a. John, to be proud of whom would be a mistake, . . .
       b. *John, to believe Mary to be proud of whom would be a mistake, . . .

Heck’s generalizations bring a lot of order and universality into the phenomenon of pied-piping. In the next paragraphs I’d like to suggest ways of understanding them in the context of the theory of locality advocated here.

First, we can explain away generalization #1 (recursive pied-piping) by distinguishing between two causes of pied-piping. The first is syntactic, and has to do with island obviation (generalization #3). The second is morphological, and has to do with certain adjacency
conditions such as the stray affix filter, which, once combined with phrase structure (constituency), will conspire to require pied-piping. For example, taking possessive ’s to be an affixal D immediately accounts for why who must pied-pipe ’s in whose book (more transparently, who’s book) did you read. It is clear that every occurrence of the relevant affix will entail recursive pied-piping. Pied-piping here need not be triggered. Wrong choices of constituent fronting will be filtered out at the interfaces.

As for the difference between regular, “well-behaved” pied-piping and “massive” pied-piping, I would like to claim that it is reducible to another factor, viz. Ross’s (1973) Penthouse principle, which states that “more goes on upstairs than downstairs.” Ross noted (see also Emonds 1976) that non-embedded domains display signs of reordering to a larger extent than embedded domains do. I assume that this greater freedom of reordering increases the likelihood of internal movement of the pied-piper.

It is worth noting that, as de Vries (2006) has shown, in languages with richer word-order permutation possibilities than English, such as Dutch (which has scrambling), massive pied-piping is available in all contexts. (Webelhuth 1992 already suggested that massive pied-piping may be related to the fact that topicalization is disfavored in embedded contexts in English.)

The remarks just made reduce the generalizations to account for to the following: (i) necessity of internal movement, (ii) repair in island contexts, and (iii) ban on specifier/complementizer internal to the pied-piped unit.

The repair effect of pied-piping can be understood as a way narrow syntax has found to avoid forming a chain that would violate QED. Like resumption (chain-splitting), pied-piping (chain takeover) frees up a certain element from too much checking. Pied-piping

---

49 As Drury (1999) shows, slight (dialectal) variations concerning the morphophonological nature of ’s will lead to absence of pied-piping. Gavruseva and Thornton (2001) document similar variations in the process of acquisition, when children are still fine-tuning the morphophonological properties of the lexical items they are acquiring.

50 I tentatively assume that the Penthouse principle is not a genuine principle of narrow syntax, but rather a filter imposed by parsing considerations.
is in some sense the mirror image of resumption.\textsuperscript{51} Instead of allowing the target of movement to abandon the domain where it would be trapped in a checking site via resumption, the target of movement takes the relevant trapping domain with it. But it does so at a certain cost, or under a certain condition.

In order for pied-piping to take place, the whole unit has to count as an appropriate checker. Syntax only offers one way to do this: feature-sharing under valuation. I would like to argue that this feature-sharing requirement is what drives internal movement of the “pied-piper.”\textsuperscript{52,53}

Since the entire projection hosting the pied-piper must form a chain of the relevant type, we can understand the no-filled specifier requirement notes by Heck as a reflex of the fact that filled specifiers

\textsuperscript{51} It is interesting to note that the domains that Heck associates with massive pied-piping correspond point by point to the contexts of (intrusive) resumption (resumption typically called for in contexts that cause processing difficulty) given in Bianchi (2004).

It is also worth pointing out that both resumption and pied-piping are often found with objects of prepositions (see Boeckx 2003\textit{a} for data on resumption, and de Vries 2006 and Webelhuth 1992 for data on pied-piping).

\textsuperscript{52} I assume that those cases where the pied-piper is in-situ are instances of hidden wh-movement (lower copy pronunciation). As Kato (2004) argues, English manifests non-echo wh-in-situ in a few contexts, which match Heck’s contexts of massive pied-piping fairly closely. It is also interesting to note that many of Heck’s instances with in-situ pied-pipers have been treated as echo-questions in the literature (see De Vries 2006), which is also the case for the examples discussed in Kato (2004). I conclude from this parallelism that instances of in-situ pied-pipers are only apparent, and claim that all cases of pied-piping require internal movement ( overtly or, in a few cases, “covertly”).

\textsuperscript{53} A similar phenomenon of internal movement obtains in the domain of parasitic gaps, a domain where “chain union” has been argued to take place (Chomsky 1986\textit{b}; Nissenbaum 2000). As was noted by Kayne (1984), no island can intervene between the parasitic gap and the head of the domain inside which the parasitic gap resides. Contrast (i)--(ii).

(i) What did you read before filing _____
(ii) *What did you read before meeting the man who filed _____

To capture this restriction on parasitic gaps, Chomsky (1986\textit{b}) argues (see Nissenbaum 2000 for renewed arguments in favor of this approach) that parasitic gaps are formed by combining two chains: a licit chain (accounting for the normal) gap, and a chain internal to the island (domain of the parasitic gap) that parallels the licit chain. From the perspective of the present work, we can think of parasitic gaps as chains formed in parallel with licit chains, hence sharing an interpretation (“licensed”) at SEM.
internal to the relevant projection would interfere with the feature-sharing achieved by internal movement of the pied-piper. The idea here is that a pied-piper clause must be as neutral as possible featurally speaking, prior to internal movement of the pied-piper, so as to be ready to assume the feature value of the pied-piper.

The no-complementizer requirement imposed internal to the pied-piped domain can be understood in the same way. The presence of a complementizer would provide a type to the pied-piped clause that may clash with the one that must be given by the pied-piper for pied-piping to be successful.

If the suggestions just made are on the right track, pied-piping can be seen as another well-behaved strategy conjured up by the resources available to narrow syntax to meet interface requirements in an optimal way (QED).

Taken collectively, island repair strategies strengthen the approach to locality proposed in this work. All island repair strategies appear to be confined to situations where chains would yield ambiguous results. We can thus think of island repair strategies as modes of disambiguation, much like cartographies were required in the context of extended projections, or, in fact, much like Agree was used to unambiguously identify the head of the Merge product. Agree, extended projections, and multiple chain formations are means that syntax resorts to to create uniformly unambiguous syntactic objects.

5.9 Final considerations

The present chapter has extended to the realm of chains the very same approach that provided an explanation for why products of Merge are labeled, and why cartographies exist. I have claimed that the QED-requirement imposed at the interfaces prevents chains from extending beyond the first checking site reached by the moving element. All islands are therefore relativized to the amount of checking relations established and their configurations. In other words, no island is absolute, although characteristic patterns of
checking led me to claim that adjoined structures as well as displaced constituents have a freezing effect that elements remaining within the thematic layer lack.

The present approach provides the basis for a unified approach to otherwise quite disparate locality constraints on chains. But it would be presumptuous of me to claim that the correctness of this unification has been demonstrated in the preceding pages. Such a demonstration requires detailed discussion of more facts than I have been able to cover here. But I hope to have shown how treating chains as projections, subject to the same ban on ambiguity as standard phrases are, captures well-known generalizations such as Huang’s CED and the that-trace effect, in addition to tying together various island “repair” strategies.

The present approach converges with the format of chains put forth in Rizzi (2006). Rizzi claims that a chain starts in an s-selectional position, and ends in a criterial position. Though adequate, Rizzi’s approach fails to explain why the s-selectional position, which is standardly phrased in terms of a criterion (such as Chomsky’s 1981 theta-criterion), does not lead to freezing. In other words, Rizzi’s approach captures the maximal size of chains, but fails to explain why some criterial positions lead to freezing, while some others don’t. The approach developed in this chapter can be seen as an extended answer to this question. For me, a chain, like any other syntactic object, can contain more than one edge, but no more than two, due to the QED-requirement.

The present approach also agrees with Boeckx’s (2003a), and Richards’s (2001) claims that chains cannot contain too many “strong” features, although I have argued that viewing this as a constraint on linearization (pronunciation instructions) is too narrow a characterization. Here the ban on ambiguous chains holds for both the mapping to PHON and the mapping to SEM.

The present approach is close in spirit to early approaches to locality like Kayne (1984), which crucially involves notions like (government-)projection, path, and “connectedness” to characterize licit displacement. The present approach is also reminiscent of
Koster (1987), whose core idea of the Configurational Matrix emphasizes the representational character of locality conditions.

The current approach is also close in spirit to models of Move $\alpha$ (Lasnik and Saito 1984, 1992), since the explanatory burden is left to a constraint (here, the QED) that acts as a filter on syntactic objects. The main difference is in the way this filter is construed, which for me involves a notion of efficient mapping. In this sense, I agree with Chomsky’s long-standing claim that islands reflect computational limitations, but, for me, such limitations are not inherent to the central component of narrow syntax (Merge, be it External or Internal), but emerge due to the fact that syntax must be mapped onto external representations of the sort I discussed in Chapter 3. In other words, locality effects could be seen as arising from the QED requirement imposed at the interfaces to ensure efficient “parseability” of syntactic objects.\[54\]

It is worth noting that the specific implementation of efficient computation advocated here is very different from the one traditionally pursued in the generative literature, beginning with Chomsky’s (1973) subjacency, and culminating in his phase-based approach (Chomsky 2000 and subsequent work). For Chomsky, computational efficiency is understood in terms of chunking. The derivation should carry along as little information as possible in its working memory. As a result, elements are forced to vacate domains, and move up the tree if they want to remain accessible to further computation. (This is the essence of Chomsky’s (2000, 2001) Phase Impenetrability Condition). Although the traditional view has been extremely useful, it leads to the somewhat paradoxical view that (i) movement is required for further movement (think of the traditional characterization of successive cyclicity), but (ii) movement tends to block further movement (Rizzi’s criterial freezing, CED-effects, etc.).

\[54\] The approach developed here is thus very congenial to the one advocated in Berwick and Weinberg (1984), although I stress that “locality” as meant here covers more than islands, and extends all the way down to labeling. Thus, although parsing efficiency plays a key role here, I do not claim that islands are performance effects of the sort advocated by Givon, Kluender, Pritchett, and others. For arguments against the latter, see Lasnik (1999b) and Phillips (2006).
By phrasing computational efficiency in terms of the QED-requirement, I hope to have shown that the paradox just mentioned ceases to exist.

Perhaps the biggest challenge faced by the present approach is the apparent lack of uniformity in the data. The clearest manifestation of non-uniformity in the data is the degree to which locality violations vary in unacceptability. But non-uniformity does not stop there. Some chains are known to be subject to both strong and weak islands, others to weak islands only (see Cinque 1990; Rizzi 1990; Szabolcsi 2006a). Some processes (wh-movement, e.g.) are subject to superiority, others (left-dislocation, e.g.) aren’t. Rightward movement is more constrained than leftward movement (Ross’s (1967) Right-roof constraint). Unembedded clauses allow a wider range of transformations than embedded clauses (Ross’s (1973) Penthouse principle), and so on.

From the present perspective, such lack of uniformity must be illusory. The illusion of non-uniformity could be the result of various factors: either some variable was not controlled for, or the wrong experimental setting was used, or the examples constructed did not constitute a minimal pair, etc.

The hope is that once all these factors are controlled for, uniformity will be manifest. Although the “uniformizing” task is vast, there are grounds for optimism. For instance, it has been pointed out (Stepanov 2001; Snyder 2000) that results of satiation tests (improvement upon repeated exposure; see Snyder 2000, Hiramatsu 2000) militate against a uniform treatment of the CED. However, Sprouse (2007) shows that, once carefully constructed, such tests no longer militate against a uniform treatment.

I should point out that it is likely that grammatical considerations alone will prove insufficient to account for the full range of acceptability facts. This is not specific to locality, and should come as no surprise once the performance-competence distinction is taken into account. Grammar is but one of the dimensions that enters into the act of judging the acceptability of sentences. Until we know more about all the other factors involved (processing, pragmatics, ease of recovery, etc.), and their respective contributions, it is not clear that a
theory of narrow syntax should account for all the variation we observe, especially if we have theory-internal reasons to believe that the variation has a non-syntactic cause.

Take, for example, the variability found in the domain of intervention effects of the sort discussed in Beck (1996).

It appears to be the case that certain quantified expressions separating a wh-expression in situ from its scope position leads to degradation. (More specifically, as Pesetsky (2000) has discussed at length, the presence of an “intervening” quantified expression appears to force a certain (single-pair) interpretation of the sentence.) Relevant examples, from French, appear in (117).

(117)  a. ??Tous les étudiants ont rencontré qui [French]
      All the students have met who
      ‘Who did all the students meet?’
  
b. ??Chaque étudiant a rencontré qui
      Each student has met who
      ‘Who did each student meet?’
  
c. ??Il admire toujours qui
      He admires always who
      ‘Who does he always admire?’

Intervention effects of this sort are known to be subject to great cross-linguistic and inter-speaker variation (see Szabolcsi 2006b), and have resisted precise characterization in terms of minimality effects.

However, Grohmann (1998, 2006) argues that it would be a mistake to question minimality-based accounts of intervention effects in general. As far as intervention effects of the sort discussed by Beck (1996) are concerned, they most likely follow from semantic-pragmatic considerations (hence the fact that they restrict the range of appropriate answers, as opposed to leaving to absolute unacceptability).55 Specifically, Grohmann argues that the relevant interveners are those that resist a topic interpretation, which is forced upon them in the relevant cases. (For converging conclusions, but different implementations, see Tomioka forthcoming, 55 For valuable discussion of pragmatic factors influencing the status of weak island violations (including the well-known “adjunct/argument asymmetry”), see Portolan (2002), Starke (2001), Frampton (1999), Postal (1998), and Szabolcsi (2006a).
Beck 2006, and Vergnaud and Zubizarreta 2006.) As Grohmann suggests, topic interpretation depends in part on discourse context and lexical factors that influence judgments and are thus the source of the cross-linguistic variation we find in this domain. So Grohmann’s conclusion is that our understanding of minimality need not be modified (i.e., complicated) to accommodate intervention effects like (117).

In a similar vein, Kato (2006) has shown on the basis of extremely detailed discussion of Japanese data that the Coordinate Structure Constraint\(^{56}\) boils down to a parallelism requirement at the level of Logical Form (as originally proposed by Munn 1993, see also Fox 2000), and should not be seen as a constraint on narrow syntax.

Kato shows that treating the Coordinate Structure Constraint as an interface condition straightforwardly captures what would otherwise constitute syntactic violations of it, while accommodating equally well cases of non-movement relations that are constrained by the Coordinate Structure Constraint. Kato’s argument is a significant contribution to a uniform theory of syntactic locality. No syntactic theory of islands I know of (certainly not the one developed here) naturally accommodates the Coordinate Structure Constraint. If Kato is right, syntactic theory need not be enriched to capture its effect.

In addition to semantic and pragmatic factors, it is likely that processing factors also give rise to non-uniform patterns of (un)acceptability.\(^{57}\)

In other situations, syntactic manipulations or refinements eliminate apparent differences among locality conditions. For example, Kayne’s (1994) Anti-symmetry hypothesis, even if it is not accepted in toto, certainly eliminates the need for an asymmetric (hence, theoretically odd) condition like Ross’s (1967) Right-roof constraint. By claiming that rightward movement is actually

---

\(^{56}\) More precisely, Kato deals with instances of extraction out of conjuncts. For cases of extraction of conjuncts, Kato suggests that a PF-condition (another interface condition) is needed (see also Merchant 2001).

\(^{57}\) Ross’s Penthouse principle may be a case in point. For remarks concerning the processing factors giving rise to “the privilege of the root clause,” see Rizzi (2007).
leftward movement followed by remnant movement, the boundedness of rightward movement processes follows from independent conditions on remnant movement (see Mueller 1998; Abels forthcoming).

Similarly, the stricter locality of head-movement (see Travis 1984) when compared to other kinds of movement is no longer problematic if head-movement is re-analyzed as the morphological reflex of head-to-head Merge (made possible under Bare Phrase Structure; see Harley 2004; Boeckx 2006a; Brody 2000; Boeckx and Stjepanović 2001; Chomsky 2001).

Likewise, a better understanding of the properties of the features driving movement has led to the trivialization or elimination of syntactic differences between control and raising (see Boeckx and Hornstein 2004), or between types of A-bar dependencies (see Bošković 1999, 2002b, and Boeckx 2003b on superiority and lack thereof with wh-movement, topic movement, and focus movement).

Finally, in some cases, non-uniformity disappears once better examples are constructed.58 Take, for example, the fact that super-raising examples (minimality violations under A-movement) are typically judged more deviant than wh-island cases (minimality violations under A-bar movement).

The examples typically used to illustrate the contrast are given in (118)–(119).

(118) **John is likely that it was told t that Mary left
(119) *What did John ask whether Bill cooked t

Though real, the asymmetry between (118)–(119) should not lead us to question the idea advocated in this chapter that A- and A-bar chains are subject to the same locality conditions.59 If we assume

58 Lasnik (2005) makes a similar point regarding Merchant’s (2001) non-uniform treatment of island repair under sluicing.

59 My claim that A- and A-bar chains are subject to the same locality does not mean that there is no difference between A- and A-bar chains. For one thing, they involve different features, which may lead to different interpretations at SEM (e.g., operator-variable structures in one case, but not in the other). This may be the source of well-known discrepancies between A- and A-bar chains regarding Weak Cross-Over, parasitic gap licensing, etc.
that that-clauses require Case (contra Stowell 1981, but see Bošković 1995, and Kiguchi 2003), we can capture the contrast by noting that movement of John in (118) crosses not only the embedded A-subject (it), but also the head of the clause (that), which competes for the subject position (cf. [That John was told that Mary left] is likely). No such double crossing arises in (119). As a matter of fact, when double crossing does arise in the context of wh-dependencies, the result is sharply deviant. Consider cases like (120), discussed by Manzini (1992) and Starke (2001).

120 *Which books do you wonder to which students to ask whether to give t t

In sum, when various factors are controlled for, the desired uniformity resurfaces. Although I find this result very encouraging, only further testing will reveal whether the degree of idealization that the present study calls for is warranted.

60 Alternatively, the (null) pronominal correlate of the that-clause, if Koster’s (1978a) analysis of clausal subjects is adopted.
This page intentionally left blank
Part III
This page intentionally left blank
Kayne (1984) begins his seminal paper on “binary branching” (ch. 7, “Unambiguous Paths”) by asking a why-question:

Our point of departure shall be the question: Why should there exist a c-command requirement? That is, why should the first branching node dominating the antecedent have to dominate the anaphor, too? We shall consider that this question calls for an answer. (p. 129)

Kayne’s solution is worth quoting in full:

As a first step toward an answer to this question, let us consider a certain manipulation of the trees of [1] and [2] that suggests the existence of a relationship between the antecedent-anaphor relation and the standard dominance relation. We imagine a horizontal axis drawn through D in both [1] and [2], and a reflection around that axis of the part of [1] and [2] to the left of D. The result is [3] and [4]:

![Diagram](attachment:tree-diagram.png)
In [4], A comes to dominate C, but in [3], A does not dominate C in the usual manner. Put another way, [2] is mapped by this reflection into a phrase structure tree of normal appearance, while [1] is not.

We interpret this to mean that the relation between A and C in [2] (the licit antecedent-anaphor configuration) is “close to” the standard dominance relation, whereas that between A and C in [1] (the illicit antecedent-anaphor configuration) is not. (p. 130)

Why-questions of the type Kayne asked in 1984 have become more prominent in linguistics in the context of the development of the minimalist program. (In many ways, Kayne’s why-question about c-command—though not his answer—anticipates Epstein (1999).)

To answer the specific why-question asked in this book—Why should there be such locality conditions in syntax?—I have, in some sense, extended the logic of Kayne’s answer. Like him I have made use of symmetry considerations, and argued that locality violations amount to departures from “trees of normal appearance.” Like Kayne I have pushed the idea that ambiguity is something that the language faculty seeks to avoid.

Crucial for me were (i) the symmetry of Merge, which allowed me to extend conditions on phrase formation (projection) to chain formation (“movement”), and (ii) a Quick Edge Detection (QED) requirement imposed at the interfaces (by both SEM and PHON). I have argued that narrow syntax meets the QED-requirement by providing unique labels for each application of Merge, that for the same reason syntax is forced to form cartographies/extended projections to accommodate multiple relations involving a single element, and, finally, that for the same reason chains are severely bounded, terminating as soon as they extend to reach the first checking domain.

In short, labeling, cartographies, and bounding nodes are all names for the solutions that the language faculty has developed to ensure that syntactic objects be unambiguous, in a way that makes the mapping to SEM and PHON maximally efficient. The symmetry of Merge, in all its guises (Internal/External Merge-equivalence; equivalence of status of Merge partners), is freely exploited by
narrow syntax, but its results are evaluated within the confines of efficiency considerations.

The net result is a much more unified syntactic engine, whose outputs vary in size, but not in shape. If correct, the long argument presented in this work can be put forth as an argument in favor of the simplicity of grammar. As Chomsky (1951) already observed, if we are to reveal the simplicity of a system, one ought to pay special attention to generalizations across “modules.” Witness the following passage.

For the formulation of any relative precise notion of simplicity, it is necessary that the general structure of the grammar be more or less fixed, as well as the notation by means of which it is constructed. We want the notion of simplicity to be broad enough to comprehend all those aspects of simplicity which enter into consideration when linguistic elements are set up. Thus we want the reduction of the number of elements and statements, any generalizations, and, to generalize the notion of generalization itself, any similarity in the form of non-identical statements, to increase the total simplicity of the grammar. (p. 5; emphasis mine)

Revealing the similarity behind the disparity of labeling, cartographies, and islands has been the major motif of this work.

The simplicity revealed is one of a specific kind, one that shows that the richness of syntactic objects that has been documented until now in the literature is the result of the interplay of simple processes and constraints—essentially, it is the interplay of just one process, Merge, formulated in the most unrestricted, symmetric fashion, and a ban on symmetric products; a subtle tension between symmetry (of rule) and asymmetry (of object).

Although the asymmetry of syntactic objects has long been documented (see especially Kayne 1994), I have tried to show in the preceding chapters that asymmetries need not be multiplied to characterize the form of phrases, chains, and domains. If I am right, it is the same asymmetry that pervades these various modules of grammar, because it is the same process that is responsible for them in the first place. The asymmetries we find are all aspects of a (spontaneously) broken symmetry.
Symmetry is a term that has figured prominently in the previous chapters because it has long been recognized to be the scientist’s favorite tool to not only unify (symmetry does that by definition: it shows that two objects are really one and the same), but also to understand aspects of the world. Symmetry is known to be able to not only systematize (bring together), but also rationalize a vast amount of results. What is symmetric is invariably characterized as elegant, beautiful, perfect, and so on. It does not require further explanation. It is departures from symmetry that require explanation. As Pierre Curie once noted, asymmetries create phenomena (prior to asymmetry, there is nothing to explain). Saying that Merge applies symmetrically follows as a matter of course. It’s only when the product of Merge is asymmetric that one begins to ask why the asymmetry went this way, and not that way.

In accordance with minimalist guidelines, I have suggested we understand the asymmetry of syntax as a departure from the symmetry of Merge dictated by efficiency (/economy) considerations. The intuition has been that symmetric products of Merge would render the mapping to SEM and PHON more difficult. By being symmetric, the products of Merge understood as instructions to the external systems do not specify a unique place to start—all options are equally valid. Computations would thus be too time-consuming. Labeled phrases, bounded chains, and cartographies are the price to pay to avoid the computational burden that would otherwise be too severe at the mapping stage. The reason so few chain types succeed is because so few unambiguous structure types can be built.  

Ambiguity reduction is not a new concern I have developed here. It is in fact presupposed in Chomsky’s earliest writings. For example, Chomsky (1955: 309) simply states that “[t]he first requirement of

---

1 I stress that the kind of symmetry I have emphasized here is a symmetry of rules, not a symmetry of objects, which is often how symmetry is used in the linguistic literature (for example, in the context of “parallelism” (see Fox 2000; Kato 2006), or “shape conservation” (see N. Richards 1999; M. Richards 2004; Williams 2003; Fox and Pesetsky 2005; among others)).

2 This formulation echoes Hale and Keyser’s (1993) answer to the question why so few theta-roles exist.
grammatical transformations is that the result of a transformation must be unambiguous. As I noted in previous chapters, ambiguity reduction is also a concern in Chomsky’s first attempts to characterize islands (Chomsky 1964, 1973). It also figures prominently in Kayne’s (1984) influential study, and is also mentioned on the first page of Rizzi’s (1990) work on relativized minimality. But I would argue that ambiguity reduction acquires a special status in a minimalist context, where it is made part of legibility conditions. Ambiguity reduction could in fact be regarded as a “third factor” principle of good design (in the sense of Chomsky 2005) that acquires special significance when fine-tuned to the context in which it is expressed (in this case, in the context of the mapping from syntax to PHON/SEM). Indeed I have argued that it is because of the very texture of our language that ambiguity reduction is required. If syntax did not have to reduce the dimensionality of the products of Merge to map them appropriately to PHON and SEM representations, perhaps labeling, cartographies, and bounding nodes would not be called for. It is the very texture of our language faculty that makes Merge symmetric, and causes its symmetry to break.

---

3 Thanks to Howard Lasnik for helping me find this quote.
4 Uniqueness considerations also play an important role in the formulation of “criteria” (see Chomsky (1981) on the theta-criterion, May (1985) and Rizzi (1996) on the wh-criterion, etc.), and in various characterization of licensing principles such as Koopman and Sportiche’s (1982) Bijection principle.
5 On “third factor,” see Boeckx (2006a: ch. 4), and, avant la lettre, Uriagereka (1998).
6 Chomsky’s (2004) claim that once you get External Merge, it would take an extra condition to exclude Internal Merge must therefore be understood in the context of linguistic Merge. An operation that combines elements into sets (call this “generic” merge, concatenate, group, etc.) will only give rise to Internal Merge if it can operate on a plane (2D-representations), the same way that complex numbers immediately arise in mathematics only when a plane (the “complex” plane) is formed by the intersection of two (number) lines (for illuminating discussion, see Derbyshire 2006: 12). For me, Internal Merge follows from External Merge not only because Merge is source-independent (as in Chomsky 2004), but also because chains can be formed once phrases can be formed. In other words, though broken upon merge, the symmetry of Merge manifests itself in subtle form in the parallel formation of projections and chains.
I have argued in Chapter 3 that narrow syntax consistently relies on lexical specifications to break the symmetry of Merge. At the initial state, upon lexical selection, Merge members share features, but some (the “Probes”) have features that lack values. A mechanism of Agree is required to remedy this deficiency lest syntactic objects violate the principle of Full Interpretation (which, I assume, demands, among other things, that elements mapped onto SEM and PHON be fully specified).

I have claimed that syntax takes its lead in the formation of asymmetric objects (e.g., uniquely labeled sets) from the asymmetry that arises in the context of featural specification. It is as if, once again, syntax economizes in this domain: having to build asymmetric representations, it seeks to minimize the sources of asymmetry, and resorts to a pre-existing asymmetry (the Probe-Goal asymmetry). We could see in this a sign of good design. We can certainly see in this a reflex of Inclusiveness: syntax merely rearranges terms; asymmetric products are built from lexical resources. In this sense, the character of our language faculty, its emergence in the species, crucially depends on the way conceptual information is lexically encoded. The formation of a digitized lexicon must therefore be regarded as a momentous event, which, once combined with unrestricted Merge, made it possible to map syntactic objects onto SEM and PHON. Syntax can be said to be bare, as the whole of syntax can be said to be the result of initial featural specifications drawn from the lexicon, symmetric Merge, and bare output conditions requiring efficient mapping.\(^7\)

Islands, labels, and extended domains of projection emerge as “bottleneck” effects, the results of the fact that the patterns of subordination must be turned into uniform patterns of coordination at the interfaces.\(^8\)

---

\(^7\) Being bare, syntax, as understood here, should not have required too many evolutionary stages—a good thing, given the short window of evolutionary time usually agreed upon for the emergence of the language faculty (see Boeckx (forthcoming) for relevant discussion and references).

\(^8\) If correct, the present line of inquiry casts doubt on frameworks like Jackendoff (1997, 2002, 2007), for whom the relations among the representations in syntax, at SEM, and at PHON, are much less constrained.
Because the products of Merge must be kept maximally asymmetric (easy to process), I have suggested that asymmetries involving a single element must be spread out over the course of the derivation. Accordingly, I have suggested that each product of Merge gets mapped onto the interfaces (see also Epstein and Seely 2002, 2006). Spell-Out is, then, very cyclic. In other words, each instance of Merge could be said to be a phase, so long as phase is understood as a mapping unit.

It has often been observed that treating each instance of Merge (the most minimal unit that syntax can form) as a phase is arguably the most natural definition of what counts as phase (see Boeckx 2007a; Epstein 2007; Bošković 2007). But Chomsky is right to point out (see, e.g., Chomsky 2007: 17; see also Ott forthcoming) that taking each Merge product to be phasal would not have the properties he wants to ascribe to phases, specifically in the context of locality.9

The reason we can entertain the idea that each product of Merge is a phase in the present framework is because I have dissociated phases from their alleged consequences for locality.

A major motivation behind the notion of phase in Chomsky’s system is the idea that phases reduce computational complexity. Islands emerge as syntax tried to forget (and map onto PHON and SEM) a maximal amount of information—minimizing what it must retain in working memory. Locality is thus a matter of derivational dynamics (see also Frank 2002, 2006).

For me, locality has a much more representational character. It has to do with the format of the products of Merge. Put differently, locality arises from the way syntax minimizes the burden on the working memory of the systems it interfaces with. It is interesting to note that this view on locality leads to a much more local window of opportunity (essentially, the phrase) than in a Chomsky/Frank-style system, for which the domain of locality (e.g., for successive cyclicity) always spans more than one phrase.

In this context of framework comparison, I should note that because so much in the previous chapters depends on lexical

---

9 For Chomsky, if each Merge product would be phasal, all instances of extraction would be ruled out, as each complement of the phase head would be frozen in place.
specifications and the way lexical information percolates up the tree node by adjacent note, the architecture I am advocating bears an obvious resemblance to HPSG-models (see, e.g., Pollard and Sag 1994; Sag 2005; Levine and Hukari 2006), as well various Categorial Grammar models (see, e.g., Steedman 1996, 2000), which have consistently relied on Gazdar’s (1981) insights in this domain. Because the present work emphasizes the need for clear, unambiguous instructions to the interfaces, it also enjoys a certain degree of affinity with LFG-treatments of island effects as resulting from “functional uncertainty” (as in Kaplan and Zaenen 1989; Bresnan 1995, 2001; Falk 2001). At a more basic level, by implicating feature-sharing (Agree) in structure-building, the present work incorporates some aspects of unification-based models of grammar (see Shieber 1986).

I find such relationships with other frameworks pleasing, as it reveals that by searching for unity within grammar (across modules), one also finds unity across models of grammar, which is exactly what one would expect if progress is being made, and investigations are on the right track.

Put in very general terms, this book has sought to achieve a simplification of concepts and objects by generalizing, or extending the range of applicability of the well-known X-bar schema (suitably derived from interface considerations). My hope in closing this long argument is that the hypothesis put forth here should be of immediate relevance to virtually all linguistic frameworks, given the centrality (and general adoption) of X-bar-like structures. Since it makes use of such a central concept, the present hypothesis ought to be testable from a variety of perspectives. Whether such tests prove the hypothesis right or wrong, only time will tell.

10 In some sense, the Bare Phrase Structure approach advocated here is lexically constrained, much like the original X-bar theory was. The difference is that I have been able to avoid having to posit bar levels as lexical formatives. Features that have been independently involved in Agree (Case, φ) have proven sufficient.

11 Much like Chomsky pointed out that you can’t avoid having Internal Merge if you have External Merge, there is a sense in which, if the symmetry argument made here is correct, once X-bar structures are adopted, locality in long-distance dependencies are expected to follow. Since Merge contains the seeds of islandhood, the present work could be said to embrace Postal’s (1997) admonition to take domains to be islands by default.
References

*Linguistic Variation Yearbook*, 7.
References


— In progress. Linguistic unity and language diversity—A minimalist program for comparative syntax. Ms., Harvard University.


— —— —— In prep. Control as movement—One long minimalist argument. Ms. Harvard University, University of Maryland, and University of São Paulo.
References


Butler, J. Forthcoming. What semantic concerns can tell us about syntactic structure (or, towards deriving something a little bit like a Cinque hierarchy). *Linguistic Variation Yearbook*. 
References

References


References 261


—— 2003. On the syntax of locative and directional adpositional phrases. Ms., CUNY.


References


References

References 265

Hiramatsu, K. 2000. Multiple Agree. Ms., MIT.


References


—— 2005. The projection of DP. Ms., SUNY, Stony Brook.


References

References


—— 2000. Words. Ms., MIT.

—— 2006. Phases and words. Ms., NYU.


References


References

References


References

References


—— 2005. Person-Case effects in Tagalog and the nature of long-distance extraction. Ms., MIT.

—— 2006. A distinctness condition on linearization. Ms., MIT.


Snyder, W. 1992. Wh-extraction and the lexical representation of verbs. Ms., MIT.


—— 2005. Division of labor between Merge and Move: Strict locality of selection and apparent reconstruction paradoxes. Ms., UCLA.


——. 2006b. How unitary are intervention effects? Ms., NYU.
References

283


This page intentionally left blank
Index

A-bar 53–5, 104 n., 136, 151, 155, 166, 168, 169 n., 172–6, 182, 183 n., 185, 192–5, 198, 203–4, 211, 224, 238
Abels, K. 54 n., 110, 238
Accessibility condition 94
adjacency 13, 77, 108, 121, 229
adjacency condition 69, 108
adjunction 16, 26, 47 n., 53, 73, 90, 94 n., 98–106, 108–9, 113, 125, 132, 168, 171, 179
adposition 128, 146, 160 n.
Long-distance Agree 77 n., 96 n.
multiple Agree 113 n., 114, 116 n., 117
agreement 36, 40 n., 53, 82, 93, 104, 113–14, 134, 138, 144–5, 156, 177–187, 189–191, 208–9, 225
complementizer agreement 138, 184 n., 185, 187
Aktionsart 133, 151
Alexiou, A. 111, 113 n.
Anagnostopoulou, E. 38, 111, 113 n., 114
Anti-Interference 52, 55
Anti-Overcrowding 52, 54–5
Anti-Reversal 52, 54–5
A-over-A 17–19, 55, 180
Aoun, J. 94, 177, 212, 214
applicative 132
Arabic 142, 184, 186
Iraqi Arabic 213
Standard Arabic 134
Lebanese Arabic 212, 214
asymmetry 80–2, 84, 87, 90–1, 97, 99, 103, 106–7, 114–15, 119, 164, 173, 202, 236–7 n., 245–6, 248 see also symmetry
Baker, M. 7–8, 69, 74, 76, 83 n., 86–8, 93, 110, 122, 137, 145, 147, 161, 182 n., 194 n.
Bantu 104, 136, 182 n.
bare output conditions 5, 11, 59, 65, 248
see also legibility conditions
bare phrase structure 5, 36 n., 48, 51, 60, 75–7, 85, 109, 120–1, 134, 161, 238, 250 n.
barrier 19–20
Basque 114, 115 n., 177, 182 n., 190
Bavarian 184 n.
Beck, S. 236–7
Bejar, S. 114
Belletti, A. 138, 160 n., 207
Benincà, P. 130–1
Berber 177, 183–4
Berwick, R. 21 n., 234 n.
Bhatt, R. 38–9, 77 n.
binary branching 86, 90, 155, 163–4, 175, 243
binding 43–4, 47 n., 115 n., 205–6, 223–5, 227
Condition A 137
Condition C 42, 137
Condition D 137 n., 224
biology 4, 8, 119 n.
Bloom, P. 11 n.
Borer, H. 9, 68, 74, 76–7, 83 n., 136 n., 140, 177 n.
Distributed Morphology 68–70, 78 n., 94 n.
D-linked reading 215 n., 217, 220
dominance 29, 31 n., 55, 107, 243–4
Donati, C. 37–9, 84 n.
Doron, E. 212
Drury, J. 189, 230 n.
Dutch 130 n., 135, 177, 184–5, 199–200, 202, 230

economy 59, 107, 246
deg 80, 91, 157, 159, 163, 168, 170, 172, 179, 233
degree-feature 78 n., 93 n.
degree-property 78
Edo 86, 177, 186
Einstein, A. 9
Elsewhere principle 219
Embed 67
Embick, D. 4 n., 69
Emonds, J. 30–1, 197, 230
endocentricity 13–14, 51–2, 104
Epstein, S. D. 42–4, 48, 54, 67, 81, 97 n., 150 n., 158 n., 244, 249
Ernst, T. 100, 102
Etxepare, R. 77 n.
Evenki 142
event 72, 78, 98–9, 133–5, 150–1, 155 n.
evolution 4, 78, 102, 120, 248 n.
Ewe 186
experiencer 42–3
extended projection see projection
Extended Projection Principle (EPP) 165, 169, 194
external systems see interface

Faculty of Language (FL) 3, 5, 7–8, 10, 20, 98, 102, 119–20, 244, 247–8
Broad Faculty of Language (FLB) 20
Narrow Faculty of Language (FLN) 20
Falk, Y. 250
Fassi-Fehri, A. 122, 184 n., 186
Case-feature 37, 94, 172, 179, 250
feature inheritance 158, 199
feature-sharing 14, 77, 92 n., 160, 231–2, 248, 250
lexical feature 48, 76–7, 151–2
phi-feature 40 n., 94, 104 n., 146, 155–6, 187, 192, 209–10, 226, 250
theta-feature 76
Finnish 222
Fiorentino 182, 184
Fitch, W. T. 4 n., 9, 20, 146
Fortuny, J. 76 n., 80, 100 n., 139 n., 151–2, 170, 198 n.
Fox, D. 67, 158 n., 223 n., 237, 246 n.
fractal 159–60
Frampton, J. 77 n., 97 n., 236 n.
French 7, 13, 114, 134–5, 140 n., 143, 148, 185, 213, 236
Fujii, T. 180–1, 207
Fukui, N. 36, 55 n., 84, 89 n., 93 n., 122
Full Interpretation 87, 248
Galician 130
Gallego, A. 25, 57 n., 76 n., 94 n., 166, 185, 197, 198 n., 201 n., 202, 204–5
Gavruseva, E. 56 n., 185 n., 230 n.
Gazdar, J. 250
German 148, 221
Germanic 116 n., 221
Giorgi, A. 135, 228
Goal 77, 85 n., 92, 94, 97, 100, 106, 117, 149–52, 154, 157, 170–2, 174, 179, 188–9, 248
Government-and-Binding (GB) 8 n., 20, 32, 52, 77, 101, 213
graph 45, 175, 194
Gracanin-Yuksek 186
Modern Greek 180–1
Grimshaw, J. 14–15, 59, 126–7, 133, 149
Guimaraes, M. 81
Gullah 186
Gutmann, S. 77 n., 97 n.
Huang, C.-T. J. 186
Hinzen, W. 168
Heim, I. 213
Hausa 181 n.
Hauser, M. D. 4 n., 8–9, 20, 78 n., 146
Hebrew 16, 44, 156, 180, 212, 222
Heck, F. 228–9, 231
Heim, I. 71 n., 135
Henderson, B. 207
Hendrick, R. 82 n., 84
Hindi 213
Hinge 118
Hinzen, W. 12 n., 71 n., 76 n., 78 n., 84, 89 n., 137
Hiraiwa, K. 113 n., 117, 126 n.
Hmong 224
White Hmong 186
Hoekestra, E. 52
HPSG 21 n., 250
Hukari, T. 250
Iatridou, S. 38, 211
immediate containment 13, 17
Inclusiveness 47–8, 76 n., 80, 85, 87, 89, 108, 224, 248
conceptual-intentional interface (C-I) 20 n.
conceptual repertoire 64–5, 74
PHON 61–6, 70, 74, 84, 88–9, 91, 97–8, 102–3, 106–7, 112–13, 115 n., 117–19, 134, 135 n., 137, 150–1, 157, 158 n., 165, 207, 216, 233, 244, 246–9
sensory-motor interface (S-M) 20 n.
SEM 63–5, 70–1, 74, 76, 78–9, 82–4, 88–9, 91–3, 97–9, 102–3, 106–7, 115 n., 118–19, 131 n., 134, 135 n., 137, 150, 154, 157, 158 n., 165, 207, 216, 227, 231 n., 233, 238 n., 244, 246–9
syntax-semantics interface 41, 44, 67, 93 n.
interpretability 40 n., 93 n., 97 n.
intervention 40 n., 42, 59, 82–3 n., 236–7
defective intervention 55 n.
Modern Irish 186, 191 n., 209, 211
Irurtzun, A. 40 n., 82 n., 84, 101 n.
“is-a” relation 32, 81–3, 106–7
island 203 n., 205, 215–20, 222, 223 n., 228, 232, 238 n.
strong island (SI) 21–3, 25–7
weak island (WI) 21–3, 235, 236 n.
wh-island 21–2, 55 n., 213, 214–15 n., 238
Italian 116, 130–1, 136, 139–141, 180 n., 185, 203–4
Izvorski, R. 38
Jacak 183 n.
Jayaseelan, K. A. 39, 82, 107 n., 138
Jespersen’s Cycle 143 n.
Julien, M. 69, 110, 122, 128–9, 144, 156
Kandybowicz, J. 184 n., 186, 188–9 n., 190
Kaplan, R. 250
Karitiana 181 n.
Kato, M. 207
Kato, T. 203 n., 231 n., 237, 246 n.
Kayne, R. S. 7–8, 15 n., 20, 31 n., 38, 47 n., 49 n., 50, 52, 54 n., 57 n., 66–7, 73, 76 n., 81, 87, 90, 110, 114, 148, 155, 204, 224–7, 231 n., 233, 237, 243–5, 247
Keizer, E. 199–200
Keysor, S. J. 52, 58 n., 74, 76, 93, 101 n., 246 n.
Kiguchi, H. 239
Kinde 182–3
Kinyarwanda 132
Kitahara, H. 43–4
Koster, J. 20, 30, 31–2 n., 57 n., 84–5, 135, 234, 239 n.
Korean 203 n.
Krapova, I. 51
Index

Mueller, G. 57 n., 165 n., 238
Murasugi, K. 110

narrow syntax 20, 23, 64, 78 n., 80, 88, 89 n., 92, 97, 106–7, 146, 159, 216, 225, 227, 230, 232, 234, 236–7, 244–5, 248 negation 40, 136, 141, 143 n., 144, 156 neuroscience 4
Niinuma, F. 77 n.
Noyer, R. 69
Number 104 n., 127, 134, 136, 141 n., 143, 147, 209
Nunes, J. 13, 25, 47, 67, 69, 75, 98, 100 n., 102, 104, 113, 157 n., 179, 180 n., 224
Nupe 88, 186

optimal design 10–11
Optimal Packaging 118
order-independence 33, 103
overt 40–1, 43, 56 n., 58 n., 93 n., 115 n., 179, 215, 229, 231 n.

Palauan 181 n.
Parallel Morphologies 68, 83 n.
parameter 7–9, 44, 74, 93 n., 215 n.
parasitic gap 231 n., 238 n.
Penthouse principle 230, 235, 237 n.
Pereltsvaig, A. 185 n.
periscope property 13
Perlmutter, D. 52 n., 114, 186, 207
Person 37, 104 n., 114, 134, 138, 143, 146–7, 180–2, 209
Person-Case Constraint (PCC) 52 n., 114
Pianesi, F. 135
Pietroski, P. 41, 71–3, 77–8, 89 n., 98–9, 101–2, 138
phase 19–20, 25 n., 66 n., 112 n., 157–60, 199, 234, 249
Phase-Impenetrability Condition 158
phase transition 160
PHON see interface
phonology 3, 11
phrase structure 5, 15, 36 n., 48, 51, 59–60, 75–7, 81, 83 n., 85, 87, 90, 93 n., 101 n., 102, 107, 109, 120–1, 134, 161, 230, 238, 244, 250 n.
Piedmontese 141–2
Pinkster, S. 5, 11 n., 146
Platzack, C. 127 n.
Poletto, C. 130–1, 138
Polinsky, M. 77 n., 226 n.
Polish 58 n., 130 n., 222
Pollard, C. 250
Pollack, J.-Y. 122, 134, 138
position see also saturation 17, 19, 26, 32, 35, 39, 43, 45, 47 n., 53–4, 56, 104, 106, 125, 165–6, 171, 167–180, 184 n., 185, 188–90, 192–3, 201, 207, 210–11, 214, 233, 236, 239
criterial position 35 n., 36, 165, 233
strong position 36, 165–6
Potsdam, E. 77 n., 192, 226 n.
povety of stimulus 4, 6
preposition 13, 71 n., 111, 126, 136, 160 n., 199, 201, 203 n., 208, 220–2, 231 n.
P-stranding 201, 220–2
Principle of Unambiguous Binding 165 n.
Principles-and-Parameters (P&P) 6–8, 10
Probe 77, 82–3 n., 85 n., 92, 94–7, 100, 104 n., 106, 114, 149–52, 154, 157, 163–4, 170–2, 174, 179, 188–9, 248
simultaneous probing 199
Probe-Label Correspondence Axiom (PLCA) 92, 94–6
pro-drop 136, 184, 202
endocentric projection 13, 163
Index

functional projection: 16, 116 n., 121–2, 156
intermediate projection: 35 n., 51, 75, 103, 161
lexical projection: 14, 122, 161
linker: 122, 161, 182 n.
maximal projection: 13, 48 n., 51, 68, 75, 90, 112 n., 161
minimal projection: 48 n., 68, 75, 90, 161
opaque projection: 122, 161
relator: 122, 160 n., 161
reprojection: 39–44
superprojection: 44, 49
transparent projection: 122, 161
pronoun: 44, 136, 182 n., 220, 224, 239 n.
Prosodic Correspondence Axiom (PCA): 67
quantifier: 40–1, 43, 135, 138, 140 n., 212, 236
quantifier floating: 208
Quechua: 177, 190
Quinn, C.: 146–7
Rackowski, A.: 25, 191, 197
raising: 42–3, 54 n., 148, 179–81, 226 n., 238
Ramchand, G.: 74, 84 n., 128–9, 132–4, 140, 151
rationalism: 5, 33, 120, 246
reconstruction: 47 n., 81, 210
reflexive: 77, 115 n., 135, 137, 220, 224–7
Reinhart, T.: 70, 74, 104 n.
Relativized Minimality (RM): 22–3, 55, 166, 171, 247
relator: 122, 160 n., 161 see also projection
reprojection see projection
resumption: 177, 189, 201, 203 n., 205–7, 211–15, 217–20, 222–3, 225–6, 228, 230–1
Rezac, M.: 36 n., 114, 177 n.
Richards, M.: 55 n., 67, 112 n., 158
Ritter, E.: 134, 146
Rodrigues, C.: 180
Romance: 136, 141, 143, 180 n.
Romanian: 113
Sag, I.: 12 n., 250
Saito, M.: 89 n., 110, 131, 203, 213, 234
Salem, M.: 207
saturation: 36, 83 n.
Savoia, L.: 15, 138
Scandinavian: 187 n., 221
Schweikert, W.: 127 n., 136
scope: 19, 40, 43, 47 n., 138, 190, 205, 236
Scottish Gaelic: 209
Seely, T. D.: 43, 47–8, 60, 81, 82–3 n., 84, 150 n., 158 n., 249
Segal, G.: 40
Selayarase: 182–3, 186–7
select: 75, 81–2, 92–3, 156, 219
Selkirk, E.: 39 n.
SEM see interface
semantics: 3, 68, 71–3, 78, 155
serial verb: 86–7, 117
Sharvit, Y.: 212
Shieber, S.: 77 n., 250
sisterhood: 13–14, 17
Slavic: 51, 148
Soltan, U.: 77 n.
spandrel: 96 n.
Spanish: 105, 111, 202–4, 222
multiple specifiers: 51–2, 54 n., 115–16
Spell-Out: 66–7, 69, 99–100, 249
multiple Spell-Out: 25 n.
Spell-Out domain: 112, 114
Spell-Out unit: 150, 158 n.
Index

Sportiche, D. 15 n., 91, 138, 208, 247 n.
sprouting 217
Starke, M. 22, 39, 107 n., 135, 215 n., 236 n., 239
“start-as” 106–7
Steedman, M. 68, 250
Stepanov, A. 25, 27, 57–8 n., 102 n., 235
Sternefeld, W. 165 n.
Stewart, O. T. 116–8
Stjepanović, S. 238
Stowell, T. 57 n., 122, 239
Structure-Preservation Hypothesis 30, 197
subjacency 19–20, 21 n., 234
subitizing 146
superiority 55 n., 115 n., 210, 235, 238
superprojection see projection
Svenonius, P. 82 n., 129, 141 n., 144, 148–9
Swedish 221
see also asymmetry
spontaneous symmetry breaking 88
syntactocentrism 70
syntax 3, 5, 9, 12, 13 n., 20, 27, 30, 32 n., 34–5, 39 n., 40, 42, 47–8, 60, 63–5, 67–71, 74–80, 82, 84, 89, 93, 96–8, 102, 105, 107–8, 112, 118–20, 123, 133, 137, 143 n., 145–6, 150, 156, 158, 160, 163–4, 165 n., 190, 206, 216, 230–2, 234, 244, 246–9 see also narrow syntax

Tagalog 190–2
Taraldsen, K. T. 185, 207
tense 14, 81, 95, 123, 134–5, 140, 142, 146–8, 155, 158, 160 n., 162, 174, 181, 213
Thai 224
that-trace 56 n., 105, 173, 176, 186, 188–90, 233
theta 53–4, 76, 93 n., 94, 96 n., 100, 103 n., 166, 195, 205, 211, 227, 247 n.
theta-position 53–4, 56, 214
theta-role 53, 54 n., 83 n., 99, 101 n., 104, 150, 246 n.
third factor principle 247
Thornton, R. 230 n.
Tongan 142
Torrego, E. 43, 58 n., 77 n., 85 n., 92–3 n., 95, 106, 110, 157, 172, 198, 203, 204
transformation 17–19, 30, 34, 44–6, 49, 59, 235, 247
Travis, L. 74, 147, 238
Trentino 182, 184
truth 73
Turkish 139 n., 177, 183–4
Uniformity Hypothesis 9, 69 n.
Universal Grammar (UG) 6–9
Ura, H. 60, 113 n.
Uribe-Etxebarría, M. 57 n., 135
Vallader 177, 184–5
Vata 177, 184, 185 n.
vector 90–2, 100, 107
Vehicle Requirement on Merge 85 n., 92 n.
verb cluster 148
Vergnaud, J.-R. 9, 38, 68, 93 n., 172, 237
Vicente, L. 207, 222
virtual conceptual necessity 27, 33
de Vries, J. 230, 231 n.
Wagner, M. 67–8, 70, 76, 79, 118 n.
Wang, A. 215, 217–18, 220–2
Watanabe, A. 81, 104 n., 156, 213–14, 215 n., 225
Weak Cross-Over 238 n.
Webelhuth, G. 228, 230, 231 n.
Weinberg, A. 21 n., 201, 234 n.
Weinberg, S. 10
Welsh 177–8
West Flemish 220
Wiltschko, M. 135
word-order 149, 230
<table>
<thead>
<tr>
<th>Index</th>
<th>295</th>
</tr>
</thead>
</table>


Yiddish 184–5

Yimas 181 n.

Yoon, H.-S. J. 54 n., 129

Yoruba 177, 185 n.

Zaenen, A. 250

Zanuttini, R. 141–2

Zulu 132

Zwart, C. J.-W. 22 n., 25, 51 n., 52, 80, 116, 130 n., 225