Understanding Minimalist Syntax

Generative Syntax

General Editor: David Lightfoot

Recent work in generative syntax has viewed the language faculty as a system of principles and parameters, which permit children to acquire productive grammars triggered by normal childhood experiences. The books in this series serve as an introduction to particular aspects or modules of this theory. They presuppose some minimal background in generative syntax, but meet the tutorial needs of intermediate and advanced students. Written by leading figures in the field, the books also contain sufficient fresh material to appeal to the highest level.

Understanding Minimalist Syntax

Lessons from Locality in Long-Distance Dependencies

Cedric Boeckx

Blackwell
Publishing

For Youngmi,
sine qua non

Contents

Acknowledgments

The idea of writing this book emerged from discussions I had with Luigi Rizzi and Norbert Hornstein in the summer of 2005. Both Luigi and Norbert gave me a sense that a detailed, extended reflection on successive cyclicity and locality might prove useful in the current minimalist context.

I was fortunate to be able to discuss the preliminary results of my investigation with my students Clemens Mayr and Dennis Ott, and with my good friends Kleanthes Grohmann and Adam Szczegielniak. Their writings and their comments have been invaluable to me. I am especially grateful to Kleanthes for writing a paper with me that constitutes the basis of chapter 3.

I have had the privilege discussing foundational issues pertaining to minimalism with Noam Chomsky over the years. Invariably, I have benefited immensely from Noam's insights. Although I am sure he won't be convinced by some of the arguments in this book, especially when I am critical of the concept of "phase," I hope that he will recognize as challenges some of the problems I raise for phase-based computations and standard views on locality. I certainly hope he will solve these problems in future work.

My interest in locality was nurtured by my teachers Howard Lasnik and Željko Bošković when I was a student at the University of Connecticut. I am especially grateful to Željko for the countless discussions we have had on locality and minimalism, for making use of my work in his seminars and publications, and for never failing to encourage me to articulate my thoughts.

As the reader of chapters 4 and 5 will notice, the works of my student Ju-Eun Lee and of my wife Youngmi Jeong on applicatives

have played a central role in my views on locality. Without them, this book would have significantly less to offer.

Thanks are also due to Juan Uriagereka, Angel Gallego, Jim Huang, Marc Richards, Terje Lohndal, So-One Hwang, Bridget Samuels, Suleyman Ulutas, and Ivona Kucerova, for commenting on portions of this book. This work was first presented in its present form in November 2005, at the Harvard Theory Group, organized with taste and talent by Clemens Mayr. Portions of this material were also presented at the University of Connecticut, the University of Maryland, Seoul National University, and the Universitat Autònoma de Barcelona. I am grateful to the audiences of all these meetings for their comments.

As every book-writer knows, book-writing depends in no small part on individuals who do not immediately influence the content of the book, but who nevertheless play a central role in making everything simpler so as to allow the author to focus on writing. In my case, I am indebted to my department administrator and friend Cheryl Murphy, who looks after my every need on a daily basis, to David Lightfoot, who recommended my project for publication in the Generative Syntax series, to Ada Brunstein, Fiona Sewell, and Sarah Coleman at Blackwell, to Terje Lohndal for the index, and to my parents-in-law, who didn't mind my taking time away from vacation to complete a final draft of this book.

Last, but not least, I am most indebted to my wife Youngmi, who not only contributed key ideas to this book, but who also provided patience, understanding, support, and, above all, love – ingredients that not only make writing a pleasure, but constitute the very best of what life has to offer.

Chapter 1

Introductory Remarks

1.1 The Framework

About 15 years ago linguists embarked on a project called the "minimalist program" (or "minimalism"). Minimalism is a research program initiated by Chomsky in two key publications (Chomsky 1993, 1995), and pursued since then by a great many researchers (see Bošković and Lasnik 2006 for a comprehensive collection of minimalist works). Minimalism is an attempt to make sense (in a specific way which I discuss below) of the properties of Universal Grammar that previous research in generative grammar had established, especially those properties uncovered during the so-called "Principles-and-Parameters" era, which crystallized in the late 1970s and early 1980s (see Chomsky 1981).

For 50 years now, linguists and other cognitive scientists have been involved in establishing the necessity of an inborn component of our biological endowment to account for the remarkable (tacit) knowledge and ability we display when we produce and understand (spoken or signed) language. Call this inborn component Universal Grammar (UG). Fifty years of intensive research have shown beyond reasonable doubt that the core properties of our linguistic capacity cannot be acquired by any naïve theory of learning that relies on reinforcement, correction, imitation, memorization, or brute instruction.

Once the existence of an innate language faculty is granted, it is up to linguists and other scientists to determine its content. Succinctly put, the minimalist program explores the possibility that much

of the content attributed to UG by previous research follows from optimal ways of satisfying requirements imposed by the mental modules with which syntax interacts (minimally, the sound/sign and thought systems). The strongest minimalist thesis contends that UG in its entirety is shaped by such optimality requirements. It is often said that were this to be the case, UG would be a "perfect" system for pairing sound/sign and meaning.

Such an ambitious program cannot arise in a vacuum. In order to determine whether UG shows signs of optimal design, one must first establish with some certainty the gross features of UG, for it will be of these features that one will ask whether they have an optimal character. Here, minimalism takes as its point of departure what many, myself included, consider our very best bet as to what the content of UG may be. Technically, it is known as the Principles-and-Parameters approach (often abbreviated as P&P). At the heart of the P&P approach is a distinction between what is invariant across languages, specified independently of linguistic input to the child (what I will refer to as the "principles"), and what is plastic, dependent on properties of the child's linguistic environment, and ultimately what results in linguistic variation (what I will refer to as the "parameters"). In its classic instantiation (Chomsky 1981; see also Baker 2001), the P&P model provides the language learner (child) with a fixed set of principles (laws of grammar, if you wish). Many of these principles contain open values ("parameters"), which the learner must set in the course of language acquisition. One can think of these principles with open values as a menu, a set of courses that the learner can combine in a limited number of ways on the basis of well-defined properties of the linguistic input so as to match the language of her community.

The main advantage of the P&P approach is the principled distinction it draws between invariant and plastic properties of the language faculty. It allows (arguably for the first time in the history of the study of language) linguists to investigate core properties of UG by making abstraction of cross-linguistic variation. The ability to isolate invariant properties of UG was decisive in the formulation of a minimalist program for linguistic theory.

It is fair to say that the P&P approach has been remarkably successful. It has allowed linguists to cover a truly impressive range of similarities and differences across the languages of the world like never before in the history of linguistics, organizing these in a way

that is much less superficial than can be achieved by more traditional approaches to cross-linguistic variation, and in a way that makes sense of the language acquisition process (see Baker 2001, 2005, Boeckx 2006a, Cinque and Kayne 2005, and Yang 2006, for more detailed discussion and references).

The minimalist program grew out of the perceived success of the P&P approach. It took the principles-and-parameters shape of the language faculty for granted. It assumed that the generalizations uncovered under the P&P model were roughly correct. Doing so allowed linguists to focus on a different question, viz. how much of the P&P model could be the direct result of optimal, computationally efficient design. This move in linguistic inquiry is far from trivial. Just imagine: barely 15 years after formulating the first principles and parameters of Universal Grammar, linguists began to ask whether the principles they discovered can be understood in terms of higher standards of inquiry. Do linguistic principles display interesting signs of symmetry, uniformity, economy? Why do we have these principles and not others? How many of these linguistic principles follow from the most basic assumptions/axioms everyone has to make when they begin to investigate language (what Chomsky has called "virtual conceptual necessity")?

As I stated above, the minimalist program for linguistic theory adopts as its working hypothesis the idea that Universal Grammar is "perfectly" designed, that is, it contains nothing more than what follows from our best guesses regarding conceptual, biological, physical necessity. This hypothesis is probably too strong, but in practice scientists often adopt the strongest possible thesis as their working hypothesis. The strongest hypothesis then acts as a limiting case, to see more precisely where and when the hypothesis fails and how much of it may be true.

Chomsky in particular never tires of pointing out in all his writings on minimalism that the minimalist program is, as its name suggests, just a "program," a mode of investigation, and "not a theory" (see, e.g., Chomsky 2000:92, 2002:96; Fitch et al. 2004: appendix). By that Chomsky means that minimalism asks questions and follows guidelines that are broad enough to be pursued in a great many directions. This flexibility, this room for alternative instantiations of minimalism, is what the term "program" emphasizes. At the very beginning of this introductory chapter I used the term "project" to stress the fact that minimalism is neither right nor wrong,

(it may be fruitful, premature, overly ambitious, sterile, fecund, etc.). Its success will be measured, in the long run, by how many insightful hypotheses it helped generate.

The above remarks provide the minimal amount of information necessary to place the present study in its proper context of inquiry. The next section turns to the specific goals and organization of the material developed in subsequent chapters.

1.2   Outline of the Book: Goals and Structure

The principal aim of this book is to shed light on the nature of the minimalist program.

There are at least two ways of "understanding minimalist syntax." One way, which I have pursued at length in Boeckx (2006a), consists in rationally reconstructing the conceptual arguments for a minimalist program in the linguistic theory. Call this the philosophical approach. I hope to have shown in Boeckx (2006a) that minimalist concerns naturally emerge once the logical problem of language acquisition is essentially solved, as it is in the P&P approach. Once a certain level of explanatory adequacy is reached, attempts to be "beyond explanatory adequacy" (to use a phrase introduced in Chomsky 2004a) follow at once. This is by no means peculiar to the practice of the linguistic sciences; it is an inherent property of good scientific practice in general. It is what Richard P. Feynman (1965:26) expressed well in the following quotation:

> Now in the further advancement of science, we want more than just a formula. First we have an observation, then we have numbers that we measure, then we have a law which summarizes all the numbers. But the real *glory* of science is that *we can find a way of thinking* such that the law is evident.

Paraphrasing Feynman, minimalism is in many ways an attempt to find a way, or multiple ways, to make the content of UG evident.

The second way in which to lay bare the internal logic of a research program like minimalism is to simply do it – teach it by example,

as it were. More specifically, select a sufficiently complex, well-studied, and reasonably well-understood phenomenon, dissect it along minimalist guidelines, and see what remains. Call this the empirical approach.

This book takes precisely this tack. It focuses on the well-established phenomenon of successive cyclic movement (the idea, going back to Chomsky 1973, that long-distance dependency formation is actually the conjunction of short dependencies) and tries to determine why such a phenomenon exists, and why it takes the form it does. Adopting a decidedly minimalist perspective, this book is an attempt to show how successive cyclicity is grounded in deeper computational principles of the type minimalism promotes.

Beyond the narrowly empirical concern pertaining to successive cyclicity, I hope to achieve two more general goals in the pages that follow.

First, I hope to show how the phenomenon of successive cyclicity raises questions that touch on virtually all of the issues that are central to minimalist research. I hope that in so doing successive cyclicity can come to be seen as an ideal empirical case study for the program as a whole. If successful, the approach I pursue here will add empirical bite to the program, always a desirable bonus.

Second, I hope to show that although there are many ways in which successive cyclicity could be captured theoretically (I will discuss quite a few of them in the following chapters), adhering to strict minimalist guidelines constrains the choice of possible theories, and leads to better empirical coverage in several domains of grammar.

The book contains four core chapters (chapters 2 through 5), in which I decompose the phenomenon of successive cyclicity into what I will argue are its natural component parts.

Each chapter builds on the conclusions of the previous one. Each chapter is instrumental in eliminating various alternatives that have been entertained in the minimalist literature (and, in some cases, other frameworks as well), either directly or in conjunction with the conclusions I reached in other chapters. Since later chapters build on the conclusions of previous ones, the volume has a funnel-like structure that makes it necessary for the reader to read the chapters in the order I have chosen.

In the first core chapter (chapter 2) I review the reasons why successive cyclicity was proposed, and what kind of empirical evidence

is used to support it. As we will see, the evidence for successive cyclicity is quite substantial, and not limited to narrowly syntactic considerations. We will see that morphological, phonological, and semantic properties of language are best captured if successive cyclicity is assumed. Once it is established that successive cyclicity exists (i.e., that displacement in natural language is bounded), the first question to ask is what the relevant cycles, or boundaries imposed on movement, are. Does movement stop at selected points (skipping positions that appear to be available), or everywhere it can? This is the question I address in chapter 3. In chapter 4, I consider the timing of the intermediate steps of movement. When exactly does the moving element start moving: as soon as it can, or not until a fair amount of structure is built? In chapter 5 I try to determine whether intermediate steps of movement, the conjunction of which results in long-distance dependencies, are triggered by some requirement such as feature checking (a prime motivation for movement in the minimalist program), or motivated in some other way. The core structure of the book, then, looks like this:
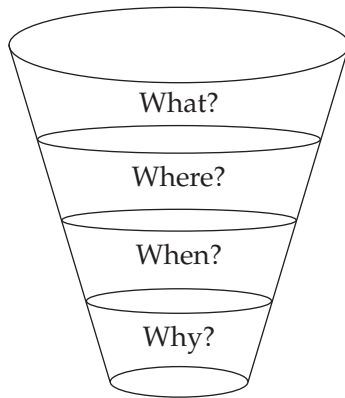
It is important to bear in mind that the questions raised in the following chapters are not new. Most, arguably all of them, were raised in some fashion in Chomsky (1973), the work that introduced the concept of successive cyclic movement. What is new is the theoretical context in which such questions are raised, and the type of answers that minimalism favors.

In chapter 6 I consider a few salient alternative views on successive cyclicity that are compatible with most of the conclusions reached in the previous chapters. After briefly sketching each of them, I provide conceptual and empirical arguments against them, and show the (conceptual and empirical) superiority of the analysis developed here.

Chapter 7 touches on a very broad question: what is the relationship between the type of locality concerns responsible for the phenomenon of successive cyclic movement and other types of locality (island effects, intervention/minimality effects, etc.)? Can we bring all types of locality under a unified theory? These are clearly questions that fall beyond the range of investigation that can be undertaken here. It requires developing equally detailed analyses of islands, intervention, etc. Since the nature of islands and intervention has been the focus of my work until now, I draw on the results I have achieved so far, and sketch ways in which the view on successive cyclicity I propose here can be integrated with them. Needless to say, the conclusions reached in this chapter will be tentative, and will require further extensive investigation.

Chapter 8 is a brief concluding chapter, which summarizes the major results of this study, and asks whether successive cyclicity meets genuinely minimalist expectations.

Let me conclude this introductory chapter by saying a few words about the intended audience for this work. I agree with Chomsky (2000:141n.13) that "[i]t is a misunderstanding to contrast 'minimalism and X,' where X is some theoretical conception. . . . X may be pursued with minimalist goals, or not." It is essentially for this reason that I think that many of the questions that I raise here are, if not theory-neutral, at least relevant to many frameworks. Displacement is a fact about natural languages that all frameworks have to come to grips with. All frameworks have to take a stance on how long-distance dependencies are formed. All frameworks will therefore be interested in the evidence I will use below to justify some of the conclusions I reach.

The book presupposes only minimal exposure to syntactic theory: basic knowledge of Phrase Structure, transformations, binding, etc. Familiarity with the material discussed in introductory syntax textbooks such as Carnie (2002) will be assumed throughout. Familiarity with the minimalist program is, of course, always desirable, but by

no means necessary. Although the book does not intend to provide a comprehensive introduction to the minimalist program (for such comprehensive introductions, I refer the reader to Lasnik et al. 2005 and Hornstein et al. 2006), each key notion of the minimalist program used here is introduced in an exhaustive fashion when it is mentioned for the first time, making the book relatively self-contained.

Chapter 2

The Marks of Successive Cyclicity (The *What*-Question)

For all their modernity and insights into the fundamental workings of language, beautifully retraced and highlighted in Lasnik (2000), Chomsky's early writings (Chomsky 1955, 1957) contain a curious gap: they do not contain any explicit discussion of locality. One does not even find any extensive discussion of the fact that movement appears to be potentially unbounded. This gap is all the more curious from our current perspective, where locality and long-distance dependencies are arguably the major area of study in theoretical linguistics.

2.1 Subjacency and the Emergence of Successive Cyclicity

We owe our modern interest in locality to Ross's (1967) seminal work in which the concept of island was introduced. Ross's thesis is full of examples of long-distance dependencies like (1a, b).

(1) a. Handsome though Dick is, I'm still going to marry Herman.
     b. Handsome though everyone expects me to try to force Bill to make Mom agree that Dick is, I'm still going to marry Herman.

Ross systematically investigated the fact that seemingly minute manipulations dramatically affected the acceptability of sentences. Witness (2a, b).

(2) 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.

Ross's study contains a list of contexts, technically known as *islands*, which disallow certain types of dependencies. (Ross distinguished between rules that leave a gap, so-called feature-changing rules, and rules that leave a resumptive pronoun, so-called copying rules. Only the former are subject to islands. I will come back to this issue in chapter 7. See also Boeckx 2001a, 2003a.)

   Chomsky (1973) set out to investigate what the various domains Ross identified as islands have in common. Thus began the modern study of locality, and, in many ways, the nature of current linguistic theorizing. Chomsky's central insight in 1973 is that movement is subject to the *subjacency* condition, a condition that forbids movement from crossing two bounding nodes. For Chomsky, the bounding nodes were the top clausal node (S; our modern IP) and NP (our modern DP). The condition correctly captured the unacceptability of (2b), but wrongly predicted (2a) to be out.

(3)   *[Handsome$_i$ though [$_\text{S}$ I believe [$_\text{NP}$ the claim that [$_\text{S}$ Dick
       is t_i]]]], I'm still going to marry Herman.

(4)   [Handsome$_i$ though [$_\text{S}$ I believe that [$_\text{S}$ Dick is t_i]]], I'm still
       going to marry Herman.

   To correct this undesirable effect of subjacency, Chomsky hypothesized that long-distance movement proceeds in short steps, passing through successive cycles. In particular, Chomsky postulated that movement can stop by at the edge of the clause (S' or COMP; the modern CP area). In other words, instead of moving long-distance in one fell swoop, movement first targets the closest clausal edge, and from there proceeds from clausal edge to clausal edge, typically crossing only one S(/IP)-node at a time.

(5)   [Handsome$_i$ though [$_S$ I believe [$_{S'}$ t'_i that [$_S$ Dick is t_i]]]], I'm still going to marry Herman.

Successive cyclicity may at first seem like a patch, an exemption granted to fix a bad problem (without it, the theory would wrongly rule out acceptable constructions). But subsequent research has uncovered a wealth of data, to be reviewed in this chapter, that converge and lend credence to the successive cyclic movement hypothesis, making it one of the great success stories of modern generative grammar. It appears to be the case that long-distance dependencies are the result of the conjunction of small steps.

2.2   The Evidence

The empirical evidence for intermediate steps of movement comes from all the major domains of linguistic investigation: syntax, morphology, semantics, and phonology.

To my mind, the syntactic and semantic evidence, though subtle, is strongest as it relies on reasonably well-understood concepts. Morphological and phonological effects are by nature more directly perceivable, but they largely depend on what we take to be the mapping from narrow syntax to the sensory-motor systems ("Phonetic Form" (PF)), a mapping whose nature is at the moment far from trivial.[1] Having said this, evidence is evidence, and although one must always keep an open mind, and be ready for surprises, I think that taken as a whole, the evidence I will review in the remainder of this chapter strongly suggests that successive cyclicity is a core feature of natural languages.

2.2.1   Syntax

One of the first reflexes of successive cyclicity uncovered by syntacticians is the presence in embedded clauses of operations that are conditioned in main clauses by wh-movement (the prototypical instance of long-distance movement). For example, in English, wh-movement in main clauses triggers subject–auxiliary inversion. The effect of wh-movement is very local. The presence

of wh-movement in an embedded clause does not trigger wh-movement in a matrix clause:

(6)   I told you John is asking who left.

(7)   *I told you is John asking who left.

From this contrast syntacticians typically conclude that subject–auxiliary inversion takes place in the context of wh-movement only in the region where the wh-phrase ends up being pronounced; more technically, in those wh-movement contexts where the C-position which the auxiliary moves to also hosts a wh-phrase in its specifier.

Interestingly for our purposes, some languages are more revealing than English about the derivational history of the wh-phrase that ends up in SpecCP. Indeed many languages not only invert the subject and the auxiliary where the wh-phrase lands, they also invert the subject and the auxiliary (or the entire verbal complex) in all the clauses that wh-movement crosses.

The first example of this highly revealing phenomenon was provided by Kayne and Pollock (1978) in their seminal study of so-called French stylistic inversion (an operation which inverts the usual subject–verbal-complex order in various contexts such as wh-movement, and thus very similar to English subject–auxiliary inversion). The examples in (8a) through (8c) show that French stylistic inversion is sensitive to the presence of wh-movement.

(8)   a.   Jean est parti.
           Jean is   left
           'Jean has left.'
      b.   *Est parti Jean?
           is   left   Jean
           'Has left Jean?'
      c.   Où   est allé   Jean?
           where is   gone Jean
           'Where did Jean go?'

The example in (9) shows that inversion takes place not only in the matrix clause but also in the embedded clause from which the wh-phrase originated.

(9) Où   crois-tu   qu'est allé Jean __ ?

where believe-you that is gone Jean
'Where do you believe that Jean went?'

Spanish (Torrego 1984; Baković 1998) and Basque (Ortiz de Urbina 1989) exhibit a similar phenomenon.[2] Here are a few examples. (For brevity's sake, I omit examples illustrating inversion in simple, mono-clausal contexts.)

(10) a. <u>Que</u> **dijo** Luis [que la gente **decia** [que el diario
         what said Luis that the people said   that the paper
         habie publicado]]?
         had   published
         'What did Luis say that the people were saying that the newspaper had published?'
     b. <u>Que</u> **dijo** Luis [que **decia** la gente [que **habia publicado** el diario]]?

(11) <u>Nork</u> **uste duzu** [**esan du-ela**   Mikelek [**idatzi du-ela**
     who think aux   say   aux-that Mikel   write   aux-that
     eskutitza]]?
     letter
     'Who do you think that Mikel said wrote the letter?'

As a matter of fact, even dialects of English like Belfast English (Henry 1995) offer similar data.

(12) a. Who did John hope would he see?
     b. What did Mary claim did they steal?
     c. I wonder what did John think would he get.
     d. Who did John say did Mary claim had John feared would Bill attack?

The successive cyclicity hypothesis readily accounts for such phenomena: wh-movement triggers subject–auxiliary inversion upon passing through each SpecCP position. I find it in fact very hard to even think of an alternative explanation for such facts.[3]

2.2.2   Morphology

In addition to this piece of syntactic evidence, some languages show morphological reflexes of successive cyclicity.

2.2.2.1   Complementizer alternations in Irish

The first kind of evidence I will discuss comes from Irish. (For a standard description of the facts reviewed here, see McCloskey 1979. For a more comprehensive view, see McCloskey 2002 and Boeckx 2001a, 2003a.) The language has a complementizer *go*, equivalent to English *that* in *John said that Mary likes sushi*.

> (13)   Creidim gu-r   inis sé bréag.
>   I-believe **GO**-[Past] tell he lie
>   'I believe that he told a lie.'

The complementizer receives a different shape (a^L, L indicating lenition) when wh-movement takes place.

> (14)   an ghirseach a ghoid na síogaí *t*
>   the girl   **a^L** stole the fairies
>   'the girl that the fairies stole away'

Again, remarkably, it is not just the complementizer closest to the wh-phrase that is affected by this change, but all complementizers on the path of the moving wh-phrase that turn into a^L.

> (15)   Aon bhliaim déag is   dóigh liom   a
>   one year   ten   **a^L**-COP [Pres] I-think **a^L**
>   deireadh   m'athair   a bhí sé nuair . . .
>   say [Past-Habit] my father **a^L** was he when
>   'It's eleven years old that I think my father used to say that
>   he was when . . .'

   More complex patterns of wh-movement in Irish provide even stronger evidence for successive cyclic movement. In addition to the a^L-pattern, Irish also shows complementizer alternations in the domain of resumption; that is, in contexts where a pronoun fills

the gap left by wh-movement. In such contexts, the complementizer hosting the wh-phrase is a^N (N indicating nasalization), as seen in (16).

(16) an fear **a** bhuail tú   é
      the man a^N struck you him
      'the man that you struck'

In long-distance dependencies involving resumptive pronouns, Irish only requires that the complementizer closest to the ultimate landing site of the wh-phrase be changed to a^N. Other complementizers along the path of wh-movement can surface as GO.[4] Witness (17).

(17) cúpla muireara **a** bhféadfá   a rá **go** rabhadar bocht
      couple household a^N you-could say GO were   poor
      'a few household that you could say were poor'

From the examples in (16) and (17) one can conclude that the complementizer hosting a wh-phrase linked to a resumptive pronoun (even across separate clauses) surfaces as a^N. With this generalization, let me now turn to the so-called mixed patterns uncovered by McCloskey (2002).

The example in (18) shows that a series of a^L-complementizers can also surface along the path of wh-phrase+resumptive pronoun dependency.[5]

(18) rud   **a** raibh coinne   aige   **a** choimhlíonfadh____
      thing a^N was expectation at-him a^L fulfill
      an aimsir
      the time
      'something that he expected time would confirm'

Interestingly, as the example in (19) shows, the opposite pattern (a^N-complementizer along the wh-movement path, a^L at the final landing site) also exists.

(19) aon duine **a** cheap   sé **a** raibh ruainne tobac   aige
      any person a^L thought he a^N was   scrap   tobacco at-him
      'anyone that he thought had a scrap of tobacco'

Now, if the generalization that the complementizer hosting a wh-phrase linked to a resumptive pronoun surfaces as a^N is correct, one is forced to conclude from examples like (19) that (at the very least some) long-distance wh-dependencies are formed by the conjunction of smaller steps.

2.2.2.2 Agreement with the moved wh-phrase in Kinande

Kinande (Schneider-Zioga 1995 and subsequent work) appears to behave like Irish.[6] Local wh-movement obligatorily triggers agreeing complementizer (unlike many other operations that affect complementizers in the same morphological fashion, wh-movement in Kinande triggers agreement on the complementizer; that is, the shape of the complementizer is sensitive to the peculiar features of the wh-phrase). Relevant examples appear in (20).

(20) a. EkIhI **kyO**   Mary 'a-ka-langIra   *t*?
          what.7 Comp.7 Mary.1 AGR.1-Pres-see
          'What does Mary see?'
      b. aBahI **BO**   Yosefu   alangIra *t*?
          who.2 Comp.2 Joseph.1 saw
          'Who did Joseph see?'

It will come as no surprise to the readers that wh-movement in Kinande (at least optionally) has the ability to affect all the complementizers on its path.[7]

(21) a. EkIhI **kyO**   Mary' akaBula [CP nga-**kyO**
          what.7 Comp.7 Mary wonders   if-Comp.7
          Yosefu a-ka-langIra   *t*]?
          Joseph.1 AGR.1-Pres-see
          'What does Mary wonder if Joseph sees?'
      b. EkIhI **ky'**   u-ka-BUla   Yosefu [CP ng' aha *t*
          what.7 Comp.7 2-Pres-wonder Joseph   if   gave
          ky'
          AGR.
          What do you wonder if Joseph gave to the children?'

Facts like (21a) straightforwardly follow if long-distance dependencies are formed step by step.

2.2.2.3   *Wh-agreement*

Since I just mentioned instances of agreement with a moved wh-phrase, let me mention other instances of morphological markings that are standardly mentioned in this context. I have in mind here the so-called "wh-agreement" (also called "Op(erator)-agreement") phenomenon found in various Austronesian languages (but not restricted to this language group).[8]

The best-studied case here is Chamorro (Chung 1982, 1994; Chung and Georgopoulos 1988; all data are from Chung 1982 unless stated otherwise). Similar facts have been studied in detail for Palauan by Georgopoulos (1985, 1991), and for Tagalog by Rackowski and Richards (2005) and Pearson (2005, and references therein).

The "agreement"[9] patterns in Chamorro are as follows:

(i)   Wh-agreement with a subject gap is realized as *-um-* if the verb is realis transitive (instead of Ergative Agreement).
(ii)   Wh-agreement with an object gap is realized (optionally) as nominalization plus *-in-*.
(iii)   Wh-agreement with an oblique gap is realized as nominalization.

Here are some illustrations of each kind, contrasting wh- and non-wh-contexts.

Subject gaps (normal word order VSO):

(22)   a.   Hayi f-**um**-a'gasi i   kareta?
              who UM-wash   the car
              'Who washed the car?'
         b.   **Ha**-fa'gasi si   Juan i   kareta.
              E3s-wash   Unm Juan the car
              'Juan washed the car.'

Object gaps, option 1 (nominalization):

(23)   a.   Hafa f-**in**-ahan-ña   si   Maria gi   tenda?
              what IN-buy**+Nmlz**-her Unm. Maria Loc store
              'What did Maria buy at the store?'
         b.   Ha-fahan si   Maria I sanhilo'-ña gi   tenda.
              E3s-buy   Unm Maria I blouse-her Loc store
              'Maria bought her blouse at the store.'

Object gaps, option 2 (no nominalization, no special morphology):

(24)   Hafa ha-fahan si   Maria gi   tenda?
       what E3s-buy Unm Maria Loc store
       'What did Maria buy at the store?'

Oblique gaps:

(25) a. Hafa puno'-mu   ni   lälu'?
          what kill**+Nmlz**-your Obl fly
          'What did you kill the fly with?'
      b. Hu-punu' i   lälu'ni   niús.
          E1s-kill   the fly Obl newspaper
          'I killed the fly with the newspaper.'

What we see here is agreement in the sense that morphology on the verb varies depending on the presence of a moved wh-phrase. In the case of long-distance extraction, higher verbs agree as though the clause containing the gap were the gap. (This is also the case in Palauan and Tagalog.)

(26) a. Hafa **um**-istotba hao [ni   malago'-ña   i
          what UM-disturb you Comp what**+Nmlz**-his the
          lahi-mu   *t*]?
          son-your
          'What does it disturb you that your son wants?'
      b. Hayi **um**-istótotba   hao?
          who UM-disturb+1pf you
          'Who is disturbing you?'

(27)   Hafa s-**in**-angan-ña   si   Juan [pära godde-tta
       what IN-say**+Nmlz**-his Unm Juan Fut   tie**+Nmlz**-our
       ni   chiba *t*]?
       Obl goat
       'What did Juan say that we should tie up the goat with?'

Notice that the wh-phrase itself cannot trigger agreement on any clause except the one it extracts from.

(28) *Hafa hassóso-mmu   [pära fa'gase-mmu
     what think+1pf**+Nmlz**-your Fut   wash**+Nmlz**-your
     ni   kareta *t*]?
     Obl car
     'What are you thinking of washing the car with?'

Facts like (26)–(28) show that the term "wh-agreement" may not be quite adequate for Chamorro (and languages behaving similarly). Although agreement on the verb is clearly sensitive to wh-move-ment, the specific morphology on the verb is not determined by the featural content of the wh-phrase (except in the clause from which wh-movement was launched), unlike what we saw above for Kinande. Although I find instances of "wh-agreement," of both the Kinande and Chamorro types, fascinating, and clearly related to wh-extraction, I don't think they provide as compelling evidence for the formation of local movement steps, although they are repeat-edly mentioned in all studies of successive cyclicity.

Unlike the subject–auxiliary inversion cases discussed above, or the Irish complementizer facts, which appear to be truly dependent on the presence of a wh-phrase in a very local specifier, the mor-phological markings we find in Kinande and Chamorro may not provide such uncontroversial evidence for local wh-movement steps. Let me explain why.

Contrary to much work in the 1980s and early 1990s, it is now thought that agreement does not require actual displacement of the element triggering agreement. Such a conclusion was reached in part of the basis of empirical facts like these:

(i)   situations where possible movement relations would wrongly feed agreement relations that are not attested (and are correctly predicted to be impossible if agreement is determined prior to movement);

(ii) situations where it is clear that the agreeing DP has not moved anywhere close to the domain of the agreeing head (say, an agreeing verb).

I have documented such situations in Boeckx (2000a, 2003b, 2004, 2006c), and Boeckx and Niinuma (2004). I have reviewed various sorts of arguments in favor of the establishment of agreement at a distance (with no displacement) in Boeckx (2007), and won't repeat

the relevant data here. Perhaps the best-known case of long-distance agreement comes from existential constructions like (29).

(29)   There seem (to John to appear to Mary) to be two men in the garden.

To capture such examples, Chomsky (2000) introduced the operation *Agree*, which consists of an element (called the Goal) transferring its features onto another element (called the Probe), which lacks such features intrinsically. In the case of (29), *two men* acts as the Goal, and the finite verb acts as the Probe.

Suffice it to say that if agreement can be established independently of movement, under Agree, wh-agreement facts can no longer argue on their own for the creation of intermediate movement steps in long-distance dependencies. One could easily imagine a scenario under which that agreement is established at a distance, prior to any wh-movement. Movement itself would take place in one fell swoop, as depicted in the figure shown here.
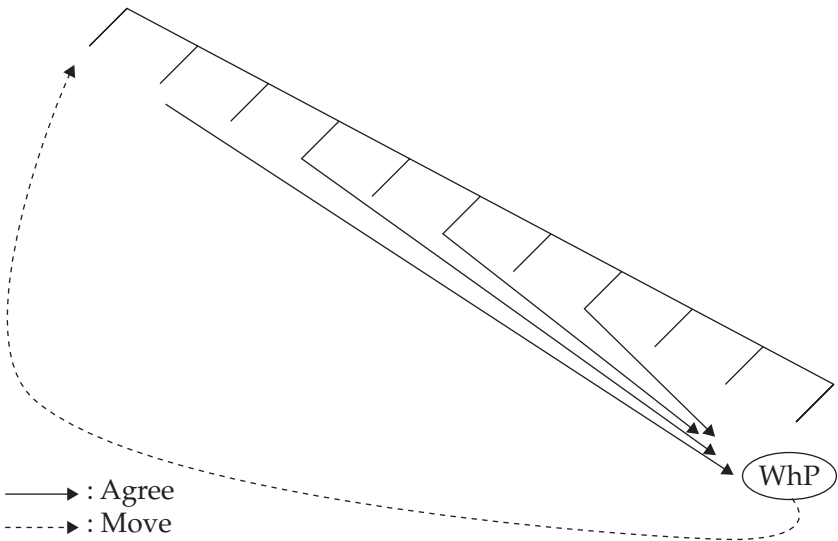
$\longrightarrow$: Agree
- - - -▸ : Move

Of course, in such a scenario, one still has to account for why agreement obtains along the wh-movement path. Here a key feature of Rackowski and Richards's (2005) analysis of Tagalog wh-agreement may fit the bill.[10]

Suppose that all finite clauses in the relevant languages are islands for extraction, blocking any kind of long-distance dependency. Suppose further, as Rackowski and Richards have argued, that agreement unlocks islands, rendering the clauses transparent for purposes of extraction. It follows that in order for long-distance extraction to be possible, the unlocking effect of agreement ("wh-agreement") will have to be present all along the extraction path, but the extraction itself may well take place in one fell swoop. This suggestion concerning wh-agreement is not outlandish, considering that the agreement in question refers to a relation between an embedding verb and the embedded clause from which extraction takes place.

I do not want to claim that this is the correct analysis of wh-agreement. To make such a claim would require me to discuss the nature of islandhood and agreement more thoroughly than I can do at this stage (see chapter 7 for additional discussion; see, especially, Boeckx 2001a, 2003a). For now, all I want to stress is that morphological reflexes of the type Chamorro exhibits are compatible with successive cyclicity, but offer less direct, compelling, theory-neutral evidence for actual successive cyclic movement than is often thought.

2.2.2.4   *Subject alternation in Ewe*

A perhaps more compelling piece of morphological evidence for successive cyclic movement comes from the Ewe data discussed in Collins (1993).

Collins (1993:157) observes that the morphology of third subject pronouns in Ewe is sensitive to the presence of a wh-phrase in SpecCP. Not only is this the case in matrix questions (obligatorily). It also holds in cases of long-distance extraction (optionally).

(30)   **E/*wo** fo Kosi.
     he   hit Kosi
     'He hit Kosi.'

(31)   Kofi gblo be   **e/*wo** fo Kosi.
     Kofi said that he   hit Kosi
     'Kofi said that he hit Kosi.'

(32)   Kofi bie   be lamata ***e/wo** fo Kosi.
     Kofi asked that why   he   hit Kosi
     'Kofi asked why he hit Kosi.'

(33) Me e   gblo be **e/wo** fo?
      who you said that he   hit
      'Who did you say that he hit?'

Admittedly we do not fully understand what triggers the change in subject pronoun in matrix questions, but that the change appears to be dependent on a wh-phrase in the vicinity of the subject pronoun suggests that movement may at least proceed in intermediate steps in long-distance cases.

2.2.3   Phonology

Phonological evidence for successive cyclic movement comes from the presence vs. absence of tonal downstep in Kikuyu (cf. Clements et al. 1983; Clements 1984; Haïk 1990). The facts are as follows.

In affirmative clauses in Kikuyu, a downstep morpheme occurs V-initially:

(34) Kariokǐ á-!t-ɛ́m-írɛ́   mo-tě!
      Karioki SubjPref-cut-Tense Class-tree
      'Karioki cut a tree.'

In questions, downstep disappears.

(35) Nó-o   o-tɛm-írɛ́   mo-te?
      Focus-who Prefix-cut-Tense Class-tree
      'Who cut a tree?'

Interestingly for our purposes, downstep disappears on all verbs between the trace and the moved wh-phrase:

(36) Nó-o   γw-eciíri-a [$_{CP}$
      Focus-who SubjPref-think-Tense
      Ngóγe a-úγ-írɛ [$_{CP}$ áte   *t* o-On-írɛ
      Ngugi SubjPref-say-Tense Comp   Prefix-see-Tense
      Kaanakɛ]]?
      Kaanake
      'Who do you think Ngugi said saw Kaanake?'

Kikuyu is the only language I am aware of that offers phonological reflexes of successive cyclic movement. But this may be due to the fact that the relation between tone and syntactic derivations hasn't been studied as much as it ought to be. Due to this fact, one ought to take this piece of evidence for successive cyclic movement with a grain of salt, until we can establish more conclusively that downstep is dependent upon the very local presence of a wh-word in the vicinity of the relevant verb. The fact that phonological relations tend to be highly local is suggestive, but again I want to warn the reader against taking for granted the standard interpretation of potentially very revealing facts.

2.2.4   Semantics

Another kind of evidence for the existence of successive cyclic movement comes from interpretive considerations; more specifically, from so-called reconstruction effects. (To the best of my knowledge, the logic of the phenomena discussed in this section goes back to Barss 1986.) Consider sentences like (37).

(37)   Who said that John$_i$ thinks that Bill$_j$ bought pictures of himself$_{*i/j}$?

The anaphor *himself* can only take *Bill* as its antecedent, not *John*. The standard explanation for this phenomenon is that antecedence is determined in terms of closest c-command. Now witness (38).

(38)   [Which pictures of himself$_{i/j}$] does John$_i$ think that Bill$_j$ bought?

Here *himself* can take either *Bill* or *John* as its antecedent. The best explanation we have for this fact was offered by Chomsky (1993). Chomsky proposed we view movement as a copying operation. Upon movement, a copy is created and remerged higher up in the syntactic structure. If we adopt the copy theory of movement, the original copy of the entire wh-phrase is located in the complement position of *bought*, as in (39).

(39)   [Which pictures of himself] does John$_i$ think that Bill$_j$
       bought <which pictures of himself>?

From this perspective (39) is not very different from (37). We there-
fore expect only *Bill* to be able to function as antecedent for the
reflexive *himself*. However, Barss (1986) observes, if long-distance
movement proceeds successive cyclically, a copy of *which pictures
of himself* will also be found in the intermediate SpecCP, as repres-
ented in (40).

(40)   [Which pictures of himself$_{i/j}$] does John$_i$ think <which
       pictures of himself> that Bill$_j$ bought <which pictures of
       himself>?

This representation makes *John* available as antecedent for the
reflexive, as it is the closest c-commanding antecedent from the per-
spective of the copy of the wh-constituent in the intermediate
SpecCP. Interpretive considerations like the one just discussed thus
offer fairly compelling evidence for the physical presence of copies
of the moved element in intermediate landing sites, precisely
where we expect them according to the successive cyclic movement
hypothesis.
    Similar considerations hold in an even stronger form in cases like
(41) (from Fox 2000:10–11; see also Lebeaux 1998).

(41)   [The papers that he$_i$ wrote for Ms. Brown$_j$], every student$_i$
       asked her$_j$ to grade.

Standard binding theoretic considerations (variable binding and
Condition C effects; see Chomsky 1981) force the interpretation of
the moved constituent in an intermediate site below *every student*
(to allow binding of *he*, which requires c-command) and above *her*
(to prevent a pronoun from c-commanding its antecedent, as man-
dated by Condition C of the binding theory).
    The conclusion reached on the basis of binding considerations also
seems to hold on the basis of scope considerations. Perhaps the clear-
est piece of evidence here comes from the argument from the inter-
pretation of nominal modifiers, as described and analyzed in Bhatt
(2002), and subsequently reanalyzed in Hulsey and Sauerland (to
appear) and Bhatt and Sharvit (2006, who address criticisms raised
by Heycock 2005).[11]

Bhatt's central new empirical point is that adjectival modifiers of the head NP of a relative clause can be interpreted within the scope of a propositional attitude verb. The modifiers with which he is most concerned are superlatives, the ordinals *first* and *last*, and *only*. He exemplifies the core empirical distinction with data like (42) and (43):

(42) the first book that John said Tolstoy had written
"High" reading:
In 1990, John said that Tolstoy had written *Anna Karenina*; in 1991 John said that Tolstoy had written *War and Peace*. Hence the NP is *Anna Karenina*.
(That is, order of *saying* matters, order of *writing* is irrelevant.)
"Low" reading:
John said that the first book that Tolstoy had written was *War and Peace*. Hence the NP is *War and Peace*.
(That is, order of *writing* matters, order of *saying* is irrelevant.)

(43) a. the only book that John said that Tolstoy had written
"High" reading:
x is the only book about which John said that Tolstoy had written x
"Low" reading:
What John said can be paraphrased as 'x is the only book that Tolstoy wrote.'
b. the longest book that John said that Tolstoy had written
"High" reading:
x is the longest book out of the books about which John said that Tolstoy had written them.
"Low" reading:
What John said can be paraphrased as 'x is the longest book that Tolstoy wrote.'

The basis of Bhatt's explanation for the existence of these different readings is an analysis of relative clauses in which the head NP and its modifiers originate within the relative clause (in the position of the "gap"); subsequently this constituent moves from this position to a position to the left of the clause. Crucially, this movement results in a collection of copies (technically known as a chain). The interpretive component is free to choose from this chain in coming up with an interpretation for the sentence. I will not go into the details

of the derivation here. The essential point is that subsequent LF processes have access to a representation along the lines of (44).

(44)   [the longest book that John said <[the longest book]> that Tolstoy had written <[the longest book]>

Low and high readings are the result of interpreting copies next to the verb they modify.

As far as I can see, the reconstruction data, rigorously established in works on the syntax–semantics interface, provide evidence of the strongest kind for the moving element passing through intermediate landing sites in the creation of long-distance dependencies.

2.2.5   Morpho-syntactic evidence from overtly stranded pieces

Additional evidence for successive cyclic movement comes from situations where pieces of the moving elements are left behind in intermediate sites. The most famous example of this sort comes from the position of floated quantifiers like *all*.

Most varieties of English allow sentences like (45).

(45)   What **all** did you get for Christmas?

McCloskey (2000) discusses data from Irish English varieties that in addition to (45) also allow (46).

(46)   What did you get **all** for Christmas?

Similar facts obtain in long-distance extraction cases. Witness (47) and (48).

(47)   What **all** did John say that Peter ate for breakfast?

(48)   What did John say that Peter ate **all** for breakfast?

If we follow Sportiche (1988) in taking quantifier float to be the result of stranding of *all*, and if we follow Bošković (2004) in saying that stranding of *all* is disallowed in positions where theta-roles are

assigned, data like (46) and (48) provide evidence for intermediate steps of movement. But, as the reader will no doubt have noticed, there are two *if*s in the chain of reasoning just provided. This means that the stranding data are good insofar as actual stranding is taking place. An alternative view on floated quantifiers (which exists; see Bobaljik 1998 for review) may nullify the evidence.

   Similarly, if Boeckx (2001a, 2003a) is right in analyzing resumptive pronouns as stranded material like *all*, data like (49) also provide evidence for the claim that long-distance dependencies result from the conjunction of short dependencies (data from Hebrew, taken from Sells 1984:92–3). (Under Boeckx's analysis, resumptive pronouns form a constituent with their antecedents, a unit split by subsequent movement of the antecedent phrase.)

(49)   a.   ha-ʔiš   še   ʔani xošev še   ʔamarta še   sara
            the-man that I   think that said-you that Sarah
            katva ʔalav   šir
            wrote about-him poem
       b.   ha-ʔiš še ʔani xošev še ʔamarta še ʔalav sara katva šir
       c.   ha-ʔiš še ʔani xošev še ʔalav ʔamarta še sara katva šir
       d.   ha-ʔiš še ʔalav ʔani xošev še ʔamarta še sara katva šir
       e.   ha-ʔiš ʔalav ʔani xošev še ʔamarta še sara katva šir
            'the man that I think that you said that Sarah wrote a
            poem about (him)'

Other potential instances of stranding under A-bar movement mentioned in Boeckx (2003a:167 n.6) include: stranding of the "the hell" part in so-called aggressively non-D-linked interrogatives "who the hell" in Hebrew (50) (a fact first observed to my knowledge by Obenauer 1994), "else" stranding in Child English (51) (Rosalind Thornton, p.c.) and in Dutch (52) (Elly van Gelderen, p.c.), and (in my dialect at least) "among"-phrase stranding in French (53).

(50)   Ma   (laʔazazel) amar jon (laʔazazel) še   mary
       what the devil   said   Jon   that Mary
       raʔata (laʔazazel)?
       saw
       'What the hell did Jon say that Mary saw?'

(51)   What do you want else to eat?

(52)   Wat wil je   anders nog   eten?
       what will you else   again eat
       'What else will you eat more?'

(53)   Qui (d'entre nous) crois-tu ?(d'entre nous) que Jean a
       who among us   believe-you   that Jean has
       dit ?(d'entre nous) que Pierre va   appeler (d'entre nous)?
       said   that Pierre will call
       'Which one of us do you believe that Jean said that Pierre
       will call?'

To repeat, the stranding data depend on controversial views about floated quantifiers and resumptive pronouns. Although it seems to me that the stranding analysis in such cases is superior to alternatives (see Boeckx 2003a:ch. 2 and Bošković 2004 for extensive discussion and arguments), the argument may be more theory-internal than one would want.

More compelling evidence of the stranding sort comes from the presence of entire copies of the moved element that are pronounced in intermediate landing sites. These are so-called wh-copying constructions, originally discussed in du Plessis (1977) and thoroughly analyzed in modern terms in Nunes (2004).

Several languages, such as Afrikaans, German, Romani, Frisian, and Child English, readily allow cases of long-distance questions like (54)–(56), in which the wh-phrase appears to be repeated in an intermediate position. Consider the following cases, from Romani (54), Frisian (55), and Child English (56).[12]

(54)   **Kas**   misline   **kas** o Demiri dikhla?
       whom you-think who the Demir saw
       'Who do you think Demir saw?'

(55)   **Wer**   tinke jo   **wer**'t   Jan wennet?
       where think you where-that Jan lives
       'Where do you think that Jan lives?'

(56)   **Who** do you think **who** the cat chased?

I am not aware of any alternative analysis of the wh-copying data. It is indeed hard to think of one.

2.3   A-movement

Up to now the evidence adduced in support of the successive cyclicity hypothesis has come from the realm of wh-movement, the prototypical example of A-bar movement. This was the original domain for which Chomsky (1973) proposed successive cyclicity, given that A-bar movement appears to be able to span long distances. By contrast, A-movement seems[13] much more local, being confined to finite clauses:

(57)   John is smart.

(58)   John seems [*t* to be smart].

(59)   *John seems [*t* is smart].

For this reason, A-movement has been much less discussed with respect to successive cyclicity. (Some of the discussion in this section relies on the excellent overview of cyclicity in Lasnik 2006.)

Chomsky (1977:74) presents the following example, noting that "If the rule of NP-movement that yields [(60)] applies successive cyclically . . . then the rule will observe subjacency."

(60)   John seems [to be certain [to win]].

Chomsky (1981:44) discusses a similar example (61), with the representation in (62).

(61)   John was believed to have been killed.

(62)   John INFL be believed [S *t'* INFL have been killed *t*].

Certainly in the theory of subjacency of the 1970s, with S (=IP) as a bounding node, successive movement is forced upon any kind of movement. But in subsequent frameworks such as *Barriers* (Chomsky 1986b), IP lost its status as a bounding node: it could only become a bounding node or barrier by inheriting barrierhood from a dominating CP, which must typically be missing for A-movement to be long-distance.

Perhaps for this reason, several authors (Epstein and Seely 1999, 2006; Castillo et al. 1998; Stateva and Stepanov in press) have argued that A-movement proceeds in one fell swoop. In so doing these authors introduce an odd asymmetry between A- and A-bar movement, which I think should be resisted on minimalist grounds. If movement exists, it should be treated uniformly (with differences between A- and A-bar plausibly relegated to the interfaces). As we will see in subsequent chapters, successive cyclicity poses interesting problems for the minimalist program, but confining it to A-bar movement does not make the problem easier (Lasnik 2006 makes a similar point). Furthermore, to the extent that successive cyclic movement can be captured within minimalism, it is hard to see what would make it restricted to A-bar movement.[14]

With such conceptual considerations in mind, it is worth considering whether A-movement is in fact successive cyclic. Chomsky (1981: 44–5) presents an empirical argument for intermediate A-traces, hence, for successive cyclic A-movement. He observes that (63) is acceptable, indicating that Condition A is satisfied.

(63) they are likely [t' to appear to each other [t to be happy]]

On the basis of this, he argues for intermediate traces (hence, for successive cyclic movement). His line of reasoning is as follows: "The [position] filled by medial traces such as t' in [(62)] may . . . be relevant to LF; for example in the sentence [(63)], . . . where the medial trace serves as the antecedent of *each other*, which requires an antecedent in the same clause in such sentences in accordance with binding theory." Interestingly, as Lasnik (2006) observes, it actually isn't completely clear that the antecedent must be in the same clause, given the binding theory in Chomsky (1981) (or those in Chomsky 1973 and Chomsky 1986a, for that matter).

Consider the Chomsky (1981) formulation:

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

(65)   SUBJECT = AGR in a finite clause; NP of S in an infinitival; . . .

(66)   γ is accessible to α iff α is in the c-command domain of γ
         and . . .

If there is a trace in the intermediate clause then that clause is the
governing category (GC) of *each other*. But that is no argument
that there is such a trace. Suppose there were none. Then the inter-
mediate clause, lacking a SUBJECT, would not be the GC, and
Condition A would, correctly, be satisfied with *they* as the binder
of *each other*.
   The result is the same under the Chomsky (1986a) formulation.
Consider the portion of the theory relevant to Condition A:

(67)   I is BT-compatible with (α,β) if:
         (A) α is an anaphor and is bound in β under I.

(68)   Licensing condition for a category α governed by a lexical
         category γ in the expression E with indexing I:
         For some β such that [i], I is BT-compatible with (α,β):
         [i] α is an anaphor . . . and β is the least complete functional
         complex (CFC) containing γ for which there is an indexing
         J BT-compatible with (α,β).

As Chomsky puts it, "for an anaphor . . . the licensing condition
amounts to saying that the relevant governing category for α is the
minimal one in which binding theory could have been satisfied under
some indexing." Once again, if there is an intermediate trace, *each
other* must be (and is) bound in that clause. But in the absence of
that trace, the licensing condition for the anaphor could not, hence
need not, be satisfied in the intermediate clause. Therefore matrix
binding is allowed. So, the sentence provided by Chomsky does
not offer conclusive evidence for successive cyclicity in the A-
domain.
   Stronger evidence comes from data like (69), discussed in Lasnik
(2006) (where the example is attributed to Adolfo Ausín) and
Castillo et al. (1998) (where it is attributed to Danny Fox, via David
Pesetsky).

(69)   John appears to Mary [to seem to himself/*herself [to be
         the best candidate]].

In the absence of successive cyclic movement, *himself* would incorrectly be predicted to violate Condition A, and, on fairly standard assumptions, *herself* to be in conformity. By contrast, if successive cyclic movement takes place, there is a point where *John* is the closest c-commanding antecedent for the reflexive.[15]

(70)   John appears to Mary [<John> to seem to himself/*herself [to be the best candidate]].

Additional evidence for the successive cyclic nature of A-movement comes from complex reconstruction effects like (71) (due to Lebeaux 1991:234; see also Nunes 1995:200–2, and Bošković 2002:180).[16]

(71)   [His$_i$ mother$_j$'s bread] seems to every man$_i$ to be known by her$_j$ to be the best there is.

For all the relevant binding properties of this sentence to be adequately predicted there must be a copy of the A-moved subject *his mother's bread* located below *every man* (so as to license the variable *his*) and above *her* (so as to avoid a pronoun from c-commanding a co-referring name like *mother*, which would cause a Condition C violation).

In addition to interpretive reflexes of successive cyclicity in A-movement, quantifier float data like (72) are a parallel to the McCloskey (2000) data discussed above in the context of A-bar movement.

(72)   a.   All the boys seem to appear to like ice cream.
       b.   The boys seem *all* to appear to like ice cream.
       c.   The boys seem to appear *all* to like ice cream.
       d.   The boys seem to appear to *all* like ice cream.

If, as argued above, resumptive pronouns are taken to be stranded elements, then copy raising provides evidence for successive cyclic A-movement. (See Fujii 2006 for evidence that A-movement is indeed involved in copy raising.)

(73)   Jan sanble [li te   renmen Mari].   (Haitian creole)
       Jan seems he Pst love   Mari
       'Jan seems to have loved Mari.'

Interestingly, copy-raising examples seem to be sensitive to the shape of the complementizer (see Rooryck and Costa 1996; Fujii 2006), much like what we saw in the context of Irish A-bar dependencies.

(74) a. There seem like/*that there are problems here.
      b. John seems like/*that he is tired.

Examples like (75) (Chomsky 2001; Groat 1999; Seuren 2004, who attributes such examples to Paul Postal) may even be taken to parallel the wh-copying examples discussed above.

(75) There look like there are problems with this machine.

Additional morphological evidence for successive cyclic A-movement comes from agreement morphology that shows up on elements that stand in between the point of origin and the pronunciation site of A-movement. Witness the past participle agreement in French (76) and the agreement on aspectual auxiliaries in Bantu (77) (from Carstens 2001).

(76) Les chaises   ont   été repaintes.
      the chairs.fem.pl have.pl. been repainted.fem.pl
      'The chairs were repainted.'

(77) (Mimi) **ni**-li-kuwa   **ni** – ngali **ni** – ki – fanya kazi.
      (I)   1SG-PAST-be 1SG-still   1SG-PERF-do   work
      'I am still working.'

As in the case of A-bar movement, some of the evidence presented here may not be as unambiguous and compelling as one would wish, depending as it does on theory-internal considerations. But the point I want to make by listing all the evidence is that much of it parallels the evidence standardly taken to suggest that A-bar movement proceeds successive cyclically.

Unfortunately, the parallelism between A- and A-bar movement is not perfect. Successive cyclic A-movement does not seem to trigger any operation equivalent to subject–auxiliary inversion in intermediate landing sites. But this should not be taken as evidence against the successive cyclic nature of A-movement. It is simply a fact that A-movement, even short A-movement, does not trigger

inversion operations. Likewise, to date, I know of no phonological evidence for successive cyclic A-movement (unlike the Kikuyu data discussed above), but notice that phonological evidence in the A-bar domain is pretty thin too. It may be that such phonological reflexes exist in the A-domain as well, but they haven't been detected yet.

In sum, I see no reason to suppose that long-distance A-dependencies are different in kind from long-distance A-bar dependencies. Both kinds of dependencies appear to be formed by a succession of highly local steps.

2.4   Conclusion

I have reviewed a wealth of evidence for successive cyclicity in the A-bar and A-movement domains. It seems to me reasonable to conclude on the basis of this evidence that successive cyclicity exists; displacement in human languages is bounded. The issue now becomes not so much whether successive cyclicity exists, but why it exists, and how it should be captured.

McCloskey (2002:184–5) puts it best:

> If locality conditions are at the heart of syntax (as increasingly seems to be the case), then the existence of apparently unbounded dependencies like [long-distance wh-movement] represents an anomaly. Since Chomsky (1973), it has come to be widely believed that the apparently distant connection between antecedent and variable position in such cases is mediated by a sequence of more local connections. . . . A much harder question is what makes this true – what property of language-design determines that this is how things work.

In the following chapters I will try to answer the much harder *why*-question.

I will proceed by first tackling the *where*-question (where does movement stop? what counts as a possible intermediate landing site?) in chapter 3, then the *when*-question (when are intermediate steps of movement taken?) in chapter 4, and then the *why*-question (are successive steps of movement driven by any formal requirement?) in chapter 5. Answering these questions will, I hope, illuminate the

broader issue of what design feature of the language faculty is responsible for successive cyclicity.

Notes

1 Although there is mounting evidence that the syntax–PF mapping is more transparent and far less arbitrary than one might have thought, at least in the case of linearization (see Kayne 1994; Chomsky 1995; Nunes 2004) and lexicalization (Baker 1985; Halle and Marantz 1993; Marantz 2000; Borer 2005), there is no denying that the syntax–PF mapping is much less straightforward (at this stage of our understanding) than the syntax–LF mapping (on the syntax–LF mapping transparency, see Chomsky 2006, Hinzen 2006, and, especially, Uriagereka, forthcoming).

2 Spanish allows but does not require inversion in embedded contexts. Spanish is not unique in this respect. Many of the reflexes of successive cyclicity discussed in this chapter appear to be optional in various contexts. As this chapter focuses on establishing that successive cyclicity exists, I will not go into why the evidence does not always surface at this point.

3 I hasten to add that when I say that I find it hard to think of an alternative account for the iterative inversion facts just reviewed, I do not mean that it is hard for me to think of long-distance dependencies as not involving movement. I am aware of the fact that many frameworks such as LFG, HPSG, and others do not resort to movement mechanisms to capture long-distance dependencies. And, not surprisingly, such frameworks provide alternative analyses of the data at issue, but these are mere technical alternatives. They all share with the standard movement-based account the idea that iterative inversion is due to the fact that the configuration involving the wh-phrase in the matrix clause is somehow mirrored in the embedded clause.

4 There exist cases where a series of a^N-complementizers are found, as in (i).

> (i) an bhean a raibh mé ag súil a bhfaighinn uaithi   é
>   the woman a^N was I   hoping a^N get   from-her it
>   'the woman that I was hoping that I would get it from her'

5 Incidentally, mixed patterns like (18) and (19) strongly suggest that, contrary to standard accounts, long-distance dependencies with resumptive pronouns involve movement. I have explored the consequences of this claim at great length in Boeckx (2003a). See also chapter 7.

6 Bruening (2001:205ff.) discusses data from Passamaquoddy that behave similarly. Agreement triggered by successive cyclic wh-movement also appears to obtain in French past participle constructions (see Kayne 1989). In this case, the agreement is not on the complementizer, but on the participle, which is standardly taken to indicate a low intermediate landing site, a fact that will be highly relevant to issues discussed in the next chapter.

    (i)   Jean a   repeint/*repeintes   les chaises.
           Jean has repainted/repainted.fem.pl the chairs.fem.pl
           'Jean repainted the chairs.'

    (ii)   Quelles chaises   Jean-a-t'il   repeintes?
           which   chairs.fem.pl Jean-has-he repainted.fem.pl
           'Which chairs did Jean repaint?'

7 This is unlike Buli (Hiraiwa 2005), where complementizer agreement is limited to the complementizer adjacent to the wh-phrase.

8 In Hausa (Tuller 1985; Haïk 1990), extraction forces the verb to take on a special modal form. Only a few speakers allow this change to affect all verbs along the extraction path. Most speakers restrict this morphological change to the clause in which the wh-phrase is pronounced. A similar restriction holds for Moore (Haïk et al. 1985; Haïk 1990), and Old Japanese (the phenomenon on Kakarimusubi discussed in Watanabe 2002). For additional examples of wh-agreement, see Reintges et al. (2006).

9 I put the term "agreement" in quotes, as it is not at all clear that the term is appropriate to describe the relevant phenomenon. That some featural relation is involved is quite clear from the morphology of the language, but exactly which relation is unclear. In the main text I will keep using the now traditional term "(wh-)agreement", but data from Selayarese suggest otherwise. Selayarese, as described in Finer (1997) and analyzed in Boeckx (2003a), appears to be the flipside of the Chamorro pattern in forcing the otherwise obligatory agreement between a verb and its clausal argument to disappear if the clausal argument contains a gap, and forcing the complementizer to be dropped.

    (i)   Ku-isseʔ-*(i) *(kuko) la-ʔalle-i doeʔ-iñjo i Basoʔ.
           1s-know-3   COMP 3-take-3 money-the h Baso
           'I know that Baso took the money.'

    (ii)   Apa mu-isseʔ   la-ʔalle _ i Baso?
           what 2FAM-know 3-take   h Baso
           'What do you know that Baso took?'

A somewhat similar effect may obtain in Bahasa Melayu (colloquial Singapore Malay), where the transitivity marker *meN* must be dropped in on any verb along the extraction path of a wh-phrase (see Cole and Hermon 1998).

(iii) a. Apa Ali (*mem)-beri *t* pada Fatimah?
            what Ali (*meN)-gave to   Fatimah
            'What did Ali give to Fatimah?'
       b. Siapa Bill (*mem)-beritahu ibunya   [CP yang *t*
            who   Bill (*meN)-tell   his.mother that
            (men)-yintai Fatimah]?
            (meN)-love Fatimah
            'Who does Bill tell his mother that loves Fatimah?'

Obviously, one would like to offer a unified analysis for Chamorro and Selayarese. For suggestions on how to proceed toward such unification, see Boeckx (2003a) and chapter 7.

10 I will discuss Rackowski and Richards's analysis in more detail in chapter 6. I want to stress at this stage that I am presenting only one aspect of their analysis in this chapter. The use I make of this aspect of their analysis does not reflect their take on successive cyclicity.

11 For additional evidence for successive cyclic movement based on scope considerations, see Fox (2000:62ff.).

12 Wh-copying is to be distinguished from scope-marking strategies (see Bruening 2006), illustrated in (i)–(ii).

(i)   Was glaubst du, mit wem er gesprochen hat? German
      what think   you with whom he spoken   has
      'With whom do you think that he spoke?'

(ii) Mit   gondolsz, hogy kit   látott
      what.ACC you.think that   who.ACC saw.3sg
      János?   Hungarian
      János. NOM
      'Who do you think that Janos saw?'

It was once popular to analyze scope-marking as partial wh-movement; that is, movement that stops short of reaching its final landing site (filled by the scope-marker). Such an analysis provided rather strong evidence for successive cyclic movement, as, under such an analysis, the wh-phrases in (i) and (ii) shows up "stuck" in an intermediate position).

However, there is mounting evidence that the partial movement analysis is incorrect (see Lutz et al. 2000), and that the pronunciation

site of the wh-phrase is the final landing site (the scope-marker being associated with the entire CP, not with the wh-phrase itself). (The position of the scope-marker in Sinhala discussed in Hagstrom 1998 and Kishimoto 2005 appears to be amenable to an analysis similar to that of wh-scope marking.)

If, however, partial wh-movement is the right analysis in at least some cases of scope-marking, as Bruening (2006) has recently argued on the basis of Passamaquoddy data, then the evidence for successive cyclic movement provided by wh-scope-marking will be as strong as in the case of actual wh-copying.

13 I use "seems" here because the copy-raising examples discussed below, if an A-movement analysis is the correct one for them, show that A-movement can extend beyond finite clauses.

14 This is especially true for Epstein and Seely (1999, 2006) and Castillo et al. (1998), who argue against successive cyclic A-movement by criticizing notions like chains and EPP-features that are equally involved in A-bar movement contexts.

Stateva and Stepanov (in press) attempt to deduce the lack of successive cyclic A-movement by arguing that successive cyclic A-bar movement is actually a sort of residual wh-scope-marking. Since Stateva and Stepanov take scope-marking to be restricted to A-bar dependencies, it follows that successive cyclicity will be restricted to the A-bar domain. However, if Williams (1983) is correct in treating expletives like *there* as scope-marker, Stateva and Stepanov's deduction does not go through. This is probably a good thing on empirical grounds, since, as we will see in the text, A-movement shows reflexes of successive cyclicity.

15 Note that (for poorly understood reasons, but see Kitahara 1997, Boeckx 1999) experiencer NPs can bind outside of their PP-domains, as demonstrated in (i). (i) is straightforwardly excluded as a Condition C violation if *her* c-commands *Mary*.

(i)   *John seems to her$_i$ to like Mary$_i$.

So, the binding possibilities in (70) cannot be dismissed on the grounds that the experiencer does not count as an accessible (c-commanding) antecedent. Examples like (ii) can be explained by appealing to the copy of *John* left by movement.

(ii)   John seems to Mary [<John>] to like himself/*herself.

16 For arguments of the same ilk, see Sauerland (2003) and Fox (2000:145–7).

Chapter 3

The Distribution of Intermediate Landing Sites (The *Where*-Question)

Having established that successive cyclicity is true of human languages, let us now examine the precise mechanisms that conspire to yield successive cyclic steps in situations of long-distance dependencies.

The first question I would like to address is how the identity of intermediate landing sites is determined. Ideally we would like more than a mere list of landing sites gathered on the basis of detailed empirical work. One would like to understand what the defining features of intermediate landing sites are. Furthermore, even if we were to restrict our ambition to providing a detailed list of intermediate landing sites, the task would be far from trivial.

In Chomsky (1973), where successive cyclic movement was first proposed, successive cyclicity was forced upon movement by subjacency. Since S (= IP) and NP were the hypothesized bounding nodes and movement was allowed to cross only one bounding node, movement had to stop right at the edge of each bounding node it encountered to avoid violating subjacency. Accordingly, the identity of intermediate landing sites, S and NP, was easy to determine. In those days, however, the number of distinct heads and associated projections wasn't as great as it is now. If, say, Cinque (1999) is right about there being 30 distinct heads in the IP-domain, and Rizzi (1997) about there being five or six distinct projections in the CP-field, it won't be easy to determine which of these many positions are targeted by movement.

It seems to me that two very general options are available to us:

- Option I: movement proceeds through all available sites that separate the original position of the moving element and its ultimate landing site.
- Option II: movement proceeds though only some designated sites that separate the original position of the moving element and its ultimate landing site.

It is the purpose of the present chapter to weigh the evidence in favor of these two options.

3.1   Punctuated vs. Uniform Paths

Addressing the very issue just raised, Abels (2003) refers to Option I as the creation of uniform (or quasi-uniform) paths, and to Option II as the creation of punctuated paths. I will adopt this terminology in what follows.

Many theories of linguistic competence outside Chomskyan circles, especially those that capture long-distance dependencies by means other than movement, such as GPSG, HPSG, Categorial Grammar, Simpler Syntax, and LFG, conceive of extraction as a relation between a base-generated filler and a base-generated gap that is mediated by a set of nodes all of which record the information that the relevant structure contains a gap that needs to be filled by an antecedent of a given type (some authors, such as Koster 2003, Neeleman and van de Koot 2002, have pursued this idea in a mainstream generative setting). Long-distance dependencies, then, are computed as part of a projection, or percolation mechanism. The most detailed version of this mechanism is the slash-category formalism of GPSG and HPSG. According to this formalism, a slash-category XP/t_{YP} designates an XP which dominates somewhere within it a trace (or gap) of category YP. The trace, which has syntactic features [YP,t], is the lowest slash-category in the tree that satisfies or saturates the slash. All the constituents dominating it up to the filler are also slash-categories. These intermediate slash-categories capture the essential properties of successive cyclic

movement and are basically notational variants of intermediate land-
ing sites.

The end result of the percolation mechanism based on slash-
categories is the creation of highly uniform paths, as schematized here
for a sentence like *who do you think that Bill bought a picture of in
Chicago*. Notice that there is no skipping of nodes between the filler
and the gap.
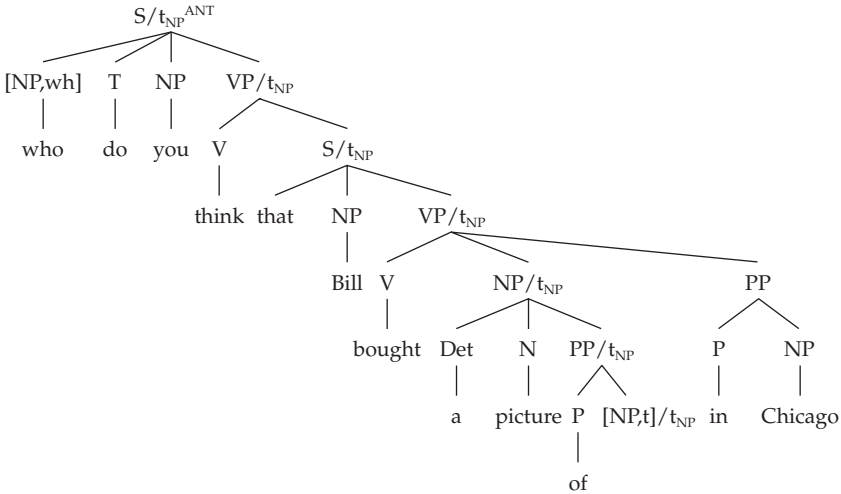
(Culicover and Jackendoff 2005:331)

Some mainstream generative studies, especially within minimal-
ism, have also formulated the idea that successive cyclic movement
proceeds through virtually all intermediate projections separating the
moved element's point of origin and its ultimate landing site. The
idea is explicit in Fox (2000), Boeckx (2001a, 2003a), N. Richards (2002),
Müller (2004), and Bošković (2002, 2006). (The spirit of Chomsky
and Lasnik 1993, Manzini 1994, Fox and Lasnik 2003, and Epstein and
Seely 2006 is very much in line with this idea as well. Takahashi
1994 is repeatedly cited in this connection, although his proposal
explicitly refers to different landing sites depending on the type of
movement involved, and as such it is much harder to categorize in
terms of Options I and II being considered here. Sportiche 1989 is
close in spirit to Takahashi 1994.)

As Abels notes, all these works argue for the creation of quasi-uniform chains. The difference between uniform and quasi-uniform chains is the following. Under a percolation mechanism like the one pursued in HPSG, *all* nodes pass on the slash-category information. This means that in a full-blown X-bar structure, with binary branching, X′ and XP will contain the relevant information; that is, X′ and XP will be intermediate "landing sites." This is not the case in the minimalist analyses just mentioned, where movement proceeds via adjunction (sister to X′) or substitution into SpecXP.

The difference between uniform and quasi-uniform chains is very subtle, and, as Abels notes, it is hard to think of an empirical test with sufficiently high resolution to distinguish between the two. And, finally, most non-mainstream analyses do not make use of binary branching, so only XP would contain the relevant information. I therefore conclude that the difference between uniform and quasi-uniform paths is negligible (but see chapter 5 for a possible argument in favor of quasi-uniform path theories). From here onwards, I will regard all the analyses mentioned in the context of uniform and quasi-uniform paths as fundamentally equivalent.[1]

Several authors (see Fox 2000; Boeckx 2003a; Bošković 2002) have claimed that path uniformity is the "simplest" option. The idea behind this claim is that any available projection is a potential landing site for long-distance movement. No special mechanism is needed to single out some sites over others. This is a very different conception from the one that underlies the punctuated path hypothesis. In the latter case, a given projection is not a potential landing site for long-distance movement, unless that projection is a bounding node (alternatively, a barrier or a phase). The punctuated path hypothesis requires a theory of what counts as a bounding node/barrier/phase. The uniform path hypothesis does not.

The intuition behind the uniform path view is that successive cyclicity is inherent to movement. That is, movement can't fail to proceed successive cyclically. (I will come back to this point in subsequent chapters.) By contrast, under the punctuated path view, movement takes place in one fell swoop unless there is an intervening obstacle (bounding node/barrier/phase), in which case movement will have to be interrupted, and then will have to resume.

The most pressing question that arises under the punctuated path view, but not under the uniform path view, is what makes certain nodes special. Why does CP count as a bounding node (barrier/phase), but AP doesn't? This is a question that the uniform path view does not have to answer. I see this as a big conceptual advantage because the question has been particularly hard to answer ever since it emerged in Chomsky (1973). In fact, I will argue in this chapter that it has yet to receive a satisfactory answer. Absent such an answer, the uniform path view is to be favored from a minimalist perspective.

I should note that although the question of what counts as a bounding node was present from the very minute the idea of bounding emerged, it was not as pressing a question as it is now in the context of the minimalist program. The ambition of the minimalist program is to find out why the language faculty is the way it is. Although it may well turn out that the answer will be found in the vagaries of chance, in accidents that by nature give rise to arbitrary outcomes, minimalism proceeds on the assumption that the answer to the *why*-questions will be deep, possibly the result of computational laws at work in complex systems ("the strong minimalist thesis").

3.2   The Difficulties Faced by Punctuated Path Hypotheses

In recent years, Chomsky has elaborated a version of the punctuated path hypothesis under the rubric of "phases." Because phase-based computations have received a lot of attention in recent years, let me examine the concept of phase more closely.

I will first briefly review what it means for derivations to be phase-based. Then I will review the empirical and conceptual arguments for phases and their specific properties that are currently entertained, pointing out parallels with earlier notions like bounding node or barrier. Ultimately, building on Boeckx and Grohmann (in press), I will show that none of the arguments for phases actually goes through, due in large part to the fact that proponents of phases have borrowed mechanisms from previous frameworks that remain stipulative.
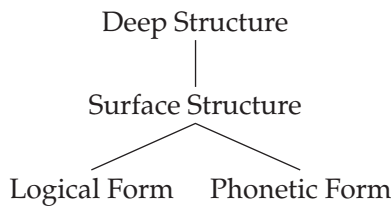
3.2.1 *Phases: an overview*

Chomsky (2000) argued for a more radical departure from the standard model of grammar inherited from the Extended Standard Theory and Government-and-Binding eras than the position he took in the original minimalist paper (Chomsky 1993) and the early stage of minimalist research (Chomsky 1995). Chomsky proposed that syntactic derivations proceed in incremental chunks, called *phases*, with each phase built from a separate lexical sub-array.[2] According to Chomsky, the phases are *v* and C. V and T aren't phases, and other categories are not discussed (although Chomsky 2006:17–18 suggests that D may also be a phase, following work by Svenonius 2004).
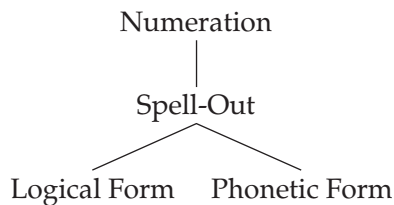
The general idea behind phases is that once these domains have been built from a lexical sub-array, much of their content is immediately transferred to the interfaces and can therefore be "set aside" for computational purposes, thereby alleviating the burden imposed on the computational system. Accordingly, instead of interfacing with sound/sign and thought systems only once, via levels like LF and PF, as in the standard Y-model, phases enable multiple access to external systems.
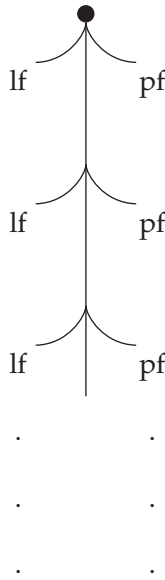
Here are schemas of the models:

Standard Y-model

```
              Deep Structure
                    |
            Surface Structure
                  /     \
        Logical Form    Phonetic Form
```

Minimalist Y-model

```
               Numeration
                    |
               Spell-Out
                  /     \
        Logical Form    Phonetic Form
```

Multiple Interface Access model

lf   pf

lf   pf

lf   pf

.   .

.   .

.   .

   In the context of phases, Chomsky (2001) distinguishes between the phase complement (sister of the phase head), the phase head itself (*v* or C), and the edge of the phase (specifier domain).

Phase

XP

YP

X°   WP ——— Phase complement

Phase edge
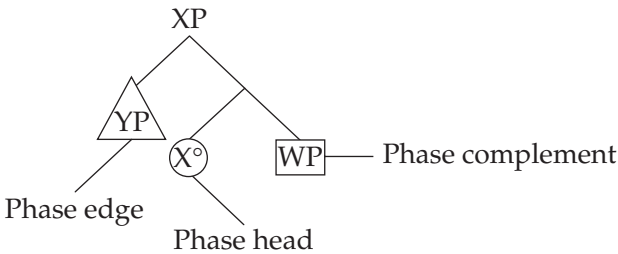
Phase head

With such notions in place, Chomsky formulates the *Phase Impenetrability Condition* (PIC), which says that once a phase has been completed, the internal domain of a phase (i.e., the complement of the phase head) is transferred to the interfaces, and thereby becomes inaccessible to operations outside the phase.[3] This in effect means that the edge of the phase and the phase head remain accessible to

material outside the phase (until the next higher phase is completed, according to Chomsky 2001). If nothing else were added, complements would not be available for extraction – clearly an undesirable outcome empirically. Accordingly, Chomsky proposes that material inside the internal domain of a phase can bypass the effects of PIC by moving to the edge of the phase via "indirect feature driven movement." This movement is triggered by a "P(eripheral)-feature" in Chomsky (2000), a "generalized EPP" in Chomsky (2001, 2004a), and an "edge feature" in Chomsky (2005). (It is not entirely clear whether the term "feature" used here is a feature in the technical sense of the term as used in minimalism – i.e., a feature subject to checking/valuation – or whether the term is simply taken to be a grammatical property of some sort, such as a licensing condition that a specifier be filled – a criterion in Rizzi's 1996, 1997, 2006 sense. I will come back to this point later in this volume.) In other words, the system allows the edge of a phase to act as an escape hatch for material that would otherwise be trapped inside the phase due to PIC. This designated escape hatch strategy immediately yields a version of successive cyclic movement, as it ensures that long movement proceeds phase edge by phase edge. In fact, due to PIC, movement is forced to be successive cyclic – in a specific sense: movement forms a punctuated path, where the "periods" are defined by the phases along the way.

3.2.2   *Conceptual arguments for phases*

As Chomsky repeatedly points out in his recent writings (see Chomsky 2000 through 2006), with phases, syntactic computation gains in cyclic character. The alleged computational cost of carrying the entire derivation up to the end, as in previous minimalist incarnations, is reduced by transferring portions of the derivation step by step to the interfaces.

   Computational cost reduction is the prime conceptual advantage and motivation for phases (see Chomsky 2000:110–12). More recently, Chomsky has also noted that the adoption of a phase-based derivation leads to a drastic reduction in the number of independent cycles that have to be assumed to be part of the computational system (see Chomsky 2004b:151, 2005:11, to appear:4). He says:

   [In GB theory] there were three, four, I guess five relatively independent generative systems, each of them essentially a cycle, all doing

the same thing, but operating separately. There was one that formed D-structure by X-bar Theory, which is basically cyclic. There's the transformational cycle, which is mapping D-structure to S-structure. There's a covert transformational cycle, which is mapping S-structure to LF, with the same kinds of transformations and also cyclic. And there's something mapping LF over to the semantic interface, a compositional semantics of some kind and, again, cyclic. And there's something mapping S-structure to PF, which is also basically cyclic. (Chomsky 2004b:151)

Chomsky has here identified a redundant feature of the computational system: different components cover the same space in the same fashion. Chomsky claims that a phase-based derivation, with its cyclic transfer property, allows one to contemplate a single-cycle architecture. Previous minimalist efforts have eliminated D-structure and S-structure (Chomsky 1993) as well as X-bar Theory (Chomsky 1994, 1995). Phases allow one to go one step further and eliminate LF by mapping chunks of syntax directly to the interfaces, providing for a less redundant, computationally more efficient system.

Chomsky provides another argument for phases, specifically for cyclic transfer, on the basis of considerations of legibility at the interfaces. Chomsky (1995) takes it to be a fact that there are interpretable and uninterpretable features on lexical items. As Brody (2003) has noted, interpretable features are canonically interpreted on the lexical items that carry them. Uninterpretable features typically aren't. As a matter of fact, Chomsky argues that considerations of Full Interpretation dictate that uninterpretable features be absent at the interfaces. That is, uninterpretable features are the engine of the derivation; their elimination prior to reaching the interfaces is what drives the computation. Chomsky assumes that the operation Spell-Out strips away features from the syntax and transfers them to the interfaces.

A look-ahead problem arises in this context, as Spell-Out must know which features are uninterpretable in order to strip them away and prevent the derivation from crashing at the interfaces. But in order for Spell-Out to know which features are uninterpretable, interpretation must have taken place. To avoid this look-ahead problem, Chomsky (2000) introduces a further distinction within features: valuation – features are either valued or unvalued. This distinction is assumed to be available to Spell-Out by inspection without recourse to the interfaces. Chomsky furthermore claims that there

is a one-to-one correlation in the lexicon between valued/unvalued and interpretable/uninterpretable. This correlation offers a way for Spell-Out to target the right features (uninterpretable ones), while being blind to interpretive considerations, that is, by targeting unvalued features.

At this point, a timing paradox arises. (For extensive discussion of this timing paradox, see Epstein and Seely 2002.) It is the business of the syntactic operation Agree, which I introduced in the previous chapter, to value lexically unvalued features in the course of the derivation. But once unvalued features have been valued, how can Spell-Out distinguish between valued and unvalued features, and transfer the right ones to the interfaces? Chomsky's solution to this timing issue is to have Spell-Out apply right at the moment when valuation takes place, when the valued/unvalued distinction is still visible. Spell-Out must therefore be cyclic, and apply each time valuation takes place, which Chomsky assumes is at the phase level.

3.2.3   *Arguments against phases*

Although I think that Chomsky has identified important issues in the context of phases such as computational efficiency, redundancy, locality, and legibility of features, I don't think that phases as currently conceived offer insightful ways of solving these problems.

Consider the memory load reduction argument. Bearing in mind the possibility that "our initial intuitions about perfection vs. imperfection (in this technical sense [of computational efficiency]) are not fully reliable" (Kayne 2004:5 n. 4), while granting Chomsky the possibility that cyclic transfer reduces computational load, it is not clear at all that phases as defined by Chomsky and currently used in the literature achieve the desired reduction. As Bouchard (2002:343) notes, it is still the case that the computational system must be able to retrieve "previously spelled-out material" to provide a complete, coherent surface string – so it can't just forget about spelled-out elements.

One may say that cyclic Spell-Out provides a way of forgetting the internal content of spelled-out element, but even that is not clear. As Norbert Hornstein (p.c.) has pointed out, interfaces appear to examine the internal content of full representations for specific processes. For example, the semantic component needs to see multiple

spelled-out chunks for pronominal binding. And PF quite possibly needs full clauses to determine intonational patterns, such as falling intonation ending up with a declarative (statement) or rising intonation yielding an interrogative (echo question) even for a simple expression like *John likes Mary* vs. *John likes Mary?* Arguably, it may still be possible in the latter case to just look at spelled-out domains as opaque chunks or "giant compound" words, as Uriagereka (1999a) has argued, but questions arise as to why the recombination of these chunks happens the way it does. I am not asking simply how this recombination takes place – Chomsky may be right in thinking that some such recombination is "easy enough to formalize" (p.c. to Dennis Ott, cited in Ott 2005). I am asking a more minimalist question: why the recombination takes place, and why it takes place the way it does. Notice that the problem is far from trivial. Recombination cannot amount to the cumulative outcome of cyclic Spell-Out because syntactic derivations proceed in parallel. Some algorithm will therefore have to be found, and, more importantly from a minimalist perspective, justified. It is unclear to me what optimal justification could be advanced here. I know of none in the current minimalist literature.[4]

The recombination problem carries over to the redundancy of levels of representation advanced by Chomsky in favor of phases. As the term "recombination" makes clear, it is still the case that the interfaces will do some combinatorial work that will no doubt mirror, or at least follow very closely, the work that the syntactic component performed – just as LF cycled over the same chunks of structure that, say, S-structure did. It appears that this kind of redundancy cannot easily be eliminated. Phases merely appear to hand over the redundancy to other components.

Turning now to the argument for phases based on the valuation of features, I want to start by noting that the argument, even if it goes through, again rests on assumptions that are questionable. Specifically, it rests on the claim that valuation and interpretability go hand in hand. As Pesetsky and Torrego (2004) have pointed out (building on work by Brody 1997 on the non-existence of genuinely uninterpretable features), the biconditional is odd and unlikely to be true. Why should the lexicon couple the disparate properties of interpretability ("Does the item have a message to send to the semantics?") and valuation ("Are any syntactically relevant properties of the lexical item left unspecified?")?

Furthermore, as Epstein and Seely (2002, 2006) argue at length, it is not clear that the mechanics Chomsky resorts to achieves the right result. As Epstein and Seely show, there is no point in the derivation where Spell-Out can distinguish between valued and unvalued features without either look-ahead or backtracking. In the absence of such undesirable computational strategies, Spell-Out will always be "too early" or "too late."

To conclude, none of the conceptual motivations for phases is compelling. This is very significant for our purposes, for recall that the punctuated path option bears the burden of proof.

3.2.4   *Old problems for phases*

The previous section has argued that it is not so easy to motivate the need for *some* phases. This section goes on to argue that, even if we could motivate the need for some phases in some fashion, it is less clear that one can motivate their current properties (phase edges, PIC, etc.) in a non-arbitrary fashion, as minimalism would like. The problem here revolves around two related questions, crucial in the context of successive cyclicity.

(i)   What exactly is the identity of phases?
(ii)   What exactly are the properties of phases?

Chomsky (2000) argues for C° and v° being phase heads on the basis of the following considerations: he claims that unlike T° and V°, C° and v° are isolable, coherent units at the interfaces. On the PF-side, they can occur in fragment answers, be clefted, etc. On the LF-side, they form full-blown, thematic and discourse, propositional units.

Legate (2003) observes that some of the tests used by Chomsky to justify his characterization of C and v, to the exclusion of V and T, as phases fail to make the relevant cuts. In particular, she shows that raising and passive verbs pass three diagnostics for phasehood argued for in the literature in the same way transitive verbs do: (i) they show instances of reconstruction of *wh*-phrases at their edge; (ii) they provide a target for QR; and (iii) they license parasitic gaps at their edge. (In addition, Legate shows that passive and raising verbs are as PF-isolable as full-fledged transitive verbs.)

These tests suggest, contra Chomsky's claims, that raising and passive verbs are phases too. To deal with Legate's findings, and while at the same time maintaining that transitive v is special, Chomsky (2001) makes a distinction between weak phases (raising/passive verbs) and strong phases (his original phases, C and v). As far as I can tell, this distinction plays no role in the theory, except that it restricts phase properties to strong phases. Weak phases act as if they weren't phases; in particular, they don't count as domains for the application of Spell-Out or PIC.

Matushansky (2005) casts further doubt on Chomsky's criteria for phasehood by looking at the behavior of DPs. She shows, on the basis of the various tests offered in the literature, that DPs would count as phases on the basis of PF-considerations, but would not count as phases on the basis of LF- (and purely syntactic) diagnostics. Moreover, and not surprisingly, several authors have explored the claim that the identity of phases may be parametrized across languages (see Gallego 2006, building on work by Juan Uriagereka) or that some domains may count as phases for LF but not for PF, or vice versa (Marušič 2005).

Chomsky (2000:107–8) also argues that phases can be identified because they are isolable at the interfaces. But it is not clear why that is so. As Grohmann (2000, 2003a), Bošković (2002), Abels (2003), and Epstein (to appear) have independently noted, (strong) phases shouldn't be isolable at the interfaces if the mechanism of Spell-Out/Transfer assumed by Chomsky (2001, 2004a) is correct. According to Chomsky, once a phase is completed, its complement domain is transferred to the interface and frozen syntactically via PIC. If so, the edge (i.e., the head with all specifiers and adjuncts) and the complement of a phase are spelled out at different times! In other words, complements of phases should be isolable units at the sound and meaning sides.

In addition, Epstein (to appear:12) also notes that the specification of a phase as having "*full* argument structure" (the term used by Epstein as introduced in Chomsky 2004a, original emphasis) cannot mean that all relevant θ-roles are in fact discharged. This would have the unintended result that raising TPs as well as passive and unaccusative vPs, for example, are phases, since all θ-roles associated with the head *are* discharged. "So, full argument structure must be a *translexical* notion" (Epstein, to appear:12, original emphasis), clearly an undesirable conclusion.[5]

The upshot is that Chomsky's characterization of C and v as phases does not receive independent support from interface diagnostics, and boils down to a stipulation. A similar conclusion obtains when it comes to properties of phases: why is the edge accessible, where does the EPP-feature come from, what is being spelled out, etc.?

I do not think that these problems are details of implementation. I regard them as persistent problems, and very significant ones in the context of successive cyclicity. As I have already pointed out above, and as should be obvious to anyone who is familiar with the history of generative grammar, phases are in many ways reincarnations of bounding nodes and barriers. For example, both in a phase-based system and in *Barriers*, an element can be both inside and outside a given domain (defined in terms of barrier or phase). Barriers and phases are notions that inherently trap elements inside them; they impose a very strict locality on syntactic computation. But in both systems there exists a possibility of circumventing this locality by adjoining to a barrier-projection – or moving to an additional specifier slot created by the phase head (the generalized EPP-feature). In both cases, this additional adjunction/specifier position extends the syntactic life of a moving element. In *Barriers* (Chomsky 1986b:6–7), this was justified by redefining adjunction in the light of May's (1985) discussion on scope and *wh*/quantifier-interactions, and also based on Koopman and Sportiche's (1982) arguments from quantifier raising.

Another, related parallelism between barriers and phases touches on the notion of successive cyclic movement. If there is no phase (boundary) intervening between extraction site and landing site, there is no need for an intermediate touch-down. Movement can take place in one fell swoop. The same holds for the *Barriers*-framework, where clause-internal *wh*-movement of an object, for example, had to adjoin to VP (a barrier) prior to moving to SpecCP, while a VP-adjoined adverbial or the subject (from then SpecIP) were free to move in one go.

For illustration, the derivations for *Who did John kiss?* and *How did John kiss Mary?* in the two frameworks are provided below, indicating the parallelism between the barrier-defined framework (1a, 2a) and the phase-based system (1b, 2b).[6]

(1)   a. $[_{CP}$ who$_i$ did-C $[_{IP}$ John$_k$ I $[_{vP}$ t$_i$ $[_{VP}$ t$_k$ $[_{VP}$ kiss t$_i]]]]]$
      b. $[_{CP}$ who$_i$ did-C $[_{TP}$ John$_k$ T $[_{vP}$ t$_i$ $[_{vP}$ t$_k$ v $[_{VP}$ kiss t$_i]]]]]]$

(2) a. [$_{CP}$ how$_i$ did-C [$_{IP}$ John$_k$ I [$_{vP}$ [$_{VP}$ t$_k$ [$_{VP}$ kiss Mary]] t$_i$]]]
     b. [$_{CP}$ how$_i$ did-C [$_{TP}$ John$_k$ T [$_{vP}$ [$_{vP}$ t$_k$ v [$_{VP}$ kiss Mary]] t$_i$]]]

A third parallelism concerns the choice of barriers or phase heads. In both approaches, the relevant natural barriers/phase heads are C and v (topmost V in *Barriers*). In *Barriers*, IP is defined as a "defective" projection (Chomsky 1986b:15): it is not an inherent barrier and can only become one by inheritance. The same is true of the phase-based approach: T (I) is not a phase-inducing head and in this sense is defective. In other words, in both approaches, TP (IP) is special.

Yet another parallelism comes from the fact that, although in both *Barriers* and recent phase work Chomsky restricts his attention to clausal properties, Chomsky (1986b:80) opens the door for other barriers, such as AP. Incidentally, while painting a "simple" picture in which v and C are the only (strong) phase-inducing heads, Chomsky (2001, 2004b) alludes to the possibility that other heads may be phasal as well, such as D or P, and is even more explicit concerning D in Chomsky (2005, 2006).[7]

It can in fact be argued that the notion of phase reinstates the notion "Complete Functional Complex," the part of the derivation in which "all functional roles [are] satisfied" (Chomsky 1986a:15). For example, as we saw, for Chomsky (2000), v counts as a phase by virtue of being the domain where all theta-roles are assigned.[8]

The major condition on phases, PIC, has its antecedents. As Abels (2003) correctly points out, the following two definitions show that the current version of PIC is essentially a modern restatement of van Riemsdijk's (1978) Head Constraint.

(3) *Phase Impenetrability Condition* (Chomsky 2000:108)
     In phase α with head H, the domain of H [= complement of H] is not accessible to operations outside α [= HP], but only H and its edge [= H plus any/all of its specifiers].

(4) *Head Constraint* (van Riemsdijk 1978:169)
     No rule may involve X$_i$ (X$_j$) and Y in the structure . . . X$_i$ . . . [$_\alpha$ . . . Y . . .] . . . X$_j$ . . . if Y is c-commanded by the head of α; α ranges over V''', N''', A''', P'''.

The above parallelism establishes a close connection between phases and barriers.[9] As I have shown, virtually all the properties

ascribed to phases in the current literature have been recycled from the very first theoretical attempt to make sense of such phenomena as islands or successive cyclicity (Chomsky 1973). In many ways, phases are to minimalism what bounding nodes and barriers were to the Extended Standard Theory and Government-and-Binding Theory, respectively. In and of itself, the fact that phases have theoretical antecedents is not a bad thing. The "eliminative" or "therapeutic" effect of minimalism (Chomsky 1995:233–4) does not entail that theoretical linguistics must restart from scratch (Lasnik 1999a highlights some old features that have been made new again in minimalism, and have been put to good explanatory use), but it does entail that all the tools we make use of must be well motivated conceptually, and that their properties must follow from virtual conceptual necessity. I think that this is far from being the case when it comes to phases. Like bounding nodes and barriers, phases beg questions that lead to persistent problems. For instance, Lightfoot and Weinberg (1988) criticize Chomsky's (1986b) mechanism of adjunction to void or escape barrierhood as a stipulation (see also van de Koot 1989). This point certainly carries over to the use of EPP-features in a phase-based theory. Accordingly, phases do not enhance our understanding of syntactic phenomena like successive cyclic movement, and locality more generally. They simply recode insights from the past.

Maybe this is the best we can do at this point, but recall that the punctuated path option is not forced upon us. And the uniform path option does not have to answer questions about designated nodes, PIC, etc. So the real question at this stage, the only argument left for phases at this point, is: are there empirical arguments against the uniform path option, and for the punctuated path options? This is the question I examine in the next section.

Note, once again, that if the punctuated path option prevails, nothing more can be said about successive cyclicity until we can offer conceptual justifications for phases and their properties. Until then, successive cyclicity will be a brute-force phenomenon resisting minimalist analysis.

3.2.5   *No empirical argument for phases*

Chomsky (2000) provides one empirical argument for phases.[10] The argument consists in showing how phases solve a problem that arises

in the context of Chomsky's (1995) Merge-over-Move principle (henceforth, MOM; see Castillo et al. 1998 for extensive discussion). Consider the data in (5):

> (5)   a.   There seems to be a man in the garden.
>   b.   *There seems a man$_i$ to be t$_i$ in the garden.

If, as Chomsky assumes, there is a feature-checking requirement on the intermediate infinitival T, a question arises why the requirement cannot be checked by the movement of the indefinite (5b). Chomsky proposes an account of the impossibility of (5b) in terms of a preference for Merge over Move. According to Chomsky, at the point when the embedded clause is built we need to insert something into the infinitival SpecTP in order to satisfy the EPP. We have two possibilities for doing this in (5). We can either insert *there*, which is present in the numeration, into SpecTP or we can move the indefinite to this position. Chomsky argues that lexical insertion is a simpler operation than movement. Therefore, the possibility of expletive insertion into the embedded SpecTP, which for Chomsky takes place in (5a), blocks the indefinite movement to the embedded SpecTP, which takes place in (5b). Here Chomsky makes use of his definition of Move as a combination of Copy plus Merge to argue that Move is more complex than Merge, and therefore its use is restricted to the cases where simpler options are not available.

Although MOM accounts for (5), it faces a problem in contexts like (6) (as was observed independently by Juan Romero and Alec Marantz).

> (6)   There was a rumor that a man$_i$ was t$_i$ in the room.

Here, an indefinite NP has moved to SpecTP although an expletive was present in the numeration and available for lexical insertion.

To deal with this type of construction Chomsky (2000) introduces the concept of sub-numeration, defined over or equated with phasal units. More precisely, Chomsky proposes that each phase has its own sub-numeration or lexical sub-array (LA). Since the expletive is not present in the LA corresponding to the embedded clause (recall that CPs are phases), the option of expletive insertion is not available, and MOM is satisfied. Although the argument based on (6) is really an argument for LA, and not for phases *per se*,

Chomsky turns the argument into one for phases by associating LAs with phases: not only are phases points of cyclic transfer, they also constitute points of cyclic access to the lexicon.

When it comes to the MOM argument, as we noted above, Chomsky's reasoning is really about LAs. That the point(s) of access to the lexicon correspond(s) to points of cyclic transfer is an additional assumption that may be correct empirically, but one that appears to be arbitrary, hence suspect from a minimalist perspective. Furthermore, the line of argumentation crucially relies on the existence of a non-finite SpecTP, and the possibility of non-finite T bearing an EPP-feature.

One cannot fail to notice that in recent years, MOM has become quite suspect. This is in part due to the fact that, since MOM was proposed, the conception of Move has dramatically changed. In more recent theorizing, "Move" is just one of the forms of the basic operation Merge. As Chomsky states:

> NS is based on the free operation Merge. [The strong minimalist thesis] SMT entails that Merge of α, β is unconstrained, therefore either *external* or *internal*. Under external Merge, α and β are separate objects; under internal Merge, one is part of the other, and Merge yields the property of "displacement" (Chomsky 2001:7–8, emphasis in the original)

The difference between internal Merge and external Merge is now restricted to the relation between objects to be merged. Under this conception of Merge and "Move," it becomes less clear how the economy-based conceptual argument motivates MOM. Specifically, internal Merge ("Move"), as defined by Chomsky in the above quotation, does not appear to be more complex than external Merge any longer. Thus, Merge-over-Move loses its conceptual argument.

Empirically, Castillo et al. (1998) have denied the claim that deficient, non-finite T bears an EPP-feature (i.e., makes available a specifier position); see also Epstein and Seely (1999, 2006), Boeckx (2000b), Hornstein (2001), and Bošković (2002), among others. If non-finite T does not have an EPP-feature, its specifier need not be filled.[11] Accordingly, the derivation underlying the licit (5a) above is as in (7), without any relevant movement steps:

(7)   There seems [$_{TP}$ to be [a man in the garden]].

If true, this has eliminated the sole empirical data provided as evidence for MOM. If MOM does not have to be evoked, the Romero/Marantz problem doesn't arise, and the need for phases (*qua* cyclic access to the lexicon) is severely weakened.

Abels (2003) provides a different empirical argument for phases, and against the hypothesis that each maximal projection is targeted by successive cyclic movement. The argument directly addresses the central issue of this chapter, hence deserves close attention. Abels notes that much work on reconstruction effects has looked at where such effects arise to identify intermediate landing sites. Abels correctly notes that this does not distinguish between uniform and punctuated paths. To tease the two options apart, one must look at where reconstruction does not obtain. If reconstruction obtains everywhere, it is an argument for path uniformity. If it fails to obtain somewhere, it is an argument for punctuated paths.

Abels claims to have found one argument of the latter type. His argument rests on the following paradigm, specifically, on the contrast between (8b) and (8d).

(8)   a.   *John said that Sue likes pictures of himself.
      b.   Which pictures of himself did John say that Sue likes?
      c.   *Mary seems to John to like pictures of himself.
      d.   *Which pictures of himself does Mary seem to John to like?
      e.   Which pictures of himself does it seem to John that Mary likes?

(8a) is straightforwardly accounted for: the anaphor is too far away from its antecedent. The acceptability of (8b) suggests a reconstruction effect in an intermediate landing site, as we saw in the previous chapter when reviewing semantic evidence for successive cyclicity. The structure for the interpretation of (8b) would be as in (9).

(9)   Which pictures of himself did John say <which pictures of himself> that Sue likes <which pictures of himself>?

(8c) is excluded by the fact that the original copy of the raised subject *Mary* intervenes between the anaphor and its antecedent. Note that the sentence is not ruled out because the experiencer DP is

embedded inside a PP. The acceptability of (8e) indicates that experiencers of this kind can license anaphors (see Pesetsky 1995, Kitahara 1997, and Boeckx 1999 for extensive discussion of how to achieve the relevant c-command relation.)

The key example is (8d). On the basis of the unacceptability of this example, Abels argues that wh-movement does not target the non-finite SpecTP in (8d); if it did, the anaphor would be expected to be licensed, in a manner parallel to what we find in (8b) and (8e). Abels concludes that the lack of reconstruction (understood as the lack of an intermediate landing site) follows under a phase-based or punctuated path theory (non-finite T is not a phase, whereas finite C is), but doesn't under a theory that assumes that each maximal projection is targeted as an intermediate landing site (non-finite T would be targeted under a uniform path hypothesis, and the contrast between (8b,e) and (8d) would be unexpected).

The logic of Abels's argument is very clear. Unfortunately for the punctuated path hypothesis, it faces significant problems. As Norbert Hornstein points out (p.c.), cases like (10) cast doubt on the paradigm in (8).

(10)   *Which pictures of himself does Mary seem to Susan to have told John that she likes?

If the paradigm in (8) were robust, we would expect (10) to pattern on a par with (8e), not (8d).

But above and beyond empirical considerations, Abels's argument is based on a false premise (which it shares with Sportiche 2005; Aoun and Li 2003). Whereas the copy theory of movement readily accounts for reconstruction by involving the interpretation of unpronounced copies, we cannot conclude from this that if no reconstruction effect is found, no copy is available at the relevant site. All we can conclude from the absence of reconstruction is either that there is no copy present, or that a copy was created, but for some (perhaps interpretive) reason cannot be interpreted in the relevant position.

There is indeed ample empirical evidence against Abels's premise that if no reconstruction effect obtains, no copy was there. Consider, for example, the absence of reconstruction effects in various A-movement contexts recently discussed by Lasnik (1999b) (see also Boeckx 2001b). A relevant case is presented in (11).

(11)   No one is certain to solve the problem.

The sentence in (11) cannot be paraphrased as "it is certain that no one will solve the problem" (with the raised subject *no one* interpreted under the scope of *certain*). As Boeckx (2000c, 2001b) has argued (contra Manzini and Roussou 2000), anti-reconstruction effects like (11) cannot be analyzed in terms of non-movement, since some instances of A-movement (especially movement involving indefinites) reconstruct. If this conclusion is correct, Abels's premise cannot be right.

Another argument against Abels comes from the fact that even weak islands, those that are standardly taken to permit a significant degree of movement, do not accommodate reconstruction, as witnessed in (12).

(12)   a.   Which of his$_{1/*2}$ pictures did Bill$_1$ ask me why nobody/everybody$_2$ hated *t*?

   b.   Which of his$_{1/*2}$ pictures doesn't Bill$_1$ think that everyone$_2$ liked *t*?

In a similar vein, Bobaljik and Wurmbrand (2005) provide compelling evidence that the lack of reconstruction effects in restructuring contexts in German (and other languages), illustrated in (13), cannot be analyzed in terms of absence of movement/copying.

(13)   . . . weil   er alle Fenster   vergessen hat [<alle Fenster>
         because he all   windows forgotten has
      zu schliessen]
      to close
      '. . . because he forgot to close all the windows' (alle
      Fenster >> vergessen; *vergessen >> alle Fenster)

Additional arguments against equating lack of reconstruction with lack of movement/copying come from work by Nevins and Anand (2003), who argue that subjects which do not trigger agreement, such as ergative subjects in languages like Hindi, don't reconstruct for scope, whereas nominative subjects, which trigger agreement on the verb, may reconstruct and be outscoped by the object, as the following contrast illustrates.

(14) Kisii Saayer-ne har   ghazal likʰii.   (Hindi)
      some poet-ERG every song   write.f-Perf
      'Some poet wrote every song.' (some > every; *every > some)

(15) Koi Saayer   har   ghazal likʰtaa   hai.
      some poet-Nom every song   write.m-Impf be-Pres
      'Some poet writes every song.' (some > every; every > some)

In a similar vein, Boeckx and Hornstein (to appear), building
on theoretical claims in Boeckx (2003a) and empirical facts from
Lebanese Arabic discovered by Aoun and Li (2003), provide evid-
ence for wh-movement in resumptive structures like (17), as evid-
enced by the presence of superiority effects (signaling that the closest
element has moved) (16), despite the fact that reconstruction fails
to obtain in such contexts.

(16) a. Miin ʔənbasatto laʔinno saami ʕarraf-o
          who pleased.2pl because Sami introduced-him
          ʕa-miin?
          to-whom
          'Who were you pleased because Sami introduced (him)
          to whom?'
      b. *Miin ʔenbasatto laʔinno saami ʕarraf   miin
          who   pleased.2pl because Sami introduced whom
          ʕəl-e?
          to-him
          'Who were you pleased because Sami introduced who
          to him?'

(17) *ʔAyya taalib min   tulaab-a$_i$ ʔənbasatto laʔinno
      which   student among students-her pleased.2pl because
      kəll   mʕallme$_i$ ħatnaʔ-ii?
      every teacher.fs will.3fs.choose-him
      'Which of her$_i$ students were you pleased because every
      teacher$_i$ would choose (him)?'

To sum up, it appears perfectly reasonable to claim that recon-
struction effects signal movement (copying), and anti-reconstruction
effects do not necessarily signal lack of movement/copying.
Therefore, the data in (8), especially (8d), are silent on the issue of

uniform vs. punctuated path in the context of successive cyclicity. Of course, one would like to know why (8d) is unacceptable. But this is true also of all the anti-reconstruction effects just mentioned. I will not speculate here as to what the relevant factors may be. The point I want to make is that the unacceptability of (8d) is not incompatible with the uniform path hypothesis.

3.3   Conclusion

In light of all the problems with the best-worked-out version of punctuated paths (phase-based computation), and the absence of empirical arguments against alternative conceptions of successive cyclic movement, I conclude that the simplest assumption – i.e., that paths are (quasi-)uniform – is to be adopted.

Notes

1 Culicover and Jackendoff (2005) claim that an HPSG-style analysis in terms of uniform paths is better motivated on parsing considerations, but I fail to see why this is so (see N. Richards 2002 for arguments that a quasi-uniform path is well motivated for parsing). Since they do not provide any detail, I will ignore their claim here. (For arguments against deriving locality conditions in syntax from performance factors, see Lasnik 1999c; Phillips to appear.)

2 For related proposals, see e.g. Epstein et al. (1998), Uriagereka (1999a, 2003), Grohmann (2000, 2003a), Platzack (2001), McGinnis (2001), Johnson (2002), Bošković (2006), Fox and Pesetsky (2004), Wagner (2005), and Epstein and Seely (2006).

3 This definition is an updated version of PIC originally introduced in Chomsky (2000:108), following arguments put forward in Nissenbaum (2000).

4 For a valiant attempt, see M. Richards (2004). The attempt, however, fails, as it assumes the existence of phases in the first place, and tries to deduce their properties from first principles. But it is the very existence of phases that one must start by justifying. (Mayr 2006 makes a similar point, and goes on to show that Richards's proposals face some conceptual and empirical difficulties, which I won't review here.)

5 Epstein introduces this notion to express that one can't just look at the verb in the sentence to know whether it is a *v*-phase or not

– one has to look at the entire verbal paradigm (hence, *trans*-lexical): for example,

> in executing the derivation of passive, I look up the lexical entry of the active form, notice that the passive form has morphologically reduced theta assigning properties (contra Baker, Johnson, and Roberts 1989) as compared to active, then return to passive, categorize it as *not* full argument structure, and then assign (somehow) *no separate array* for this not-yet-constructed construction. (Epstein, to appear:12, slightly modified)

6   The *Barriers*-derivations are modeled on the discussions on p. 29 and p. 19 of Chomsky (1986b), respectively. For simplicity, I indicate all lower copies/traces as *t* and indicate the predicate-internal subject in Spec*v*P in representations of the barrier-defined framework as well.

7   For a systematic exploration of this possibility, see e.g. Abels (2003), Svenonius (2004), and Lee-Schoenfeld (2005).

8   Recent work has explored the possibility of a phase being the domain in which all relevant features have been checked (Felser 2003; Svenonius 2004).

9   Regarding locality, Chomsky (1986b) assumed two types: *barriers* (boiling down to subjacency: in the ideal case, movement does not cross any bounding node, or barrier) and *minimality* (a closer governor takes precedence over a potential governor further away). The latter condition was then relativized by Rizzi (1990), and relativized minimality has enjoyed tremendous acceptance all the way to the current minimalist approach(es). In fact, up to Chomsky (2000), this was the only valid or relevant condition on locality, usually taken to be some form of closest movement (Shortest Move of Chomsky 1993, Minimal Link Condition and Attract Closest of Chomsky 1995). However, as we have seen, PIC imposed on phases now offers an approach to CED phenomena (see Chomsky 2004a). As Norbert Hornstein has pointed out to me (p.c.), phases in this respect look more like rigid bounding nodes than barriers, which could acquire their "bounding node" status in the course of the derivation. (For recent work that suggests that phasehood may also be determined, at least in part, derivationally, see Bobaljik and Wurmbrand 2005 and Gallego 2006.)

10 In addition, Chomsky (2000) notes that PIC not only yields a version of successive cyclic movement as an immediate consequence, it also "suggests a new approach to some Empty Category Principle (ECP) issues, such as subject extraction" (Chomsky 2000:144 n. 46, crediting Idan Landau, p.c.), an approach explored in Chomsky (2005).

Here too, it is far from clear that phases achieve the right results. The problem with CED effects in a phase-based system is made clear in Ceplova (2001).

> In the current theory [Chomsky (2001)], all phase-boundary-inducing heads can have P-features. A head with a P-feature can attract elements with unsatisfied uninterpretable features to its specifier, with the result that the P-feature is checked by the attractee, and the attractee is in a position from which it can move further to satisfy its uninterpretable feature (and thus prevent the derivation from crashing). The *problem* that arises by this proposal is that now *nothing should be an island* if all strong phases allow movement out of them (due to P-features). (Ceplova 2001:2–3; emphasis mine)

Faced with such a situation, Ceplova (and many others before her; cf. Chomsky 1986b) investigates "a possibility of restricting the distribution of P-features that depends on structural position of the category, a possibility reminiscent of L-marking in Chomsky (1986[b])". I know of no non-arbitrary way of doing that.

In addition to this conceptual problem, Hiraiwa (2003), Boeckx (2003a), Jeong (2006), and Broekhuis (2006) have identified technical problems for a phase-based account of locality. Not surprisingly, even people who assume some notion of phase argue it has nothing to do with locality (see e.g. Bošković 2006; Fox and Pesetsky 2004; Boeckx 2006b).

11 I depart here from Castillo et al. (1998) (as well as Hornstein 2001, Epstein and Seely 2002) in using the weaker term "need not" as opposed to "is not," as alternative conceptions of successive cyclic movement and expletive insertion may allow for non finite T to be filled (see e.g. Boeckx 2000b, Grohmann 2003b, and Bošković 2002, 2006).

Chapter 4

The Timing of Intermediate Steps of Movement (The *When*-Question)

4.1 Early vs. Late Successive Cyclicity

The present chapter focuses on one question that many may well regard as either misplaced, or obviously settled: when are intermediate steps of movement taken? Obviously, I intend to show that the question is meaningful, and far from settled. I also intend to provide (novel) empirical arguments that bear directly on the issue and help decide how to answer the question.

I should point out right away that the question of when intermediate movement steps are taken only makes sense in a derivational context, where syntactic operations are temporally ordered. The question is meaningless in a representational framework of the type advocated in Brody (1995, 2003). The majority of minimalist investigations (the present work included), however, assume a derivational model. But even in derivational frameworks the question of when intermediate steps of movement are taken is rarely asked. In a framework where trees are built from bottom to top in a step-by-step fashion, as standardly done in minimalism, it is equally standardly, and quite reasonably, assumed that intermediate steps of movement are taken before the final landing site of movement is reached, as soon as intermediate landing sites are created. The assumption, I suspect, tracks the working of what Chomsky (1993) calls the Extension Condition, the idea that syntactic material may be added only at the root (topmost position) of the tree, not inside the already constructed structure, as schematized in (1)–(2).

(1)

OK

(2)

Not OK
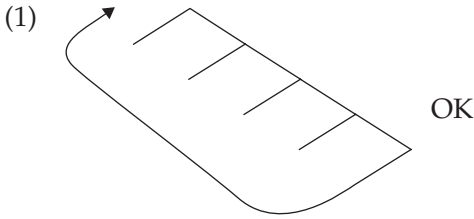
If the Extension Condition is adopted, intermediate steps of movement must be taken as the syntactic structure is being built. The creation of intermediate movement steps cannot be delayed, as these would force a merge operation that would fail to extend the tree, as schematized in (3). (The numbers in brackets refer to the timing of operations.)

(3)   Insert final landing site (#1)

(#3)

Intermediate movement step (#2)
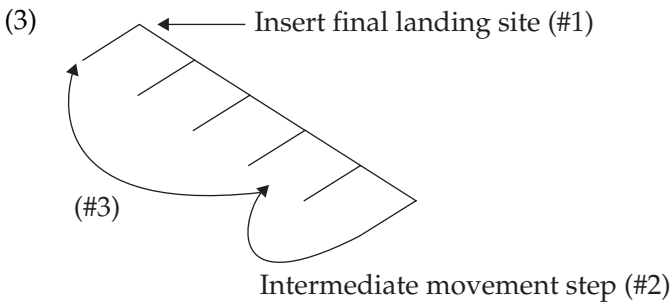
Since the Extension Condition appears to be a fairly natural condition in a minimalist context (see Chomsky 2005; Hornstein 2005), and has been widely adopted, it is no surprise that linguists would consider the timing issue at the heart of this chapter a non-issue – except in the literature focusing specifically on the nature of successive cyclicity. There, the timing issue is far from having been trivially settled. As a matter of fact, the best-worked-out version of successive cyclic movement, going back to Takahashi (1994), argues that

intermediate steps of movement are formed at a fairly late stage in the derivation, not before the ultimate landing site of the moving element has been reached.

In this chapter I will argue against Takahashi, and in favor of what most linguists working in a derivational context already regard as the correct hypothesis – the early formation of intermediate steps of movement. My arguments will be based on novel empirical findings and generalizations in the realm of ditransitive structures. In previous work of mine (Boeckx 2001a, 2003a) I provided empirical arguments in favor of the late formation of intermediate steps of movement. Although these no longer seem to me to be as strong as they once were – more specifically, they do not seem as strong as the arguments in favor of early formation of intermediate steps which I will provide below – they cannot simply be ignored. Instead they must be recast, a task I will also undertake in the present chapter.

I should point out that if my reasoning proves correct, the present chapter can in turn be seen as an argument for a derivational view of syntax, since I will provide an empirical argument for an issue that is only meaningful in a derivation context, and for a solution to an empirical problem that has no obvious representational alternative.

4.2   Takahashi (1994)

Let us begin with a review of Takahashi's (1994) conception of successive cyclic movement, if only to show that the late formation of intermediate steps can be made fully consistent with minimalist desiderata.

Takahashi's starting point was that intermediate steps of movement were not triggered by any feature-checking requirement (an issue I will come back to in the next chapter, where I will strengthen Takahashi's position; see also Boeckx 2001a, 2003a; Bošković 2002, 2006). Instead, Takahashi argued that successive steps are taken simply because of the requirement that movement be local. Takahashi adopted Chomsky and Lasnik's (1993) Minimize Chain Links Condition, which demands that each link of a chain must be as short as possible, one of the earliest economy conditions formulated

in the early 1990s. According to Takahashi, the Minimize Chain Links requirement forces element X undergoing movement of type Y ({A;A-bar}) to stop at every position of type Y on the way to its final landing site independently of feature checking. (For a related proposal, see Sportiche 1989.) Boeckx (2001a, 2003a) argues for a version of Takahashi's view that does not discriminate between different movement types and force movement to stop by all intermediate positions available, making the movement path (quasi-)uniform, fully in line with the conclusion reached in the previous chapter.

It is crucial to note here that Takahashi assumed (along with Chomsky and Lasnik 1993) that the relevant operation underlying movement is Form Chain. If this is so, Form Chain is the operation that ought to be subject to Last Resort. In other words, what must be motivated (typically, in terms of feature checking) is the entire movement path, the formation of an entire chain (collection of copies created by movement), not the formation of individual links of a chain (i.e., movement steps). More succinctly: Takahashi claims that the formation of a chain – i.e., movement in general – must have feature-checking motivation, but the formation of chain links – i.e., each movement step – need not.

Takahashi's assumption is not innocuous. If chain formation (Form Chain) counts as a single operation, formation of a chain cannot be interrupted by any other operation, under the reasonable assumption that operations cannot be interleaved (see Collins 1994 for an early defense of this assumption). This is just another way of saying that intermediate steps of movement will have to be taken at the same time as the final step of movement, with no other operation separating them. Put differently, it is only once the motivation for movement, the final landing site, is introduced that the relevant element will start moving, creating chain links, as the chain is being formed, because of the requirement that movement (not Form Chain as such) be local.

Takahashi's assumption also nullifies a thorny issue raised by Zwart (1996). Zwart argues that the Shortest Move condition, the later incarnation of the Minimize Chain Links condition, is not obviously part of virtual conceptual necessity, for one can think of an equally economical condition, Fewest Steps, that says that the steps of movement must be as few as possible. The Fewest Steps condition goes directly against the formation of intermediate steps of movement. However, Zwart's point is moot if what counts for economy purposes

is the formation of a chain, not chain links – chain links formation not being an independent operation. If Form Chain is taken to be the fundamental theoretical concept, any application of Form Chain counts as one step (see Chomsky 1993:182 and Uriagereka 1998:512 n.2), the fewest number of steps any operation can contain.

From the present perspective, the crucial ingredient of Takahashi's (1994) proposal is that an element does not move until its final landing site has been introduced into the tree, and attraction takes place. This proposal is still consistent with Chomsky's (1995:233−4) assumption that an element must reach its target by the very next step after introduction of the Probe, once Form Chain is taken to be the relevant operation. Chain formation will be the very next step following the introduction of the relevant target.

Chomsky's assumption that checking must be done as early as possible has come to be known as the "virus theory" of "strong feature," given in (4).

(4)   A strong feature must be checked as soon as possible after being introduced into the derivation.

(4) is easily reformulated in terms of EPP requirement (the requirement that forces the creation of a specifier position within the projection of a given head H°), as in (5).

(5)   The EPP requirement of a head H must be satisfied as soon as possible after H° has been introduced into the derivation.

Chomsky (1995:233) phrases this as follows: "Suppose that the derivation D has formed Σ containing α with a strong feature F. Then, D is canceled if α is in a category not headed by α."

Chomsky (1995:233) observes that under the virus theory "Cyclicity follows at once." This is because this conception of strength disallows acyclic checking of heads with strong features. Typically, if overt insertion of a head α with a strong feature takes place acyclically then the derivation will immediately be cancelled, since, by the hypothesis, α is contained in a category not headed by α. Similarly, if overt movement takes place acyclically, then the strong feature F of α driving the movement must not have been checked until a later structure has been created headed by something other than α.

So, the virus theory of strong features and the Extension Condition overlap considerably in their effects. Interestingly, as several authors have noted (see Bošković and Lasnik 1999; Boeckx 2001a, 2003a; Lasnik 2006; N. Richards 1999, 2004), the virus theory of strong features defines a version of the syntactic cycle that renders some seemingly acyclic operations possible. Bošković and Lasnik (1999), in particular, compare the "virus" version of the cycle with other ways of capturing the essence of cyclicity (see Freidin 1999 for a good review of the history of the concept of the cycle), and conclude that it is superior in allowing a variety of attested operations so long as no strong features are involved – operations that would be considered acyclic under different definitions of the cycle, including the Extension Condition. Bošković and Lasnik's argument boils down to saying that the definition of the syntactic cycle is feature-based, as opposed to tree-based.

As Boeckx (2001a, 2003a) notes, intermediate steps of movement resulting from Form Chain would be exactly instances of the kind of seemingly acyclic processes ruled in under the virus theory: not being subject to any checking (let alone checking of a strong feature), they are in a sense orthogonal to cyclicity (i.e., they do not have to take place in what would traditionally be called a strictly cyclic fashion), and thereby do not violate the cycle (if they don't take place "cyclically"). Only Form Chain has to abide by the virus theory.

To conclude this overview of Takahashi's system, late formation of intermediate steps violates the Extension Condition, but the latter is not the only way of capturing cyclicity. Crucially, there are ways of making late successive cyclic movement compatible with minimalist principles. I have argued that the virus theory of the cycle meets the relevant desiderata.

Empirically, evidence for or against late successive cyclic movement is hard to come by. As the reader will recall, the evidence for successive cyclic movement reviewed in chapter 2 comes largely from interface properties such as binding and scope (on the LF-side) or prosody (on the PF-side), phenomena that under plausible assumptions are part of the interpretive components, and act on final representations, at which stage information about the timing of operations is lost. Other indicators of successive cyclic movement, such as morphological markers or the position of the verb after raising, have recently been argued to be determined at PF (see Chomsky 2001; Boeckx and Stjepanovic 2001), especially if morphology is

"distributed" as argued by Halle and Marantz (1993). If morphology essentially amounts to grafting morphemes onto a syntactic structure already formed, and if head-movement amounts to morphological conflation of heads (see especially Harley 2004; see also Boeckx and Stjepanovic 2001), then the timing of syntactic operations will be obscured.

In short, the tests reviewed in chapter 2 provide evidence for intermediate steps, but don't tell us when such steps are taken. This is even true of stranding phenomena if floated quantifiers are treated as adjuncts that are inserted late into the structure, as argued by Bošković (2004). As Bošković and Lasnik (1999) noted, if adjunction is not subject to feature checking, it will be allowed to take place "acyclically" (from the point of view of the Extension Condition), in conformity with the virus theory.

Nevertheless, I submit that it is possible to find evidence bearing on the timing issue. As we will see, the evidence appears to argue for early formation of intermediate steps, and against Takahashi's elegant view.

Evidence for early successive cyclic movement must have the following abstract form: intermediate steps of movement must be taken so as to make the ultimate landing site reachable. In the absence of an early first step, the more distant second step cannot be taken. Evidence for late successive cyclic movement must have the opposite abstract form: information about high portions of the tree are crucially needed to motivate intermediate steps, or something goes wrong if an element moves too early (say, an illegitimate structure is incorrectly ruled in). I will now turn to evidence of either kind and evaluate it.

4.3   The Evidence for Early Successive Cyclic Movement

The literature on bounding nodes, barriers, and phases abounds with evidence for early successive cyclic movement: intermediate steps must be taken to avoid an element being trapped in a phase, or to avoid the crossing of too many barriers/bounding nodes. But such evidence is too theory-internal, and, as we saw in the previous chapter, it relies on stipulations that in effect create the evidence. If there

aren't any phases/barriers/bounding nodes, there is no need to escape them, hence no need to move early.

More uncontroversial evidence for early intermediate steps comes from the realm of applicative structures. The discussion requires a brief detour into the nature of applicative constructions, since this is an area of grammar that is rarely discussed in general introductions.

4.3.1   *Background information on applicatives*

Applicatives are usually understood as constructions in which a verb bears a specific morpheme which licenses an oblique, or non-core, argument that would not otherwise be considered a part of the verb's argument structure, as illustrated in (6) (from Chaga; Bresnan and Moshi 1990:149).

(6)   N – ä – ï – lyì – à   k-élyá.
      FOC-1SUB-PR-eat-FV 7-food
      'He/She is eating food.'

(7)   a.   N – ä – ï – lyì – *í* – à   *m-kà* k-élyá.
           FOC-1SUB-PR-eat-APPL-FV **1-wife** 7-food
           'He is eating food for his **wife**.'
      b.   N – ä – ï – zrìc – *í* – à   **mbùyà**.
           FOC-1SUB-PR-run-APPL-FV **9-friend**
           'He is running for a **friend**.'

If the base verb is transitive (6), the applicative marker may super-transitivize it and produce a double object construction like (7a). On the other hand, if the base verb is intransitive, the applicative morpheme adds the transitive flavor on it, as in (7b).

By extension, the term "applicative" can also be used for oblique/indirect objects of the verb that precedes the direct object in languages even without an overt applicative marker, as Marantz (1984, 1993) and Baker (1988) have done in seminal work. Marantz (1993) was probably the first to recognize that at least some indirect objects are semantically external to the event described by VP. In other words, applicative affixes are elements which take an event as their semantic argument and introduce an individual

which is thematically related to that event. Marantz had in mind examples like (8), from Chaga (taken from Pylkkänen 2002).

(8) a. N – ä – ï – lyì – í – à   *m-kà* k-élyá.
     FOC-1SUB-PR-eat-APPL-FV 1-wife 7-food
     'He is eating food for his wife.'

   b. N – ä – ï – zrìc – í – à   mbùyà.
     FOC-1SUB-PR-run-APPL-FV 9-friend
     'He is making a run for his friend.'

For such examples, Marantz posited a structure whose modern rendering is something like (9).

(9)
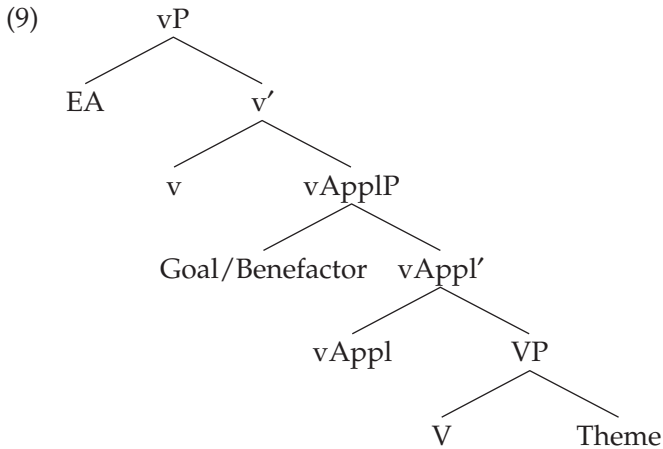
Pylkkänen (2002) demonstrates that Marantz's (1993) proposal cannot be generalized to all applicative constructions, as not all applicatives relate an individual to an event. Building on Pesetsky (1995) (and references therein), Pylkkänen notices that an interpretation where the applied argument bears no "immediate" relation to the direct object is impossible in the English double object construction. For example, in a sentence like (10), the two objects are standardly interpreted in such a way that Jane did the baking for Bill so that he would have the cake.

(10) Jane baked Bill a cake.

On the other hand, in the Chaga applicative construction given in (8a), *the wife* stands in a benefactive relation to the event of eating

but bears no relation to the object *food*. This is so because *the wife* cannot become the possessor of *the food* as a result of somebody else eating it. The same holds for the Chichewa instrumental applicative (11a) (taken from Baker 1988:230), and Albanian applicative construction with a static verb (11b) (taken from McGinnis 2001:5).

(11) a. Mavuto a – na – umb – *ir* – a *mpeni* mtsu.
        Mavuto SP-PST- mold-APPL-ASP knife waterpot
        'Mavuto molded the waterpot with a knife.'
     b. Agimi   i mban Dritës çanten   time.
        Agimi.NOM CL holds D.DAT bag.ACC my
        'Agim holds my bag for Drita.'

In (11a) *the knife* bears an instrumental relation to the event of molding but no relation to *the waterpot*, and (11b) implies that *Drita* could put something in it by an event of (*Agim*) *holding* (*my bag*).

On the basis of this, Pylkkänen (2002) argues that there are two types of applicatied arguments, which she calls high applicatives and low applicatives. The high applicative construction consists of a designated head, the high applicative head (ApplH°), which introduces the applied argument, and expresses a relation between an event (typically the semantic content of the VP) and an individual (the applied argument). Low applicative constructions consists of a low applicative head (ApplL°), which hosts the applied argument, and expresses a relation between two individuals (the direct object and the applied object).

The relevant structures, which Pylkkänen shows capture the right semantic relations cross-linguistically, are given in (12) and (13).

(12) High applicative structure:

(13)   Low applicative structure:

```
                      vP
                    /    \
                  v       VP
                        /    \
                      V      LApplP
                            /    \
                          IO     LAppl'
                                /    \
                             LAppl    DO
```
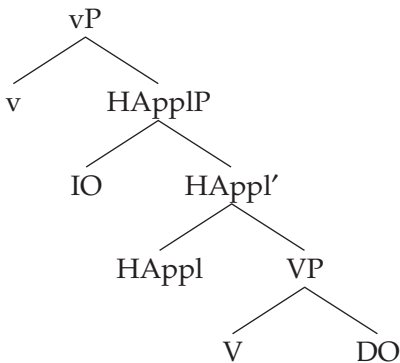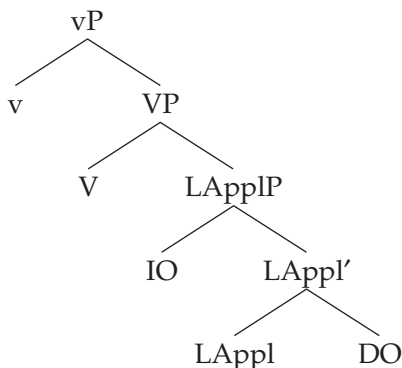
McGinnis (2001) saw the syntactic potential of Pylkkänen's proposal and tried to relate it to another distinction often made in the realm of applicatives: that between symmetric and asymmetric constructions, first due to Bresnan and Moshi (1990), and now well established cross-linguistically (for an excellent overview, see Jeong 2006.)[1]

Asymmetric applicatives are characterized by asymmetric behavior between the direct object and the applied object in such a way that only the applied object shows true object properties. In contrast, in symmetric applicatives, both the applied object and direct object behave as true objects. An example of the kind of variation that arises can be seen in the differences in the verbal agreement pattern. Contrast (14) (from Chichewa; Marantz 1993:127) and (15) (from Kinyarwanda; Kimenyi 1980).

(14)   a.   Chitsiru chi-na-*wa$_i$*-gul-*ir* – a t_i mpatso.
            fool   SP-PST-OP-buy-APPL-FV   gift
            'The fool bought them a gift.'
       b.   *Chitsiru chi-na-*i$_i$*-gul-*ir* -a   atsikana t_i.
            fool   SP-PST-OP-buy-APPL-FV girls
            'The fool bought the girls it.'

(15)   a.   Umugóre a-rá-*mui*-he-er-a t_i ímbwa ibíryo.
            woman SP-PR-OP-give-APPL-ASP dog   food
            'The woman is giving food to the dog for him.'

b. Umugóre a-rá-***bii***-he-er-a   umugabo
woman   SP-PR-OP-give-APPL-ASP man
ímbwa *t_i*.
dog
'The woman is giving it to the dog for the man.'

Whereas Kinyarwanda allows either the direct or the applied object to be incorporated in pronominal form into the verbal complex, Chichewa allows only the applied object to do so.

   The most recognized and attested difference between symmetric and asymmetric applicative constructions is in terms of A-movement (passivization) possibilities. In a symmetric applicative construction, either the applied object (16c) or direct object (16b) can raise to the subject position under passivization. (Examples taken from Chaga; Bresnan and Moshi 1990:51.) By contrast, in asymmetric applicative constructions, only the applied object can be promoted to subject-hood under passivization, as shown in (17) (for English) and (18) (for Icelandic; McGinnis 2001:5).

(16) a. N- ä -ï – lyì -*í* – à   *m-kà*   k-élyá.
FOC-1SUB-PR-eat-APPL-FV 1-wife 7-food
'He is eating food for his wife.'
   b. K-ely   k-i-lyi-i-o   m-ka   *t*.
7-food 7SUB-PR-eat-APPL-PASS 1-wife
'The food is being eaten for the wife.'
   c. M-ka   n- a-i-lyi – i-o   *t*   k-elya.
1-wife FOC-1SUB-PR-eat-APPL-PASS 7-food
'The wife is having the food eaten for her.'

(17) a. John baked Bill a cake.
   b. Bill was baked *t* a cake.
   c. *A cake was baked Bill *t*.

(18) a. Honum   var gefin *t* bokin.
him.DAT was given the book.NOM
'He was given the book.'
   b. *Bokin   var gefin honum.   *t*
the book.NOM was given him.DAT
'The book was given to him.'

Having introduced the two basic distinctions central to applicative constructions (high vs. low applicatives, and symmetric vs. asymmetric constructions), we are now ready to turn to the argument that I would like to make in favor of early formation of intermediate movement steps (chain links).

4.3.2   *The need for early successive cyclic movement*

McGinnis (2001) tried to collapse the two key distinctions introduced in the previous section and reduce the symmetric/asymmetric distinction to the high/low applicative contrast. The key feature of McGinnis's proposal is to be found in a proposal she made in McGinnis (1998), and which can also be found in Ura (1996) and Anagnostopoulou (2003). It all boils down to how to allow for passivization of the lower object in contexts of symmetric passivization.

Ura, Anagnostopoulou, and McGinnis all converge on the view that, all else being equal, passivization of the lower object in an applicative structure should be disallowed on grounds of (relativized) minimality. The basic insight behind (relativized) minimality is that "the operation [of movement] should always try to construct 'the shortest link'" (Chomsky and Lasnik 1993:89). Relativized minimality (first formulated by Rizzi 1990) is a very simple, and yet powerful idea. It accounts for why you must front the first auxiliary in an auxiliary sequence when you want to form questions (19).

(19)   a.   Has John seen it? (Cf. "John <u>has</u> seen it.")
       b.   *Seen John has it?

It also accounts for why a sentence like *somebody bought something* can be converted to a question like *who bought what?*, but not into *what did who buy?* (a so-called superiority effect).

In all these examples, you have the choice of moving either of two auxiliaries, or two wh-words, and in each case you front the "first" one (or the one closer to the target position; recall that by "first," I really mean "higher," as syntactic processes rely on hierarchical structure, not linear structure). In all cases, you take the shorter path (relativized to the type of elements that could move), as schematized in (20).

(20) $x_\alpha \ldots y_\alpha \ldots z_\alpha$

In an Agree framework, minimality is typically formulated in terms of Minimal Search. Once a Probe looks down into its complement domain to attract a matching element, it selects the first element it finds in that domain. Going back to ditransitives, we can see that by any metric, the higher object should count as the closest element to move to subject position (say, SpecTP) in both high and low applicative structures.

(21)   High applicative structure:

T° → vP

v   HApplP

   IO   HAppl'

      HAppl   VP

         V   DO

Complement domain of T°/search space

(22)   Low applicative structure:

T° → vP

v   VP

   V   LApplP

      IO   LAppl'

         LAppl   DO
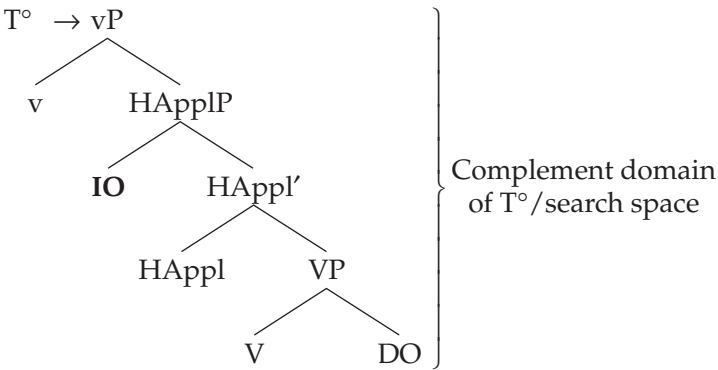
Complement domain of T°/search space
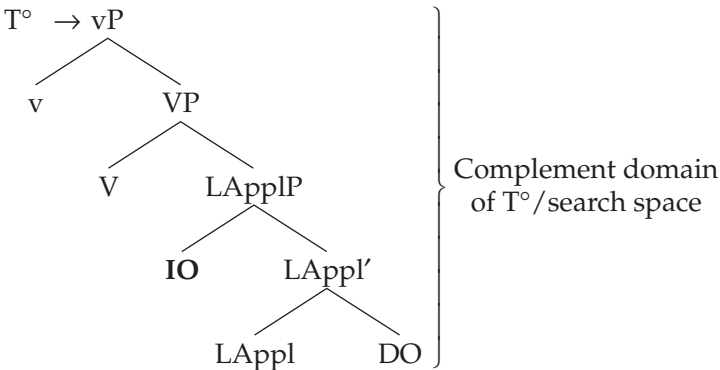
Movement of the lower object ought to constitute a violation of relativized minimality. Put differently, all else being equal, we expect

all applicative structures to display asymmetric behavior. Since this is not so empirically, some mechanism must be found to avoid the apparent minimalist violation. The mechanism proposed by Ura, Anagnostopoulou, and McGinnis is one whereby the lower object first moves to a position slightly above the original position of the higher object, without crossing the projection that hosts the latter. From such an adjoined or extra specifier position, the once lower object is able to continue its journey to SpecTP by virtue of being the element closer to T°. Specifically, given a high applicative structure like (23), the authors under discussion assume that the lower object moves to an extra specifier of ApplP, right above the specifier position occupied by the applied object. (I return to low applicative structures in the next chapter.)

(23)
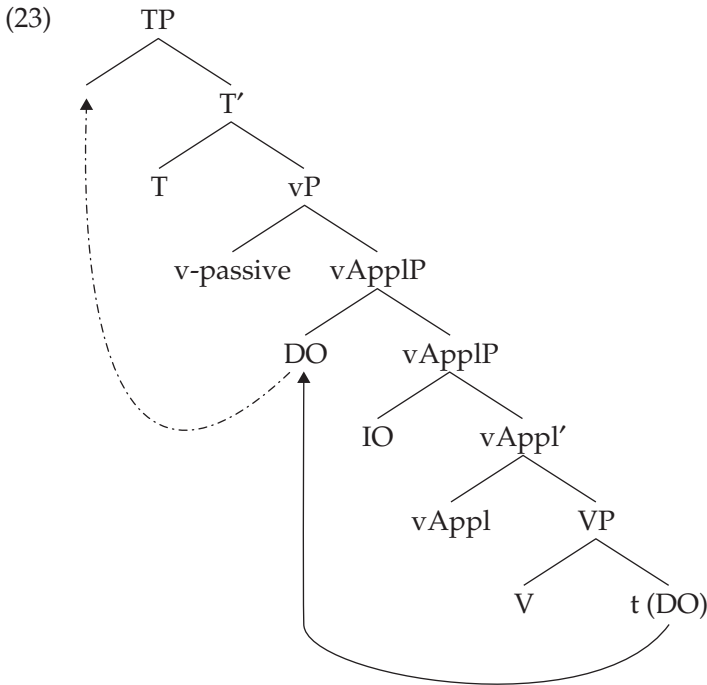
Under such a derivation, minimality is obeyed at all steps. Movement of the lower object is licit, as the lower object is the only element within the domain of Appl°, hence counts as closest. And once the lower object occupies the higher SpecApplP, it again counts as closer for attraction by T°.

It is clear that the movement that brings the lower object cannot be driven by case consideration (often, arguments move to get case, but the movement at issue takes place as a necessary first step for ultimate movement to TP, where case is typically assigned, and multiple case assignment to the same element is banned).[2] The relevant movement cannot be A-bar movement either, as the element will subsequently moved to an A-position (SpecTP), and that movement would count as "improper" (A-bar movement cannot feed A-movement: see Chomsky 1981; Boeckx 2006b).

The most likely candidate appears to be whatever drives intermediate movement steps, the topic of the next chapter.[3] The key point in the present context is that this intermediate movement step must necessarily take place prior to movement to T°. In the absence of such a first step, the second movement would be excluded as a minimality violation. We thus have evidence that intermediate steps of movement must be taken early. I regard this evidence as fairly strong since it implicates the (relativized) minimality principle, which is one of the central principles, if not the central principle, of current syntactic theory that makes eminent sense in a minimalist context.

Notice how the present discussion provides support for a derivational model of syntax. If the structure of applicative constructions is as discussed above (direct object in a lower projection, distinct from the projection occupied by the applied object), as seems forced upon us by the kind of semantic considerations discussed by Pylkkänen (2002), there is no way to rule in passivization of the lower object other than by capitalizing on a first, "leapfrogging" intermediate movement step. The other alternatives (parametrizing minimality, ternary branching structures for ditransitives, random order of base generation of arguments, etc.) would amount to the abandonment of independently well-motivated and comprehensive principles of the grammar. This would clearly be a step backward in linguistic theorizing at this stage of our understanding.

The argument I have just offered for early successive cyclic movement makes a rather strong case for models of grammar that make use of notions like sequential ordering of operations (i.e., so-called derivational models), and against strictly representational models of the type advocated by Brody (1995, 2003). (The logic of the argument is very reminiscent of the feeding effects studied in the rule-based models of phonology and syntax that were popular before the advent of the Principles-and-Parameters and Optimality-Theoretic approaches.)

Although I think we have reached the kind of evidence we want for a rather subtle theoretical question (timing of intermediate steps), I wouldn't want to leave the reader under the impression that there is no evidence going the other way (evidence in favor of late formation of intermediate movement steps). I will now turn to the evidence that some have adduced in favor of Takahashi's proposal, and show that the evidence is not incontrovertible.

4.4   Potential Arguments for Late Successive Cyclic Movement

There are three types of situations which can be (and have been) used to argue for late intermediate steps.

4.4.1   Sub-extraction out of a moved element

The first argument (discussed in Boeckx 2001a, 2003a) refers to an abstract situation like the following. Suppose one finds a phrase γ, which readily allows sub-extraction of α (say, non-specific *a picture of* α). In an approach that allows intermediate links to be formed at will, with no feature checking involved in intermediate sites (an issue I turn to in the next chapter), one could in principle allow for movement of α out of γ, followed by movement of γ to a position β out of which sub-extraction is impossible. Nothing seems to prevent further movement of α, as the latter has moved out of γ prior to the latter's fateful landing on a freezing node. An example of this scenario is given in (24), with the derivation sketched in (25).[4]

(24)   Target: *Who did [a picture of <who>] cause Bill to cry?

(25)   a.   Movement of *who* out of [*a picture of*] to YP when the *picture*-NP is in Spec$v$P (where sub-extraction is allowed).
       b.   Movement of *a picture of <who>* to SpecIP, a "freezing node."
       c.   Movement of *who* from SpecYP to SpecCP.

Nothing seems to ban the undesirable derivation sketched in (25), unless we adopt the idea that movement is initiated upon insertion of the final landing site. When C^0 is inserted, *picture of who* would already have landed on a freezing node, and sub-extraction would be doomed. This would correctly rule out (24).

I think the logic of the argument is sound, but it depends on notions that have not been as decisively and definitively characterized as the cause of minimality effects that I used in arguing for early successive cyclic movement. In particular I have in mind here the notion of freezing node or island. Several ways have been explored in the minimalist literature to capture islandhood and opacity (see Boeckx 2006b for review and discussion; see also chapter 7), and it may well be that the abstract derivation under discussion will be ruled out even if intermediate steps are taken early.

We could, for example, stipulate that no chain may contain a copy whose sister heads a phrase in SpecIP. Opacity would then have to be treated in representational as opposed to derivational terms. But this would suffice to nullify the argument for late successive cyclic movement I presented in earlier work.

Alternatively, Bošković (p.c.) suggests that the derivation under discussion may be ruled out by some version of the A-over-A condition. The A-over-A condition, first proposed by Chomsky in 1964, states that if both A and B are eligible for movement, and A contains B, only A can move. In the case at hand (24), Bošković (2005) claims that the sub-extraction step moving *who* out of *[a picture of (who)]* is rendered impossible because the entire nominal element (*[a picture of (who)]*) contains a feature that also requires it to move, and take precedence over the feature-forcing movement of *who*. So, in principle, successive cyclic movement takes place as early as it can, but in this particular case, the sub-extraction step is independently blocked.

Finally, I must not fail to mention that Chomsky (to appear) cites cases of licit sub-extraction out of displaced subjects, such as (26), that suggest that the abstract derivation under discussion must be available, as the resulting structures sometimes lead to acceptable results (but see Broekhuis 2006, Gallego 2006, and Boeckx 2006b, for arguments against Chomsky's position).

(26)   it is the truck (not the car) of which the driver was injured
         in the accident that . . .

In sum, the cause of freezing effects is not clear enough to provide a basis for a solid argument in favor of late successive cyclicity. This is, to repeat, unlike the argument for early intermediate movement steps, which is based on the principle of grammar we understand best, hence the one we should trust most: minimality.

4.4.2   Intervening traces

The second argument for late successive cyclic movement is a version of Chomsky's (2001) argument for a phase-based (as opposed to a strictly cyclic) evaluation of locality (minimality, in particular). Chomsky proposed a phase-based locality on the basis of sentences like (27).

  (27)   What did John [t_{what} [t_{John} buy t_{what}]]?

At issue is the role of the intermediate trace of *what*: why doesn't it block the relation between *John* and T? If locality were checked at each stage of the derivation (call this the strictly cyclic view on locality), blocking should take place. At the point when T° tries to reach inside the VP and attract *John*, it first encounters a copy of *what*, which should be enough to block attraction.

One could turn the issue raised by Chomsky into an argument for late successive cyclic movement. If successive cyclic steps are taken only upon the introduction of the final landing site (C[+wh], in this case), there won't be any copy of *what* at the edge of the VP-domain to block movement of *John*. Once *John* has moved to SpecTP, since it doesn't bear a wh-feature, *what* will be free to move to SpecCP, passing by the edge of the VP-domain.

Chomsky (2001), however, suggests a different tack, one that does not rely on late successive cyclic movement, hence a potential argument for the position defended in this chapter. To account for (27), Chomsky argues that the reason attraction is not blocked is that intermediate copies/traces of movement are in a sense invisible for computational purposes (an idea going back to Uriagereka 1988). The task, then, for Chomsky, amounts to turning the copy of *what* at the edge of the VP-domain into an intermediate trace. Put differently, one must evaluate locality (i.e., check that no intervener is present) after movement of *what* beyond the domain of IP

(movement to SpecCP). To guarantee this result, Chomsky claims that phases (*v*P and CP, but crucially not TP) are also the units for evaluation of minimality. This means that in a phase-based system of the kind Chomsky explores in recent work, one must suspend talk of locality until the relevant phase level is reached. Going back to (27), if minimality is evaluated at the CP level, and intermediate copies/traces of movement are invisible for computational purposes, nothing intervenes between T° and *John*. The acceptability of (26) can therefore be captured.

Readers may object to Chomsky's phase-based solution, on the grounds that it seems unfair to appeal to phases to dismiss an argument against my claim that successive cyclic movement takes place early, since I rejected phase-based derivations in the previous chapter. This is one of the places in the present book where the funnel-like structure of my argument (mentioned in chapter 1) really can be felt. Accordingly, I will sketch an alternative.[5] But I just wanted to mention Chomsky's argument to emphasize the fact that the argument for late successive cyclic movement is open to alternative accounts, hence is weaker than the argument I offered for early successive cyclic movement, for which I do not know of any serious alternative.

The alternative approach to Chomsky's phase-based solution I would like to propose makes crucial use of case. If *what* moves to the outer Spec*v*P early (i.e., prior to *John* moving to SpecIP), it is quite standard to assume that it already has its case checked when T° is introduced into the derivation. If so, one could claim that *what* doesn't count as a matching element when T° is looking for a nominal element to move to its specifier; it is therefore transparent for purposes of minimality.

4.4.3   Object agreement

The final argument for late successive cyclic movement comes from a refinement of Boeckx and Niinuma's (2004) analysis of object honorification in Japanese, proposed in Lee (2004). The basic phenomenon is illustrated in (28).

(28)   Taroo-ga Tanaka sensei-o   o-tasuke-si-ta/tasuke-ta.
       Taro-NOM Prof. Tanaka-ACC OH-help-do-PAST/help-PAST
       'Taro helped Prof. Tanaka.'

A transitive verb agrees with the Theme direct object (29a), whereas ditransitive verbs agree with the Goal indirect object in honorification, as (29b) shows.

(29)   a.   Taroo-ga   Tanaka sensei-o   o-tasuke-si-ta.
            Taro-NOM Prof. Tanaka-ACC OH-help-do-PAST
            'Taro helped Prof. Tanaka.'
       b.   Hanako-ga   Tanaka sense-ni   Mary-o
            Hanako-NOM Prof. Tanaka-DAT Mary-ACC
            go-syookai-si-ta.
            OH-introduce-do-PAST
            'Hanako introduced Mary to Prof. Tanaka.'

In another ditransitive context (30a), the NP capable of triggering honorification (i.e., Prof. Tanaka) appears as a direct object and the object honorification is blocked. The change of word order also does not help, as shown in (30b).

(30)   a.   *Hanako-ga   Mary-ni   Tanaka sensei-o
            Hanako-NOM Mary-DAT Prof. Tanaka-ACC
            go-syookai-si-ta.
            OH-introduce-do-PAST
            'Hanako introduced Prof. Tanaka to Mary.'
       b.   *Hanako-ga   Tanaka sensei-o   Mary-ni
            Hanako-NOM Prof. Tanaka-ACC Mary-DAT
            go-syookai-si-ta.
            OH-introduce-do-PAST
            'Hanako introduced Prof. Tanaka to Mary.'

Boeckx and Niinuma (2004) treat honorification as a case of (abstract) agreement on a par with the common φ-feature agreement of many other languages, and regard the phenomenon in (30) as a blocking of direct object honorification agreement in the presence of a dative argument due to a locality constraint; more specifically speaking, a minimality effect of the familiar sort. The schematic representation of this effect is given in (31).

(31) v [$_{VP}$ Dative-NP [V Accusative-NP]]

For this analysis to work, it is crucial to assume that the dative element c-commands the accusative element, since if the reverse were possible, we predict that the accusative element would be closer to v and there would be no defective intervention; hence the honorification agreement with the accusative element should be possible, contrary to fact, as (30b) demonstrates. If Boeckx and Niinuma's (2004) theory of object honorification in terms of agreement and blocking is on the right track, it entails that the relative order of the two objects in ditransitive constructions must be the <Goal; Theme> order at the point when honorification agreement takes place.

Crucially, Boeckx and Niinuma note that the constraint on object honorification treated as a minimality effect obtains even if the direct object undergoes movement (e.g., scrambling), as shown in (32).

(32)   *Tanaka sensei-o Hanako-ga   Mary-ni
       Prof. Tanaka-ACC Hanako-NOM Mary-DAT
       go-syookai-si-ta.
       OH-introduce-do-PAST
       'Hanako introduced Prof. Tanaka to Mary.'

Notice that for minimality to be relevant in such cases, it is important that the direct object not move (in successive cyclic fashion) to a position in between v and the indirect object, say a VP-adjoined position:

(33) v [$_{VP}$ Accusative-NP [Dative-NP [V $t_{\text{Accusative-NP}}$]]]

Lee (2004) argues that it is not imperative to block such intermediate movement at all costs. If successive cyclic movement happens late, i.e., after honorification agreement has taken place, minimality will still have the desired effect as in (34).

                              #2 (movement; OK)
(34) v [$_{VP}$ Dative-NP [V Accusative-NP]]
                              #1 (agreement; blocked)

If intermediate steps of movement are taken early, however, it is necessary to make sure that no site between v and the indirect object

remains accessible to the direct object. It now becomes important to examine what the precise VP structure of double object constructions may be. Jeong (2006) examines ditransitive constructions in detail and argues that Japanese makes use of two structures, given in (35)–(36) below (see Jeong's work for full exposition and justification).

(35)   [$_{vP}$ Subj $v°$ [$_{VP}$ PP [V° DO]]]   (PP-dative structure)

(36)   [$_{vP}$ Subj $v°$ [$_{VP}$ V° [$_{ApplP}$ IO Appl° DO]]
                                                    (Low-applicative structure)

Given these two structures, here are the intermediate steps for DO-movement to be avoided (marked by ⊗), assuming $v°$ to be the source for object honorification.

(37)   [$_{vP}$ Subj $v°$ [⊗ [$_{VP}$ IO [V° DO]]]

(38)   [$_{vP}$ Subj $v°$ [⊗ [$_{VP}$ V° [⊗ [$_{ApplP}$ IO [Appl° DO]]]

We could of course stipulate that the relevant sites are not proper landing sites, but needless to say such a move would be far from explanatory.

   I happen to think we can do better. Following a growing literature, I will argue in the next chapter that an element cannot move anywhere within the projection in which it originated, as that movement would count as "too close" (see next chapter for the rationale behind this condition). Accordingly, DO wouldn't be able to move to the outer SpecApplP in (38) or the outer SpecVP in (37). The landing site that remains to be excluded is thus SpecVP in (38). This is plausibly done on minimality grounds, as such a movement would cross IO. In sum, the honorification data can be analyzed in a way consistent with the claim that intermediate steps of movement are taken early.

4.5   Conclusion

To conclude, the present chapter has argued that what many syntacticians working within a generative paradigm would regard as

the most conceptually appealing answer to the question of when intermediate sites are taken (i.e., such steps are taken early, "strictly cyclically") is correct empirically. Arguments to the contrary are weaker, as they involve properties of grammar whose character cannot yet be taken as definite, and can thus be reanalyzed in a way consistent with the claim that intermediate movement steps are taken as soon as the targeted site is introduced into the derivation.

Notes

1   The symmetric/asymmetric contrast in often called a parameter distinguishing languages (see Baker 2004 and references therein), but this is clearly incorrect since several languages exhibit both types of constructions, as Jeong (2006) shows.
2   In this respect the situation under discussion is distinct from the one discussed in Bošković (1997b), where apparent superiority violations are accounted for by way of a two-step derivation very similar to the one under discussion, except that both movement steps are featurally motivated in Bošković's case.
3   For this reason, McGinnis (2001) argues that the high applicative phrase is a phase (see also Lee 2004). Jeong (2006) provides rather compelling arguments against a phase-based approach to the issue at hand.
4   Bošković (2006) discusses a similar case, viz. (i), originally due to Postal (1972).

>   (i)   *Which garage do you think that [in *t*] John found this car?
>   (cf. "In which garage do you think that John found this car?")

Postal was concerned with how preposition stranding (which is independently attested in English; cf. (ii)) can be blocked in intermediate landing sites.

>   (ii)   Which garage do you think that John found this car in?

Postal took the unacceptability of (i) to argue against successive cyclic movement. But his conclusion is too strong. One could rule out preposition stranding in (i) by appealing to late intermediate movement step formation. An operation like preposition stranding would have to take place prior to movement, if it is to take place at all, for, once chain formation takes place, it cannot be interleaved with another operation like preposition stranding.

But although late successive cyclic movement is adequate to handle (i), as it is for (24), it is not the only option available to us. Other solutions, compatible with the early successive cyclicity view argued for here, quickly emerge. One could, for example, appeal to a ban on sub-extraction from moved elements (see Takahashi 1994; Boeckx 2003a). That is, claim, as the two authors just mentioned have done on independent grounds, that once an element is moved, it becomes opaque to extraction from within it. Alternatively, one could appeal to the A-over-A principle, and say that once the system has decided to move the entire PP, extraction of the NP-complement is banned, as the entire PP is available for movement. (I should also mention in passing that apparent examples of preposition stranding in intermediate landing sites have been reported for Afrikaans in du Plessis 1977. It is, however, not clear whether Afrikaans, like German and Dutch, exhibits genuine preposition-stranding properties. See Abels 2003.)

5   There are in fact independent reasons not to rely on Chomsky's phase-based account. Evidence against Chomsky's analysis comes from the realm of raising constructions involving an experiencer. As is well known, many languages disallow the type of raising that English allows in situations like (ii).

(i)   John seems to be smart.   (OK in many languages)

(ii) John seems <u>to Bill</u> to be smart. (disallowed in many languages)

One such language is Italian. Interestingly, in this and other languages, subject raising across an experiencer is not completely disallowed either. The key factor allowing raising appears to be the surface position of the experiencer. If the experiencer is a moved wh-phrase, or a topicalized element, or a clitic (i.e., if an experiencer has obviously moved out of its base position), raising is possible. If the experiencer is a full DP (by assumption, located in its base position), subject raising is blocked, as in (iii).

(iii) a. A chi   sembra Gianni$_i$ [t$_i$ essere stanco]?
           to whom seems Gianni   be   ill
           'To whom does Gianni seem to be ill?'
       b. A Maria, Gianni$_i$ gli sembra [t$_i$ essere stanco].
           to Maria Gianni her seems   be   ill
           'To Maria, Gianni seems to be ill.'
       c. Gianni$_i$ gli   sembra [t$_i$ essere stanco].
           Gianni to-her seems   to-be ill
           'Gianni seems to her to be ill.'

(iv)   *Gianni$_i$ sembra a Maria [t̲$_i$ essere stanco].
         Gianni   seems   to Maria   to-be ill
         'Gianni seems to Maria to be ill.'

Whereas the transparency of the experiencer is fully expected under Chomsky's phase-based approach in cases where the experiencer is an A-bar moved element, the transparency of clitics is not, since for Chomsky it is critical that the potential intervener raises past the subject for raising to be disallowed.

Chapter 5

The Motivation for Intermediate Movement Steps (The *Why*-Question)

5.1 Last Resort

At the heart of the present chapter lies a simple but all-important question concerning the phenomenon of successive cyclicity: what is the rationale, the motivation for intermediate movement steps?

The minimalist insistence on the economical, computationally efficient character of syntactic operations finds its clearest and longest-lasting instantiation in the view that movement is a Last Resort operation, driven by feature checking, a hypothesis first formulated by Chomsky (1986a) to account for the unacceptability of cases like (1).

(1)   *John seems [t is ill].

Sentences like (1) are ruled out by appealing to the fact that the moved element (*John*) had satisfied all its requirements in the embedded clause (*John is ill* is a fine sentence), and, as such, had no reason to undergo further movement.

The reasoning just sketched turned out not to be restricted to case features and instances of A-movement. Rizzi (2006) and Bošković (2005) have argued that the so-called freezing effect that characterizes a wh-word that has reached a [+wh]-checking site is also a reflex of Last Resort. Thus, the sentence in (2) is excluded because once the wh-word has checked a wh-feature, it cannot check another one higher up in the tree.

(2)   *Who did John wonder *<who>* Mary saw?

The account generalizes to cases where a given element is forced to check an A-feature and an A-bar feature. In particular, Boeckx (2001a, 2003a), N. Richards (1997, 2001), Rizzi (2006), and Rizzi and Shlonsky (2005) have argued that the well-known difficulty in extracting raised subjects, illustrated in (3), is due to the fact that raised subjects, prior to movement, have reached a position that typically requires pronunciation (a so-called EPP-position, a position that demands to be filled.)

(3)   *Who do you think that <who> left?

Minimalism elevates the type of analysis just discussed to a principle of Least Effort: if you need not, you cannot. Last Resort/Least Effort considerations have led to the intensive investigation of the range of features driving movement and their properties. Fine-grained analyses of features in turn led to "cartographic" projects, attempts to identify as precisely as possible the specific locations of all possible syntactic features and their relative hierarchical positions (see Cinque 1999, 2002; Rizzi 2004a).

5.2   Problematic Cases
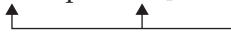
Although Last Resort appears to be a key feature of syntactic computations, there remain a few recalcitrant cases where a given element appears to enter into multiple (checking) relations of a similar kind. The cases that are always mentioned in this context are instances of concord and instances of successive cyclic movement.

5.2.1   Concord

The example in (4), from Hindi, illustrates concord properties.

(4)   Vivek-ne [kitaab parh-nii] chaah-ii.
      Vivek-erg book.f read-inf.f want-pfv.f
      'Vivek wants to read the book.'

Here the direct object of the embedded clause triggers agreement on both the embedded verb and the matrix verb. Various ways have been tried to account for concord phenomena in a way consistent with Last Resort. The analysis I favor (which I have argued for in the context of examples like (4) in Boeckx 2004) invokes the concept of Multiple Agree (the term is due to Hiraiwa 2001, 2005, but the basic idea goes back at least to Bošković 1999). The idea behind Multiple Agree is that a checking relation can involve more than two participants at a given time. That is, within a given domain, a single Probe may enter into checking relations with multiple Goals simultaneously (5a). The figure in (5b) schematizes the Multiple Agree relation responsible for the surface form in (4).

(5)   a.   Probe . . . [. . . Goal 1 . . . [. . . Goal 2 . . .]]

   b.   Vivek-ne [kitaab parh-nii] chaah-ii.

Simultaneous checking with multiple goals gives the impression that a given element has entered into distinct checking relations, involving the same feature, with several elements in an iterative fashion in violation of Last Resort, whereas in fact Multiple Agree analyses contend that only one such checking relation has been established, in conformity with Last Resort.

5.2.2   Successive cyclicity

Successive cyclic movement poses a different problem for the idea that Last Resort/Least Effort regulates syntactic computation: although the hypothesis that the final step of movement is a checking site appears to be sound, there doesn't seem to be any obvious motivation for intermediate movement steps.

5.2.2.1   Spurious features

Since the very beginning of the minimalist program, various featural options have been tried in an attempt to motivate intermediate steps of movement (see Collins 1997; Sabel 1998; Fanselow and

Mahajan 2000; Adger 2003; Rizzi 2006; to name but a few), I am inclined to agree with McCloskey's (2002:186) assessment that all the features appealed to boil down to "spurious," or "pseudo-"features (+Q, +Op, +Wh, etc. for A-bar movement; +Null_Case for A-movement; see Uriagereka forthcoming). By "spurious" features, McCloskey means movement-triggering features optionally present on intermediate landing sites, whose presence is required neither by lexical requirements nor by considerations of interpretability. They are there simply to make intermediate movement steps conform to Last Resort. (This is not to say that no other, non-featural options have been explored to capture the need for successive cyclic movement. I discuss salient alternatives in the next chapter.)

Rizzi (2006) has recently argued that although the features invoked in intermediate landing sites may appear *ad hoc*, there is actually some empirical evidence for them. In particular, Rizzi points to facts like those discussed in chapter 2. Rizzi notes, correctly, that in languages that allow subject–verb/auxiliary inversion in intermediate landing sites (Spanish, French, Belfast English, etc.), the operation is always as sensitive to the nature of the moving element as it is in the context of the final landing site of movement. That is, just as subject–verb inversion is only allowed with wh-phrases, and not with, say, topicalized material in mono-clausal contexts, so is subject–verb inversion in intermediate landing sites. Put differently, just as short topicalization fails to trigger subject–auxiliary verb inversion in Belfast English (6), in contrast to short-distance wh-movement (7), long-distance topicalization fails to trigger successive cyclic inversion (8), again in contrast to long-distance wh-movement (9).

(6) a. This book, Mary will read.
     b. *This book will Mary read.

(7) What will Mary read?

(8) a. This book, John can believe Mary will read.
     b. *This book, can John believe will Mary read.
     c. *This book, can John believe Mary will read.

(9) What can John believe will Mary read?

From this generalization Rizzi concludes that the feature (presumably [+wh]) responsible for subject–auxiliary inversion at the final

landing site must also be present in intermediate landing sites. If it weren't, how could we account for the contrast between wh-movement and topicalization? Accordingly, Rizzi concludes that intermediate steps of movement are triggered by an uninterpreted version of the feature responsible for the final step of movement in long-distance dependencies.

Although very interesting, Rizzi's argument is not forced upon us, as the generalization underlying the paradigm in (6)–(9) is subject to an alternative explanation. Specifically, the contrast between wh-movement and topicalization may be due not to the specific feature on the head of the projection hosting the moved phrase, but to the nature of the moved phrase ([+wh] vs. [+topic]). Rizzi's conclusion depends on head-movement from T° to C° being dependent on a feature on C°, that is, it relies on head-movement being feature-driven. But in recent years such a conclusion has been questioned from various perspectives (see Boeckx and Stjepanović 2001 for review). If, as Chomsky (2001), Boeckx and Stjepanović (2001), and others have argued, head-movement is a PF-operation, taking place after all syntactic operations have applied, Rizzi's conclusion rests on shaky grounds.

Since Rizzi's conclusion wouldn't provide a rationale for intermediate steps of movement being triggered by a feature similar to the feature responsible for the final step of movement (Rizzi's conclusion does not answer the minimalist why-question), I conclude that the empirical evidence he provided is not strong enough. I tentatively conclude that spurious A- or A-bar features are not present in intermediate landing sites.

5.2.2.2   Agreement

It has sometimes been suggested that intermediate movement steps, of both A- and A-bar types, are triggered by the checking of φ-features (see, e.g., Hornstein 2001:119). On the surface, such a conclusion appears to be well motivated empirically, on the basis of such phenomena as past participle agreement facts for French (Kayne 1989) and "wh-agreement" facts for Chamorro (Chung 1998 and references therein). As the following examples show, subject raising and wh-movement appear to have a clear influence on the morphology of verbal forms, specifically their phi-features (person, number, gender, case), along the movement path.

(10)   Jean a repeint les chaises.

(11)   Les chaises ont été repeint-**es**.

(12)   Quelles chaises Jean a-t-il repeint-**es**?

(13)   **Ha**-fa'gasi si   Juan i   kareta.
       E3s-wash Unm Juan the car
       'Juan washed the car.'

(14)   Hayi f-**um**-a'gasi i   kareta?
       who UM-wash   the car
       'Who washed the car?'

However, there are both conceptual and empirical problems with the hypothesis that phi-features drive intermediate movement steps. Conceptually, it is not at all clear why phi-features should be located in intermediate landing sites. Why phi-features, and no other features? Empirically, the alleged agreement facts are less obvious upon closer examination. Some of the points made about wh-agreement in chapter 2 are worth bringing up again.

First, as is very clear from the literature on the languages displaying wh-agreement, morphological changes on the verbs along the movement path are only indirectly conditioned by the moving phrase. That is, although wh-movement induces a morphological change on intermediate verbs (verbs along the wh-movement path), the morphological change refers to a special kind of "agreement" between the verb and the clause from which the wh-phrase has been extracted. In particular, when overt wh-movement takes place, the verbs along the way to the ultimate [+wh] SpecCP bear the morphology they would bear if the complement clause out of which wh-movement took place were extracted, with the exception of the clause from which movement was launched, where the morphology directly reflects the featural content of the moving element.

Interestingly, one finds a similar asymmetry between the clause containing the launching site and subsequent embedding clauses in the context of past participle agreement in French. There agreement on a past participle is restricted to the most deeply embedded clause from which movement takes place. So, while it is plausible to say that movement internal to the most deeply embedded clause is triggered

by phi-feature checking, it appears much less plausible to say that sub-sequent intermediate landing sites are triggered in the same fashion.[1]

Second, when one takes a closer look at the actual morphology that appears on the verbs along the movement path, one immediately realizes that the term "agreement" is very misleading. In some cases, the term "agreement" is the result of theoretical considerations that are far from obvious. For example, in their analysis of long-distance dependencies in Tagalog, Rackowski and Richards (2005) follow Rackowski (2002) in characterizing the morphology on the relevant verb as case (agreement) morphology. Let me provide relevant examples in (15)–(17).

(15) Sino ang   nagbigay   ng bulaklak sa   kanya?
      who ANG **NOM**-gave NG flower   DAT 3
      'Who gave him/her the flower?'

(16) Sino ang   binigyan   mo   ng bulaklak?
      who ANG **DAT**-gave NG-you NG flower
      'Who did you give the flower to?'

(17) Ano ang   ibinigay   mo   sa   kanya?
      what ANG **OBL**-gave NG-you DAT 3
      'What did you give him/her?'

Whereas such distinctions can be regarded as cases they may with equal plausibility be treated as theta-roles (see Baker 1997), especially if theta-roles are treated as features undergoing checking (see Hornstein 2001 and references therein).

In some other cases, the term "agreement" appears to be just plainly wrong. For example, in Selayarese, the otherwise obligatory agreement between a verb and its clausal argument disappears if the clausal argument contains a gap, and the complementizer is dropped, as the following examples illustrate.

(18) Ku-isseʔ-*(i) *(kuko) la-ʔalle-i doeʔ-iñjo   i Basoʔ.
      1s-know-3   COMP 3-take-3 money-the h Baso
      'I know that Baso took the money.'

(19) Apa mu-isseʔ   la-ʔalle _ i Basoʔ?
      what 2FAM-know 3-take   h Baso
      'What do you know that Baso took?'

The case of Selayarese is far from isolated. As I have documented in Boeckx (2003a), anti-agreement is often found in the context of (sub-)extraction (see also chapter 7). As a matter of fact, I know of no clear case where genuine phi-agreement facilitates extraction. For this reason, one should be suspicious of the phenomenon of wh-agreement *qua* agreement.

The only instance of genuine long-distance wh-agreement (where morphology co-varies with the featural specification of the wh-word, not the clause containing it) I am aware of is found in Kinande. As Schneider-Zioga has illustrated in a series of papers (Schneider-Zioga 2000, 2002, 2004; see also Rizzi 1990:55), the language expresses the noun class and number of the wh-phrase on the complementizer(/focus-marker) immediately adjacent to the wh-phrase (20), as well as on complementizers along the wh-path (21).

(20) a. IyondI **y0**   Kambale alangIra?
         who.1 that.1 Kambale saw
         'Who did Kambale see?'
     b. ABahI **Bo**   Kambale alangIra?
         who.2 that.2 Kambale saw
     c. EkIhI **ky0**   Kambale alangIra?
         what.7 that.7 Kambale saw
         'What did Kambale see?'
     d. EBIhI **By0**   Kambale alangIra?
         what.8 that.8 Kambale saw

(21) EkIhI$_j$ **kyo$_j$** Yosefu a-kabula [$_{whp}$ nga-**kyo$_j$** [$_{ip}$ a-kalangira t_j]]?
      what FOC J.   wonders if -FOC   agr.sees
      'What does Yosefu wonder if he sees?'

In addition to its relative isolation in the domain of markers of successive cyclicity, Kinande wh-agreement is puzzling in a number of other respects. For example, Schneider-Zioga (2006) has observed that whereas local A'-extraction in Kinande allows for a reconstructed interpretation (22), reconstruction is impossible if the displacement is long-distance (23).

(22) ekitabu kiwe$_{j/k}$ **ky'**   obuli mukolo$_j$ a.kasoma
      book   his   *wh-agr(eement)* each student agr.reads
      kangikangi.
      regularly
      '(It is) his$_j$ book that [every student$_{j/k}$] reads regularly.'

(23) ekitabu kiwe$_{k/*j}$ **kyo**   ngalengekanaya [$_{cp}$nga.**kyo**
      book   his   *wh-agr* I.think   *wh-agr*
      [obuli mukolo]$_j$ akasoma __ kangikangi].
      every student read   regularly
      '(It is) his$_{k/*j}$ book that I think [every student]$_j$ reads
      regularly.'

I have stressed in chapter 3 that lack of reconstruction should not be taken to signal lack of movement, but surely absence of reconstruction effects render the pattern found in long-distance dependencies in Kinande even more marked. Elsewhere (Boeckx 2006b), I have suggested we extend to Kinande the type of analysis proposed by Davies (2003) for long-distance wh-questions in Madurese and Javenese, a suggestion adopted by Schneider-Zioga (2006) (see also McCloskey 2002 for Irish; Zwart 1996). Davies argues that apparent long-distance wh-movements are instances of iterative prolepsis, as schematized in (24). (It is interesting to note that Kinande wh-extractions, like Madurese and Javenese wh-questions, have a cleft-like nature.)

(24) Wh$_i$ [pro$_i$ C . . . . <pro$_i$> [pro$_i$ C . . . <pro$_i$>]]

According to such an analysis, the agreement we find on complementizers could be incorporated pronouns that are linked in a way similar to what we find in the English construction in (25). (On agreement as incorporated pronouns in Bantu languages, see Bresnan and Mchombo 1987.)

(25) I believe <u>of Mary</u> that Bill said <u>of her</u> that <u>she</u> is smart.

Under such an analysis there is no agreement as such between the wh-phrase and the complementizer; it is the co-indexed pronoun in each clause that causes one to think that Kinande manifests genuine wh-agreement under successive cyclicity. The iterative prolepsis analysis could explain why no reconstruction effects are found in long-distance dependencies in Kinande, as the analysis does not require long-distance movement of the wh-phrase.[2]

Although nothing I have said so far about successive cyclicity rules out instances of genuine wh-agreement, it is not at all clear what role agreement would play in long-distance dependencies, especially

given that we find instances of forced anti-agreement in long-distance dependency formation. The iterative prolepsis analysis rationalizes the iterative agreement pattern we find, and hence ought to be adopted until evidence against it is found.

5.2.2.3 *More evidence against feature checking*

Bošković (2002) provides additional arguments that intermediate landing sites are not feature-checking sites. The clearest piece of evidence comes from the generalization, going back to Lobeck (1995) and Saito and Murasugi (1990), according to which ellipsis is licensed in the complement of a head taking part in Spec–Head agreement (feature checking). Although it is fair to say that such a generalization remains to be derived from some deeper principle, the evidence in its favor is rather strong. Relevant examples illustrating the generalization appear in (26)–(27). (Strikethrough indicates elision.)

(26) a. John's talk was interesting, but [$_{DP}$ Bill [$_{D'}$'s ~~talk~~]] was boring.
 b. *A single student came to the class because [$_{DP}$ [$_{D'}$ the ~~student~~]] thought it was interesting.

(27) a. John met someone but I don't know [$_{CP}$ who [$_{C'}$ C[+wh] ~~John met t~~]].
 b. *John thinks that Peter met someone but I don't believe [$_{CP}$ [$_{C'}$ that ~~Peter met someone~~]].

As we can see in the a-examples in (26) and (27), NP-ellipsis and IP-ellipsis (sluicing) are allowed to take place in situations where the head whose complement is elided is in a feature-checking relation with the element in its specifier. On the basis of this fact, Bošković argues that the unacceptability of (28) is puzzling if Spec–Head agreement (feature checking) takes place in intermediate C positions. All else being equal, we expect the complement of heads hosting intermediate traces of movement to be elidable, contrary to fact.

(28) *John met someone but I don't know who$_i$ Peter said [$_{CP}$ t_i [$_{C'}$ C ~~John met t_i~~]]

By contrast, if no feature checking takes place in intermediate landing sites, (28) is excluded on a par with the b-examples in (26)–(27).

On the basis of arguments of this sort, I conclude that the hypothesis that successive cyclic movement steps are feature driven poses more problems than it solves, not just empirically, but also conceptually, given the poorly motivated nature of the features allegedly involved. I therefore conclude that despite the problems it poses for the notion of Last Resort, one should resist the claim that intermediate movement steps are driven by feature-checking considerations. This is, to repeat, not to say that we should give up the idea that successive cyclic steps are motivated in some (other) way.

This conclusion is embraced by Chomsky in his recent writings (Chomsky 2000, 2001, 2004a, 2005, to appear). There, successive cyclic movement is forced not by feature checking, but by the Phase Impenetrability Condition (29). To avoid being trapped in the complement of a phase head, an element must move from phase edge to phase edge to reach its ultimate landing site. The mechanism that ensures this escape-hatch process is given in Chomsky (2000):

(29)   At the end of phase HP, the head H may freely be assigned an EPP-feature, forcing overt movement of a phrase into SpecHP.

There are two important concepts in Chomsky's formulation, both of which are actually characteristic of all versions of successive cyclic movement since Chomsky (1973): it is not a forced option (cf. "freely . . . assigned"), and it happens for "EPP"-reasons. As Lasnik (2001b, 2003) has emphasized, the EPP in Chomsky's system is not a feature in the technical sense of the term, the way, say, [wh]- or [ϕ]-features are; that is, things that are being valued or "checked." For instance, the [EPP]-feature can never be checked at a distance under Agree (unlike other "features"). It just seems that "the EPP . . . demands that certain functional heads have a specifier" (Lasnik 2001b). As such, it is more adequate to speak of an EPP-property, since the way to satisfy this property is quite distinct from valuing features at a distance ("checking"). Put differently, the mechanism in (29) simply states that the phase head is given the property of hosting an additional element in its edge domain. Needless to say,

the EPP so construed is no more than a descriptive characterization. But the remark just made about Chomsky's phase system suffices to illustrate my point that the hypothesis that successive cyclic movement is feature driven has receded into the background of many minimalist works, as an unrealized theoretical possibility.

As already stated in previous chapters, in many ways Takahashi's (1994) proposal, based on Chomsky and Lasnik's (1993:90) Minimize Chain Links Principle, which requires that each chain link be as short as possible, is the most principled motivation we have for successive cyclicity. Although we have found problems with Takahashi's specific proposal – such as his claim that only intermediate sites of the right A/A-bar type are targeted (chapter 3), and his claim that intermediate movement steps are not taken early, as soon as the relevant head is introduced – the logic of his approach deserves further examination. In particular, there is one aspect of Takahashi's formulation that I would like to explore in more detail in this chapter, in my attempt to determine the rationale behind successive cyclic movement.

5.3   Anti-locality

Consider again Takahashi's claim that chain links must kept as short as possible. What precisely does "as short as possible" mean?

As Bošković (1994) originally observed, some condition is needed to prevent the Minimize Chain Links Principle from forcing a phrase that has created its first chain link to keep adjoining the same node – creating infinitesimally short chain links. Put differently, some condition is needed to prevent chain links from being too short. Although Bošković was referring to Takahashi's version of successive cyclicity, his point applies with equal force to Boeckx's amendment in favor of (quasi-)uniform paths, discussed in chapter 3. Bošković (1994) noted that such a condition already existed in Murasugi and Saito (1995), who formulated the condition in (30).

(30)   A chain link must be at least of length 1.
         A chain link from A to B is of length n iff there are n "nodes" (X, X', or XP, but not segments of these) that dominate A and exclude B.

The empirical reason Murasugi and Saito gave for positing (30) goes back to an observation made in Lasnik and Saito (1992). Lasnik and Saito wanted to rule out cases of short-distance subject topicalization. They had independently argued that topicalization consisted in adjoining the topicalized material to IP, as illustrated in (31).

(31)   I think that [$_{IP}$ John, [$_{IP}$ Mary likes <John>]].

The derivation they wanted to exclude was something like (32).

(32)   *I think that [$_{IP}$ John, [$_{IP}$ <John> likes Mary]].

Lasnik and Saito (1992) observed that if (short-distance) subject topicalization were allowed, (33) would be predicted to be on a par with (34), contrary to fact.

(33)   *John$_i$ thinks that himself likes Peter.

(34)   John$_i$ thinks that himself$_i$ Peter likes t_i.

As (34) shows, topicalized anaphors can be bound by an antecedent in the next clause up. But subject anaphors cannot be so bound. This is unexpected if the derivation in (35) is available.

(35)   *John$_i$ thinks that himself$_i$ t_i likes Peter.

On the basis of such facts, Lasnik and Saito (1992) conclude that movement from SpecIP to an IP-adjoined position must be disallowed. Lasnik and Saito's conclusion is very reminiscent of Chomsky's (1986b) claim that vacuous movement (movement that crosses no overt material) ought to be banned.

In the same spirit, Murasugi and Saito (1995) claim that their condition in (30) is reducible to an economy guideline, viz. a ban on superfluous, vacuous steps. Bošković (1994) (see also Bošković 1997a:184n.28) argues that (30) has considerable motivation. In particular, he notes that (30) rules out adjunction of X to its own XP and substitution of X to SpecXP (situations that Chomsky 1995:321 referred to as "self-attachment"), as illustrated in the following figures.

In a similar spirit, Kayne (2005) has independently proposed ruling out movement of the complement of X to the specifier position
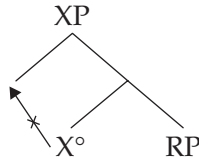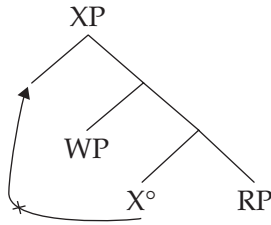
of XP, and suggests that this condition could be derived in feature-checking terms if upon Merge the maximal set of matching features must be checked. If this is indeed the case, there would be no reason for a complement of $X°$ to move to SpecXP, as there would be no feature left unchecked after the initial Merge operation.

Kayne's proposal goes beyond Murasugi and Saito's specific formulation. Combined with the latter, Kayne's proposal comes close to prohibiting movement internal to a given projection, as illustrated in the following figure.
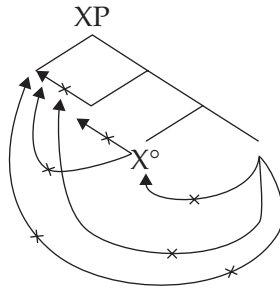
The idea that movement that is "too short" or superfluous ought to be banned has been appealed to in a variety of works in recent years. For example, Bobaljik (2000) argues, after examining a wide range of facts, that V-to-I movement never takes place in situations like (36), when VP and IP are adjacent projections.

(36)

```
        IP
       /  \
      I    VP
           |
           V
```
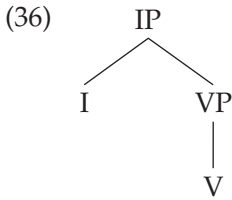
By contrast, head-movement from V to I becomes possible as soon as a projection – call it XP – separates VP and IP, as schematized in (37).

(37)

```
        IP
       /  \
      I    XP
          /  \
         X    VP
              |
              V
```
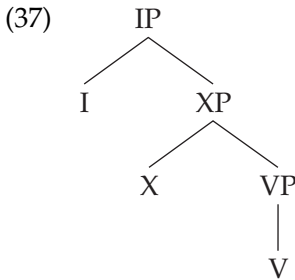
Bobaljik's claim can be deduced, as he suggests, from a ban on movement that is too short, or superfluous, if we take the label of X to be a copy of X (Harley 2004; Boeckx 2006a, 2006b; Rezac 2003). A condition in the spirit of (30) would then prohibit movement in (36), as V (*qua* VP) is already a sister of I. Head-movement from V to I would not achieve any new configuration. By contrast, no such local (sisterhood) relation exists between V and I prior to V-movement in (37), due to the presence of XP. Accordingly, movement is legitimate.

Abels (2003) provides another empirical argument against movement that is too short, or superfluous. He notes that, though mobile in general, IPs may not move across the C-head that embeds them. Consider (38).

(38) a. *[John is a fool] is believed that <John is a fool>.
      b. *[John is a fool], Mary told herself that <John is a fool>.

Likewise, Abels shows that, though mobile in principle (in some languages), VPs never strand the *v*-head that embeds them. The

generalization is harder to illustrate here, because apparent VP-movement examples exist, but, as Huang (1993) originally claimed, instances of VP-fronting are never instances of bare VP-fronting, but always consist of fronting of something bigger, which contains material that is not pronounced (such as a vP, with a null $v°$ and an unpronounced copy of the subject raised from SpecvP to SpecIP), as illustrated in (40). (I refer the reader to Huang 1993 and Abels 2003 for evidence that bare VP-fronting does not exist.)

(39)   ... [kiss Mary], John did <kiss Mary>

(40)   a.   Wrong analysis: [$_{VP}$ kiss Mary], John did [vP <John> $v°$ <[VP kiss Mary]>]
       b.   Right analysis: [vP <John> $v°$ <[VP kiss Mary]>], John did <[vP <John> $v°$ [VP kiss Mary]]>

Noting that the generalization applies to IP and VP, the complement of phases (CP and vP) in Chomsky (2000), Abels proposes the generalization in (41).

(41)   Given α, the head of a phase
       Always: *[α t]

In plain English, complements of phase heads are immobile. Abels deduces this generalization from Chomsky's Phase Impenetrability Condition. Recall that the latter makes the claim that material inside the complement of a phase head is inaccessible to further movement unless it first moves to the edge domain of the phase. Abels extends this condition to the entire complement domain of a phase, and claims that the complement of a phase is inaccessible unless it moves to the phase edge. This would in effect mean that the complement of a phase head would have to move to a specifier position of that phase head. In other words, (41), combined with the PIC, amounts to (42).

(42)   Given α, the head of a phase
       Always: *[$_{\alpha P}$ β_i [α t_i]]

(42) is just a special case of Kayne's (2005) ban on movement from complement to specifier of the same head, discussed above.

Bobaljik's and Abels's arguments make use of notions (head-movement and phase) that I have questioned at various points in this study, and, as such, they cannot be regarded as conclusive, but they are nevertheless suggestive.[3] And they are not the only arguments that have appeared recently in favor of banning phrase-internal movement. I here refer the reader to Bissell-Doggett (2004), Ko (2005), Pesetsky and Torrego (2001), Ishii (1999), Hornstein (2005), Mayr and Reitbauer (2005), Rizzi and Shlonsky (2005), Cinque (2005), Koopman (2005), Sener (2006), Den Dikken (2006), Mayr (2006), Lee (2004), and Jeong (2006).

Grohmann (2000, 2003a) is without a doubt the most systematic investigation of the claim that movement that is in some sense too short is blocked. On the basis of a wide range of considerations, Grohmann (2000, 2003a) formulates (43).

(43) *Anti-locality hypothesis*
      Movement must not be too local.

Grohmann conjectures that movement is too local if an element K has two occurrences within a given domain α (see also Ticio 2003, 2005; Henderson 2006; Cheng 2006; Schneider-Zioga to appear; Putnam 2006).[4] For Grohmann, α ranges over thematic ("*v*P"), inflectional ("IP"), and discourse-related ("CP") domains. Accordingly, no movement can take place within, say, the verbal domain (unless resumption takes place, which Grohmann takes to repair violations of anti-locality).

Grohmann's hypothesis is very desirable conceptually. It places a lower bound on movement, as Chomsky's subjacency condition places an upper bound on movement. And as locality is typically formulated in terms of domains, Grohmann formulates anti-locality in similar terms. His specific implementation of the anti-locality hypothesis in terms of three domains (roughly, *v*P, IP, and CP) has a variety of consequences in a number of domains, including successive cyclicity, to which I now turn.

5.4   Anti-locality and Successive Cyclicity

Grohmann's notion of locality bans movement not just internal to a phrase (a ban it shares with the works reviewed in the previous

section), but also internal to larger domains. I will show in this section that such a ban is too strong, and that only the weaker anti-locality condition (ban on movement internal to a projection) can be maintained.

Part of the evidence I wish to bring to bear on the issue has already been presented in the previous chapter. There I argue in the context of applicatives that the most principled account of passivization of the lower object in ditransitives consists in allowing a short movement of the lower object to a position right above the base position of the higher object. Once this position is reached, the once-lower object becomes eligible for movement to SpecTP. The derivation in (44) illustrates the relevant computational steps needed to achieve the right result.

(44)

```
                 TP
            ┌─────┴──────┐
            ↑            T′
            ┊         ┌──┴────┐
            ┊         T       vP
            ┊              ┌───┴────┐
            ┊          v-passive  vApplP
            ┊                  ┌────┴─────┐
            ┊                  DO       vApplP
            ┊                        ┌────┴─────┐
            ┊                        IO       vAppl′
            ┊                              ┌────┴─────┐
            ┊                            vAppl        VP
            ┊                                     ┌────┴────┐
            ┊                                     V      t (DO)
            └──────────────────────────────────────────┘
```
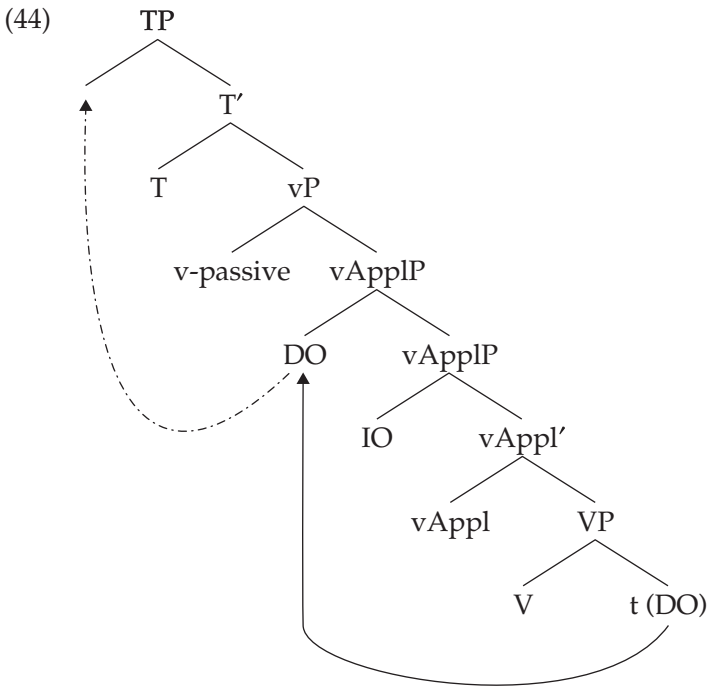
If the derivation depicted in (44) is correct, Grohmann's ban on movement internal to the thematic domain (*v*P) cannot be maintained.

The evidence against Grohmann's hypothesis does not stop here. As we saw in the previous chapter, empirical considerations have

led several researchers to claim that we need to distinguish at least two kinds of double object structures: a low applicative structure like (45), and a high applicative/dative structure like (46). (I have modified the structures in (45) and (46) slightly, eliminating structural details that are irrelevant to the issue at hand.)

(45) $[_{vP}$ Subj V $[_{VP}$ IO V DO]]

(46) $[_{vP}$ Subj V $[_{ApplP}$ IO V $[_{VP}$ V DO]]]

Those researchers have reached a conclusion that is puzzling at first. McGinnis (2001), in particular, has provided compelling empirical evidence suggesting that languages employing the structure in (46) correspond to what previous research had called symmetric languages, that is, languages that treat both objects alike for a variety of syntactic purposes such as passivization, cliticization, etc. McGinnis's conclusion amounts to this surprising statement: the closer the objects are structurally upon base Merge, the more asymmetrically they will behave at subsequent stages in the derivation. Conversely, the more distant they are upon base Merge, the more symmetrically they will behave. From a phrase-structural perspective, however, one would expect that greater distance between the two objects would increase the asymmetry between them.

   As a matter of fact, Hornstein (2005) (building on Chomsky 1993) has pursued the idea that elements internal to a projection are all "equidistant." But natural languages just do not seem to work that way. From the point of view of locality (specifically, relativized minimality/intervention effects), it is easy to understand why IO is always passivizable (*pace* independent factors that might block such movement, such as inherent case on the higher object in languages that lack non-structurally case-marked subjects; see Jeong 2006 for extensive discussion); IO always start off higher than DO, that is, closer to T°. As we saw in (44), DO is allowed to circumvent the intervening higher object via early successive cyclic movement. But, interestingly, such a derivation appears to be confined to high applicative structures. DO-passivization in lower applicative structures is severely restricted. As Jeong (2006) has shown, it is confined to those instances where the higher object is ineligible for movement (for a variety of reasons that are language-specific, and that I won't discuss here). Crucially, for our purposes, there is no evidence that

in such cases of DO-passivization, early successive cyclic movement is needed. There is, in fact, reason to believe that early successive cyclic movement of the type schematized in (44) is unavailable in low applicative structures. The reason is anti-locality.

As Lee (2004) and Jeong (2006) observe, a derivation like (44) in the context of low applicatives would amount to something like (47).

(47)

```
                    TP
                   /  \
                  /    T′
                 /    /  \
                /    T    vP
               ↑        /   \
               |       v-passive  VP
               |               /    \
               |             DO      VP
               |             ↑      /   \
               |             |    IO     V′
               |             |         /   \
               |             |        V°    t (DO)
               |             |_____/
               |_____/
```
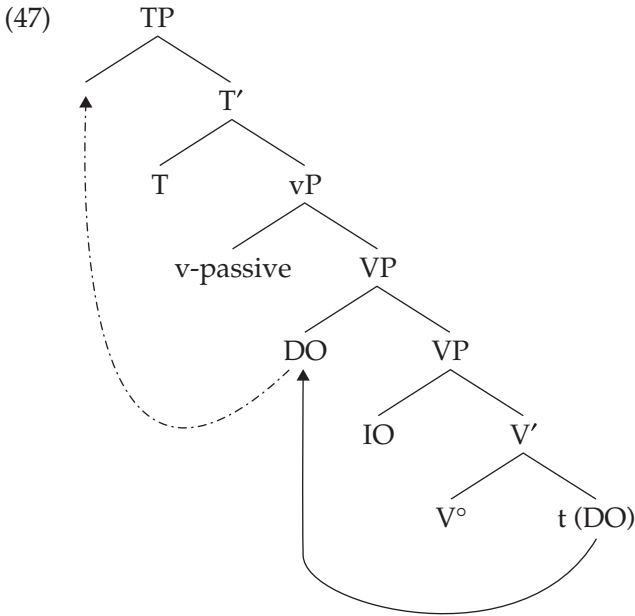
The derivation in (47) requires the first step of movement to be internal to the projection in which DO is merged. If such a movement is blocked because it is, in a sense, too short, or superfluous, it will readily explain why DO-passivization is typically banned in low applicative structures. This, in turn, strengthens our analysis in (44), and the idea of early successive cyclic movement argued for in the previous chapter.

The analysis just presented also argues against Grohmann's domain-based conception of anti-locality. If, as Grohmann claims, movement within *v*P (the thematic domain) were banned, one would not be able to distinguish between symmetric and asymmetric languages. In fact, Grohmann predicts that all languages are of the asymmetric type, since the movement step that obviates

minimality – "leapfrogging," in McGinnis's terms – would violate Grohmann's notion of anti-locality. (I see no way, under Grohmann's approach, of capturing symmetric languages in a way consistent with minimality.)

The full range of applicative data discussed here demands a notion of anti-locality that confines the effect of the condition to a single projection (as opposed to involving the notion of "domain") – something like (48).

(48)   The complement of X cannot move anywhere within XP.

More generally, we can formulate the anti-locality hypothesis as in (49).

(49)   Movement internal to a projection counts as too local, and is banned.

The next obvious question is: what is the rationale for a condition like (49)?

5.5   Anti-locality and Last Resort

Following Murasugi and Saito's (1995) lead, I would like to argue that anti-locality (as defined in (49)) makes eminent sense from a minimalist point of view. To make my case, I need to say a few words about Bare Phrase Structure, the minimalist treatment of Phrase Structure.

Most current views on Phrase Structure rely on the fundamental insights expressed in "Remarks on nominalization" (Chomsky 1970). "Remarks" made three basic claims (see Fukui 2001 for detailed overview), listed in (50).

(50)   a.   Every phrase is "headed"; i.e., it has an endocentric structure, with the head X projecting to larger phrases.
       b.   UG provides a general X-bar schema of the following sort, which governs the mode of projection of a head:
            $X' \rightarrow \ldots X \ldots$
            $X'' \rightarrow \ldots X' \ldots$

This gives rise to the now familiar X-bar schema:

(51)   XP

(Diagram showing XP branching to X', and X' branching to X)
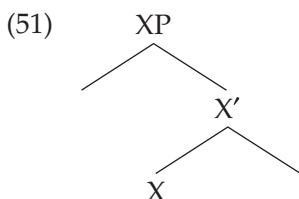
Chomsky (1994) formulated a Bare Phrase Structure theory as part of the minimalist program. In particular, Chomsky has argued against a rigid definition of the three juncture types X, X', and XP, and in favor of a more relational definition of them. Chomsky's proposal is best understood in the context of the following question, stated clearly in Hornstein et al. (2006:ch. 6): how should we conceptualize the difference between X, X', and X"? The "rigid" way of conceptualizing them would be to claim that they have different intrinsic features, the way, say, Nouns and Verbs have. Alternatively, they may differ in virtue of their relations with elements in their local environment, rather than inherently. On the first interpretation, bar levels are categorical features; on the second, they are relational properties.

Chomsky (1994) claims that a Bare Phrase Structure theory should treat bar levels as relational properties. Specifically, the following relations ought to be recognized:

(52) a. *Minimal projection* ($X^{(0)}$)
          A minimal projection is a lexical item selected from the lexicon.
      b. *Maximal projection* (X")
          A maximal projection is a syntactic object that doesn't project any further.
      c. *Intermediate projection* (X')
          An intermediate projection is a syntactic object that is neither minimal nor maximal.

The relational view on bar levels immediately eliminates "spurious" or "vacuous" projections that were standard under the rigid view. According to the latter, a bare element like *John* invariably projected as in (53).

(53)   NP
         |
        N′
         |
        N
         |
       John
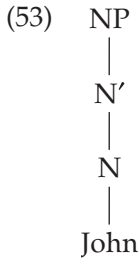
Under the relational view, an element like *John* is ambiguously min-
imal and maximal, its status being determined only once it enters
into a syntactic relation, since bar levels are reflexes of the position
of a syntactic item with respect to others. Put differently, under a
relational or "Bare" Phrase Structure theory, there can be no non-
branching projections, since projection is a reflex of syntactic com-
binations, and those by definition necessitate two elements.

The key aspect of Bare Phrase Structure for present purposes is
the fact that there is no intrinsic distinction between specifiers and
complements. Such terms are mere descriptive devices, akin to the
status constructions like passive and raising have in modern gen-
erative grammar. Accordingly, in such a framework, moving a
complement of $X°$ to the specifier of XP would be an instance of
superfluous, vacuous movement, a violation of Last Resort.
Nothing (barring look-ahead) would be beneficial for either the mover
or the host of movement. No new phrase-structural status is
acquired by performing such a movement. So the anti-locality con-
dition on movement follows immediately from minimalist guide-
lines, specifically from Last Resort considerations. Put differenty,
minimalism allows us to explain why the first intermediate step of
movement must target a position outside of the projection from which
the moving element originates.

We have thus reached a stage where we have shown the work-
ings of Last Resort on successive cyclicity despite rejecting the idea
that intermediate movement steps are not driven by immediate
feature-checking considerations. Admittedly, the conclusion we have
reached still does not not answer the question of why intermedi-
ate steps are taken. It only answers the question of why some inter-
mediate steps are disallowed. It is now time to address the
why-question.

5.6   The *Why*-Question

It may be helpful to take stock of the conclusions we have reached so far, before proceeding.

I have argued in this and the previous chapters that there is ample evidence that long-distance dependencies are not taken in one fell swoop. It appears that movement cannot be too long. For some reason, the links of a chain must be kept short. Since I found no reason to abandon the claim that movement paths are (quasi-)uniform, I have adopted it: once movement is launched, it targets all possible landing sites along the way. I have found empirical evidence that movement is launched as soon as possible, as soon as the first target of movement is created. In this chapter, I have argued that the first target of movement cannot be internal to the same projection in which movement was launched. Last Resort bans gratuitous remergers. Accordingly, an element cannot merge with the same head twice.

Notice that anti-locality considerations of the type discussed in this chapter are likely to prove problematic for uniform path hypotheses of the type entertained in HPSG (cf. chapter 3), since under such hypotheses the "gap" information (slash-category) cannot skip any node, and is therefore found on multiple nodes internal to a projection. Accordingly, anti-locality may provide an argument in favor of quasi-uniform path hypotheses.

At this point, I would like to suggest a partial reconsideration of Last Resort within minimalism. Specifically, I would like to propose that intermediate steps of movement are taken not because they are triggered by some feature, but because they are not forbidden. In particular, intermediate movement steps must be compatible with Last Resort considerations: the moving element cannot take superfluous steps. In addition to banning movement steps that would be too short, compatibility with Last Resort also means that the moving element must contain within its feature set at least one feature that remains unchecked at the point at which movement is launched. Given that I take the operation Agree to be part of the grammar, this means that at the point when movement is launched, no Probe capable of checking the moving element's unchecked feature at a distance must be present in the tree.

Intermediate movement steps must also be compatible with locality conditions on the grammar. Locality conditions have been subject to debates since Ross's landmark (1967) study. Although we have made progress in establishing the kind of locality conditions human languages are sensitive to, exactly which locality conditions are the right ones remain a subject of debate (a topic I return to in chapter 7).

The only locality condition one can be pretty confident about is adjacency, expressed phrase-structurally in terms of sisterhood. Theta-role assignment appears to require such a notion, and no other. Sub-categorization considerations likewise are best expressed in terms of structural adjacency. I would like to claim that the length of intermediate movement steps argued for in this book is also a reflex of adjacency, albeit viewed not horizontally (sisterhood), but vertically; in terms of "next projection." An argument must be as close as possible to a verb to receive a theta-role – it can't be inside the verb (due to a condition of lexical integrity of sorts) – and so it must be the sister of the verb; similarly, movement must target a position close enough to its launching site. Since movement internal to a projection is banned, the moving element must target the very next structural possibility available: the next projection up, and so on, iteratively, until the element has checked all its features.

In many ways, vertical adjacency is what Chomsky (1973) dubbed subjacency. Chomsky defined subjacency in terms of layers of clausal embedding, not in terms of projection. But defining subjacency in terms of projection is just a more general, comprehensive version of the original definition of terms of clauses, just as our modern version of Merge is but a more general, comprehensive version of Chomsky's (1955, 1957) Generalized Transformation operation (also formulated in terms of clauses, or kernel sentences.) To put this differently, just as the elements of Merge must be close enough (under the same projection), so the elements of move (the copies or "occurrences" of a given element) must be close enough (under adjacent projections). The conclusion we appear to have reached is that if movement is launched, it must be compatible with locality and Last Resort, but the launching action is not forced by them. It is merely given as a derivational option. In the absence of a Probe capable of checking all the unvalued features of an element, the system offers that element the possibility of moving, as long as no principle of the grammar is disallowed.

This state of affairs will no doubt look to many like a return to the GB-notion of movement as part of the overarching operation Affect α. The latter stood for the idea that if something isn't disallowed, it can be made use of. Minimalism counterbalanced this intuition by imposing a Last Resort condition on syntactic operations. Minimalism adopted the following admonitions:

(i)   If you need (locally), you must.
(ii) If you must not (locally), you cannot.

I say "admonitions," and not principles, because, as Chomsky has always stressed, notions like economy are guidelines for specific analyses, not principles of actual theories. It would therefore be wrong to block intermediate movement steps if no feature-checking operation can be found to legitimize them. All that is required is that the operation be compatible with economy guidelines.

It is also important to understand guidelines like Last Resort in a local manner. If at the point at which movement is launched, no Probe is "in sight," the moving element does not know if there will be a Probe "down the road" ("up the tree"), inserted in subsequent derivational steps. But even if the insertion of a Probe were somehow "guaranteed," the moving element cannot know whether the Probe will merely require an Agree relation, or will have an EPP-requirement forcing movement. If EPP-requirements didn't exist, no movement would be allowed, as all checking could be done long-distance (ignoring intervention effects, which may be obviated by altering phrase structural relations, as we saw in derivations like (44)). Since EPP-requirements exist (whatever their ultimate nature), launching movement is indeed a way of maximizing derivational options, not so much for the moving element as for the Probe to be able to meet its EPP-requirement, if it has one. Absent movement, and given locality (*qua* adjacency/subjacency), the Goal may be too far away from the Probe to move into the latter's specifier.

I conclude therefore that if the domain within which an element finds itself does not contain a Probe, the system allows remerger of the element in the next higher projection, because at that point the element doesn't know whether it will participate in an Agree relation. Movement is thus allowed "agnostically," as a way of maximizing derivational options. (The term "agnostic movement"

is borrowed from Franks and Lavine to appear, to which I return in the next chapter.) In other words, just as Chomsky (2000) claimed that matching effects (checking relations) ought to be maximized, I am suggesting that displacement may take place to maximize subsequent derivational alternatives (about which nothing is known at the point where movement is taken, no look-ahead being involved). To put this yet another way, I am claiming that in addition to the Last Resort admonitions highlighted by the minimalist program, we need to recognize the following guideline:

> If you don't know (locally), you can do everything you can (so as to maximize your options).

Put differently, if you don't know that you "must not," you don't know you "cannot," hence you "can." Although this may signify a departure from the minimalist guidelines advanced in the literature, it is fully compatible with minimalist tenets, and in no way undermines the Last Resort character of syntactic operations which is essential in explaining the anti-local nature of movement (ban on vacuous movement) discussed in this chapter.

Movement is, thus, in a sense, optional. The characterization of symmetric applicative structures discussed in the previous chapter indicates that this is the right conclusion empirically. Although the lower object can move higher than the applied object, and thereby become eligible for passivization, such movement need not take place. If it doesn't take place, passivization of the applied object remains the only derivational option.

The system just sketched is fully compatible with Frampton and Gutmann's (1999, 2002) idea that minimalist derivations are crash-proof (computationally efficient). Frampton and Gutmann claim that computationally efficient derivations rule out comparisons among derivations, look-ahead, backtracking, and derivational impasses. A crash-proof syntax requires that whatever derivational option be taken, there should exist a way of making that derivation ultimately converge. In the context at hand, it means that a crash-proof syntax must ensure that once movement is launched, a Probe be merged at a subsequent derivational stage. In some cases, this will mean that at a given stage in the derivation, various elements of the same type will be moving, but I don't see this as a disadvantage, as such situations are attested in natural languages (absorption phenomena

in multiple wh-questions, composed chains in parasitic gaps, multiple scramblings, etc.), and various devices already exist in the literature to preserve relative hierarchy among moving/"active" elements (see N. Richards 2001; Jeong 2004; Williams 2003), or ensure that only one of the moved elements gets pronounced in the moved position (see Groat and O'Neil 1996, Bobaljik 2002, Pesetsky 2000, and Bošković 2001, for various cases of moved elements that are pronounced as if they hadn't moved).

5.7 Conclusion

I have no doubt that the hypothesis proposed here will require detailed empirical considerations before it can be adopted, but I hope to have shown that conceptually it is certainly a theoretical option that exists within the minimalist program. Since, as its name indicates, minimalism is a program, not a theory, it is an option that one ought to examine. As Chomsky (2000) stated, "there are minimalist questions, but there aren't any minimalist answers, except those that emerge from pursuing the program." As Avery Andrews once said, "there is little point in doing theoretical work if one is not willing to be surprised by the analysis one eventually comes up with."

The minimalist question I started off with in this chapter was: what is the motivation for intermediate movement steps? I have shown that in addition to locality considerations that have long been involved in addressing this question, anti-locality considerations allow us to reveal one key aspect of successive cyclicity. Having shown intermediate movement steps to be fully in line with locality and Last Resort, I have proposed that movement steps are taken to maximize options that may be beneficial at later derivational stages.

Notes

1 I here set aside differences among the relevant languages pertaining to whether "wh-agreement" on intermediate verbs is obligatory (Tagalog) or optional (Chamorro) depending on the nature of the moving wh-phrase (D-linked or not).
2 Although I hasten to add that the prolepsis approach is not incompatible with movement (for movement-based approaches to pronoun–

antecedent relations, see Kayne 2002 and Boeckx 2001a, 2003a). If movement is adopted, absence of reconstruction must be explained in some other way.

3   For alternative analyses that retain Bobaljik's and Abels's generalizations without appealing to some of the problematic concepts just mentioned, see Alexiadou and Fanselow (2000) and Wurmbrand (2004).

4   Grohmann notes that just as movement that is "too long" can be "salvaged" (impressionistically speaking) by resumption (i), movement that is too short can, too. For justification of the derivations in (ii), see Grohmann's own work.

(i)   ?Which woman did you claim that Peter met the man who saw ~~<which woman>~~ *her*?

(ii)   a.   John [$_{VP}$ <John> likes ~~<John>~~ → himself].
       b.   [$_{XP}$ Diesen Mann, [$_{XP}$ ~~<diesen Mann>~~ → den
             this.acc man   that-one.acc
             kenne ich nicht].   German
             know I   not
             'This man, I don't know him.'

For alternative views on resumption (including reflexivity) that don't require Grohmann's specific version of anti-locality, but instead rely on agreement and case, see Boeckx (2001a, 2003a) and Hornstein (2001), respectively.

Chapter 6

Alternative Views on Successive Cyclicity

I do not want to give the reader the impression that the view on successive cyclic movement developed in the previous chapters is the only minimalist way of analyzing long-distance dependencies. The reader should always bear in mind that minimalism is a research program, open to many alternative formulations. In this chapter I would like to consider alternative ways of approaching the phenomenon of successive cyclic movement. As we will see, some of these alternatives share a few aspects with the proposal I have defended in this book; others depart more radically from the axioms adopted here. After sketching the main features of each alternative, I will indicate why I think these alternatives are not superior to the account proposed here.

So far, the only alternative account of successive cyclic movement I have discussed is the standard phase-based account. I hope that in considering, albeit briefly, other, less often-discussed, research paths, the reader's critical eye will be sharpened. I will start with accounts that are furthest from the account I developed, working my way toward an account that comes closest to my own proposal.

6.1   TAG-based Accounts

Castillo and Uriagereka (2002) (see also Drury 2005) attempt to integrate the insights of TAG-grammar (Kroch and Joshi 1985;

Frank 2002) into the minimalist program and propose that lexical insertion proceeds in a fashion that N. Richards (1997, 2001) calls "tucking-in." Richards's tucking-in operation is best illustrated by comparing the two notions of cyclicity already discussed in earlier chapters – the Extension condition and featural cyclicity – in the context of multiple movements targeting the same projection.

The idea behind the Extension condition is extremely simple: operations always extend their target (Chomsky 1993:190). Accordingly, movement should always make the tree "bigger," not "thicker." Featural cyclicity works differently. It says that a strong feature must be checked as soon as possible after being introduced into the derivation (i.e., before the head containing the strong feature is embedded).

The two notions of cyclicity make different predictions in a number of domains. One of them concerns situations of multiple movement. Consider a situation in which two elements α and β relate to and move to a target (head of projection) γ. If movement is constrained by minimality (attract closest X), Extension predicts the derivation in (1), in which movement of α (AP) happens first, and movement of β (BP) next. Crucially, movement of BP necessarily targets a position higher than the landing site of AP.

(1)
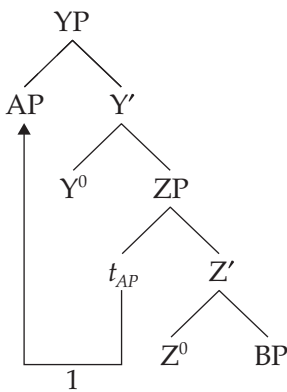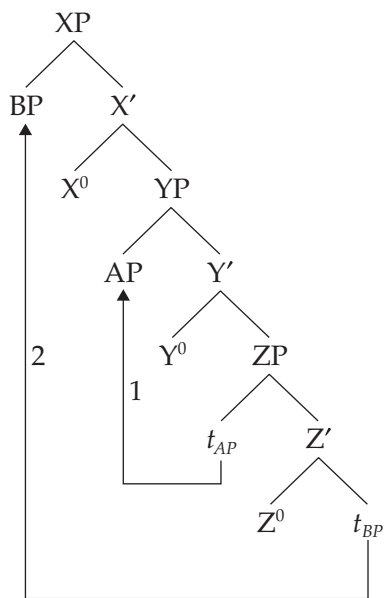
In other words, Extension requires that movement paths nest, as schematized in (2).

(2) $[_{XP}$ BP $[_{XP}$ AP $[_{X'}$ X^0 $[_{YP}$ t_{AP} $[_{Y'}$ Y^0 $[_{ZP}$ t_{BP} $[_{Z'}$ Z^0 . . .]]]]]]] *nested paths*
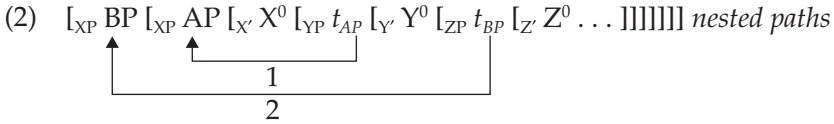
Things are different if Featural cyclicity is involved. Featural cyclicity is compatible with the derivation in (1), but it is also compatible with the derivation given in (3), in which the second movement (movement of BP) targets a position higher than the attractor, but lower than the position occupied by AP.

(3)
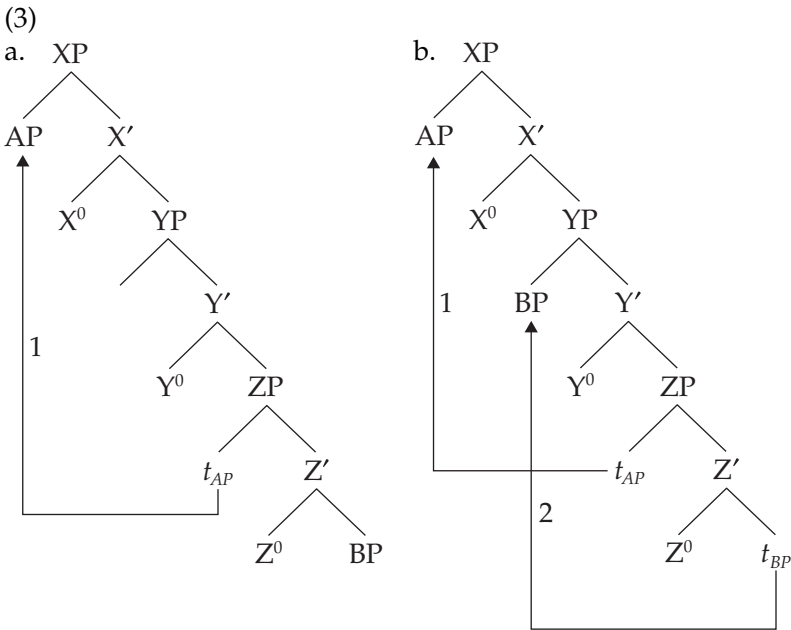
In other words, Featural cyclicity allows movement paths to cross, as represented in (4).

(4)
$[_{XP}$ AP $[_{XP}$ BP $[_{X'}$ X^0 $[_{YP}$ t_{AP} $[_{Y'}$ Y^0 $[_{ZP}$ t_{BP} $[_{Z'}$ Z^0 . . .]]]]]]] *crossing paths*
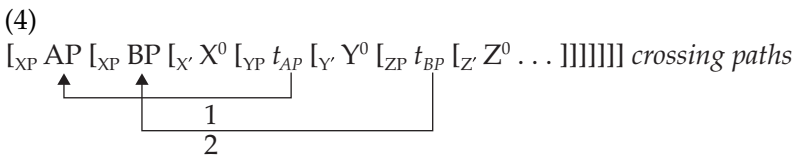
The derivation in (4) is what Richards calls "tucking-in." Richards argues that Featural cyclicity, once combined with the requirement that movement be as short as possible, makes the prediction that movement of multiple specifiers to a single head is forced to "tuck in," i.e., that movement paths cross.

The definition of Shortest that Richards adopts is given in (5).

(5) *Shortest*
A pair P of elements (α, β) obeys Shortest iff there is no well-formed pair P' that can be created by substituting γ for either α or β, and the set of nodes c-commanded by one element of P' and dominating the other is smaller than the set of nodes c-commanded by one element of P and dominated by the other.

This definition of Shortest ensures that in the case of multiple movement to X (target; head of projection), the closest matching element will be moved, followed by movement of the next matching candidate. Crucially, because each movement should cross as few nodes as possible, the second movement will "tuck in," as schematized in (6).

(6) a. $[_{XP}$ YP $[_{X'}$ X $[_{WP}\ldots t_{YP}\ldots [\ldots t_{ZP}\ldots]]]]$

 b. $[_{XP}$ YP $[_{XP}$ ZP $[_{X'}$ X $[_{WP}\ldots t_{YP}\ldots [\ldots t_{ZP}\ldots]]]]]$ *crossing paths*
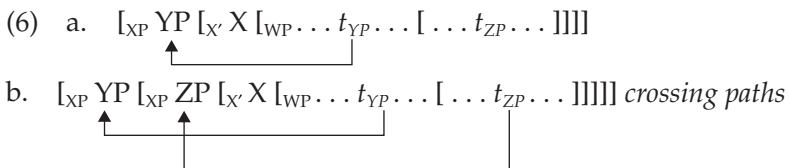
Thus, Featural cyclicity *plus* Shortest predicts that movement to multiple specifiers of a single head should maintain the c-command relation that obtains between the moving elements prior to movement to the relevant target.

Richards offers several pieces of evidence that Featural cyclicity + Shortest is empirically adequate. Let me briefly discuss one case: multiple A-bar movement. The starting point of Richards's demonstration is the well-known superiority effect, here illustrated on the basis of English wh-movement (7) and Japanese wh-scrambling (8).

(7) a. Guess who *t* bought what.
  b. *Guess what who bought *t*.

(8) a. Dare-ni John-ga [Bill-ga *t* [Mary-ga nani-o
who-dat John-nom Bill-nom Mary-nom what-acc
tabeta to] itta to] omotteiru no?
ate   that said that thinks   Q
'To whom did John think that Bill told that Mary ate
what?'

b. *Nani-o John-ga [Bill-ga dare-ni [Mary-ga *t* tabeta to] itta
to] omotteiru no?

Building on work by Rudin (1988) and Bošković (1998), Richards
shows that in the case of multiple wh-movement in languages like
Bulgarian, wh-movement is constrained by superiority: the highest
wh-phrase prior to movement must be the first to move, and the
second movement tucks in. This is illustrated in (9) for Bulgarian.

(9) a. Koj$_i$ kogo$_j$ vidjal t_i t_j?
who whom saw
'Who saw whom?'

b. *Kogo$_j$ koj$_i$ vidjal t_i t_j?

c. Kogo$_i$ kakvo$_j$ e pital Ivan t_i t_j?
who   what   is asked Ivan
'Who did Ivan asked what?'

d. *Kakvo$_j$ kogo$_i$ e pital Ivan t_i t_j

Similarly for multiple wh-scrambling, illustrated in (10) for
Japanese.

(10) a. Dare-ni$_i$ nani-o$_j$ John-ga   [Tanaka-sensee-ga
who-dat what-acc John-nom Tanaka Prof-nom
t_i t_j yomaseta to] itta no?
read.caus   that said Q
'To whom did John say that Prof. Tanaka made read
what?'

b. *Nani-o$_j$ dare-ni$_i$ John-ga [Tanaka-sensee-ga t_i t_j yomaseta
to] itta no?

The conclusion Richards draws from such examples is that in the
case of multiple A-bar-movements to a single landing site, the
paths must cross. This is correctly derived by the combination of
Featural cyclicity and Shortest Move.

Having clarified the notion of tucking-in, let us now go back to successive cyclicity. Castillo and Uriagereka (2002) propose that wh-movement is clause-bounded (see also Frank 2002), the appearance of long-distance movement being the result of insertion of clausal material below the moved element, as illustrated in the derivation in (11). (I ignore the phenomenon of I-to-C movement and *do*-support. Castillo and Uriagereka are not specific about this, and could easily assume it to be a fact determined post-syntactically, in the PF-component.)

(11)   Step 1: Who you Past talk to <who>?
       Step 2: Who {John Past say that} you talked to <who>?
       Step 3: Who {Bill Past ask that} John said that you talked to <who>?

More graphically, for part of the derivation:
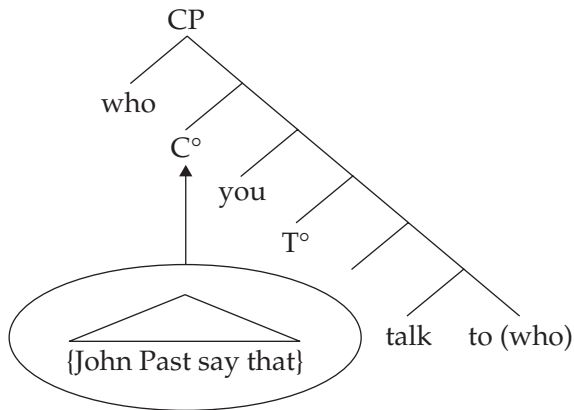
Facts like the wh-agreement data in Kinande discussed in the previous chapter, especially when viewed under the iterative prolepsis analysis I proposed, appear to fit Castillo and Uriagereka's proposal like a glove, since, under such an analysis, successive cyclic movement is an illusion, resulting from very local clausal sub-embedding.

An obvious question that arises in the context of Castillo and Uriagereka's proposal is what drives lexical insertion. Castillo and Uriagereka assume that merger takes place in an arbitrary fashion, bad instances being filtered out by selectional, thematic, or case/agreement requirements.

Though ingenious, Castillo and Uriagereka's analysis suffers from the fact that we saw evidence that the cycles for movement are very small. In particular, there is evidence (reconstruction effects, quantifier floating, past participle agreement etc.; see chapter 2) suggesting that a wh-element stops by at least some position above its launching site and below CP (the final landing site of movement under Castillo and Uriagereka's analysis). Put differently, there is evidence for successive cyclic movement not just across clauses (a fact Castillo and Uriagereka's analysis is designed to capture), but also clause-internally.

How to capture this fact under the approach under discussion is not obvious, since under a TAG-style analysis, part of the domain within which movement takes place would have to be inserted (counter-cyclically) after movement, if intermediate landing sites amount to acyclic insertion under tucking-in. Castillo and Uriagereka do not address this problem, as they assume that the first landing site (for a wh-word) is SpecCP, but, given the evidence in the literature for multiple clause-internal intermediate landing sites, I regard this as a serious problem.

6.2   An Agreement-based Account

Rackowski and Richards (2005) develop a theory of wh-extraction which crucially involves Probes agreeing with multiple Goals, a phenomenon already discussed under the term "concord" in chapter 5. The theory is meant to deal with the phenomenon of wh-agreement, discussed at various points in previous chapters, and also with islandhood.

Rackowski and Richards's discussion is based on facts from Tagalog. The reader will recall from the previous chapter that in this language, extraction of a DP requires the verb to agree with that DP:

(12) a. Sino ang   nagbigay   ng bulaklak sa   kanya?
          who ANG **NOM**-gave NG flower   DAT 3
          'Who gave him/her the flower?'
      b. Sino ang   binigyan mo   ng bulaklak?
          who ANG **DAT**-gave NG-you NG flower
          'Who did you give the flower to?'

     c. Ano ang ibinigay mo   sa   kanya?
        what ANG **OBL**-gave NG-you DAT 3
        'What did you give him/her?'

In cases of wh-movement across clause boundaries, the verb of the most embedded clause, from which movement was launched, must still show agreement with the extracted phrase. All higher verbs must agree with the clause from which extraction is taking place:

  (13) a. Sino ang sinabi   ng magsasaka [na kumain
         what ANG **ACC**-said NG farmer   that **NOM**-ate
         ng bulaklak]?
         NG flower
         'Who did the farmer say ate the flower?'
      b. Ano ang sinabi   ng magsasaka [na kinain
         what ANG **ACC**-said NG farmer   that **ACC**-ate
         ng kalabaw]?
         NG water-buffalo
         'What did the farmer say the water-buffalo ate?'
      c. *Ano ang sinabi   ng magsasaka [na kumain
         what ANG **ACC**-said NG farmer   that **NOM**-ate
         ang kalabaw]?
         ANG water-buffalo
      d. *Ano ang nagsabi   ang magsasaka [na
         what ANG **NOM**-said ANG farmer   that
         kinain   ng kalabaw]?
         ACC-ate NG water-buffalo

(13a–b) show extraction of an embedded subject and an embedded object, respectively. In both, the higher verb *sinabi* 'ACC-said' agrees in case with the complement clause (and crucially not with the extracted wh-phrase). The embedded clause, on the other hand, has a verb which does agree in case with the extracted phrase: *kumain* 'NOM-ate' for subject extraction, and *kinain* 'ACC-ate' for object extraction.

   Rackowski and Richards account for this pattern of facts by positing a version of locality which makes crucial use of Chomsky's (1964) A-over-A condition, such that (according to Rackowski and Richards) when a wh-phrase is embedded in a CP, the CP will be closer to Probes outside the CP than the wh-phrase will. Extraction of the relevant DP is therefore expected to be impossible, contrary

to fact. To allow for extraction, Rackowski and Richards build on N. Richards's (1997, 1998, 2001) Principle of Minimal Compliance (PMC), given in (14).

(14)   *Principle of Minimal Compliance*
      For any dependency D that obeys constraint C, any elements that are relevant for determining whether D obeys C can be ignored for the rest of the derivation for purposes of determining whether any other dependency D' obeys C.

The net result of the PMC is that once the Probe has agreed with this closer potential Goal, it is free to Agree with Goals that are further away. Once the DP has entered into an agreement relation with v, it is allowed to move to Spec$v$P, and from there on to its final landing site. As the PMC must apply at each clausal juncture, successive cyclicity emerges, albeit in a slightly different guise from the typical conception of successive cyclic movement, since, for Rackowski and Richards, Spec$v$Ps, and not SpecCPs, or any other Specs, act as intermediate landing sites.

The following figure schematizes a long-distance movement derivation along the lines of Rackowski and Richards's account.
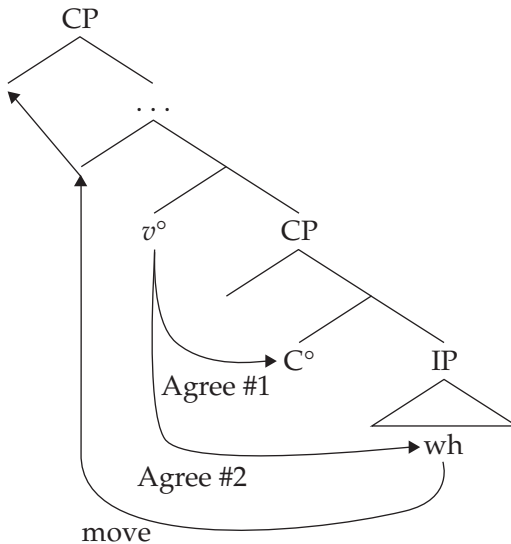
The upshot of Rackowski and Richards's analysis is that in order for the v of the matrix clause to Agree with the wh-phrase, causing

it to move out of the embedded clause, the matrix v must first Agree with the embedded CP. This has two consequences. First, in Tagalog, the first Agree relation determines the morphological form of v, correctly giving the result that extraction from an embedded clause will require v to Agree with that clause. Second, Rackowski and Richards argue that their approach yields a version of various island effects. Specifically, for them, islands are those domains with which v cannot enter into an Agree relation. In short, Rackowski and Richards (2005) claim that movement across a clause boundary involves two Agree relations by v, one with the embedded clause, and a second one with the moving XP. Local extraction, by contrast, involves only a single Agree relation with v.

Although Rackowski and Richards elegantly capture the Tagalog facts, it is unclear how their analysis would extend to what we may call wh-anti-agreement languages like Selayarese, where clausal agreement along the movement path must be dropped for extraction to be allowed. It is also not at all clear why agreement with v is so crucial to A-bar extraction. As already pointed out in chapter 5, the morphology on Tagalog verbs in the context of extraction need not be equated with agreement for case. The morphology may simply encode thematic relations, or inherent case marking. This would, in fact, make more sense than the agreement view Rackowski and Richards adopt, as there is a fair amount of evidence (reviewed in detail in Boeckx 2003a) that agreement between, say, a verb X and an element Y bleeds (and not feeds) sub-extraction from Y. (For the role of anti-agreement in extraction, see Boeckx 2003a. See also chapter 7.)

Perhaps, then, Tagalog wh-agreement really amounts to anti-phi-agreement, as inherently case-marked or theta-marked elements typically fail to enter into phi-agreement relations. This would make Tagalog identical to Selayarese. Furthermore, Rackowski and Richards' analysis crucially depends on the Principle of Minimal Compliance, which is a poorly understood principle that doesn't seem to make a lot of sense, as it allows for any constraint to be violated by X, once the constraint has been satisfied by Y. Finally, there is also a fair amount of detailed, careful evidence that intermerdiate SpecCPs are targeted by successive cyclic movement (see Den Dikken 2006 for recent evidence), so Rackowski and Richards' theory of impoverished intermediate landing sites does not appear to be empirically adequate.

6.3   Prolific Domains

Grohmann (2003a, 2003b) defends an approach to successive cyclicity within his framework of prolific domains. Like the approach developed in this book, Grohmann's takes intermediate steps of movement not to be driven by featural requirements. Also like the present approach, Grohmann's makes use of anti-locality considerations in the context of movement.

Grohmann builds successive cyclicity into his definition of locality. His take on locality is domain driven. He assumes that a clause is divided up into three domains (Θ-,Φ-, and Ω-domains; roughly, *v*P, IP, and CP) which constitute cycles. He furthermore assumes that as a matter of locality/cyclicity, an element is required to pass through each domain on its way to its final landing site, except the domain from which movement was launched – for anti-locality reasons. Although Grohmann avoids the need for spurious features driving movement, his approach crucially relies on axiomatic concepts like domain and locality which one would like to derive, and not merely stipulate. At the moment, his approach suffers from some of the problems that plague phase-based derivations. Just as the identity of phases is a fact we would like to understand, the identity of domains and their properties remain generalizations to be deduced.

Furthermore, by forbidding movement internal to each prolific domain, Grohmann offers an impoverished theory of intermediate landing sites, which appears quite problematic in light of the fact that movement internal to *v*P (and presumably other domains) appears necessary to capture facts about natural languages (cf. chapter 4).

6.4   Greed-based Approaches

The remaining accounts of successive cyclic movement I would like to review all share the idea that intermediate movement steps are triggered by the fact that the moving element moved because it had an urge to do so (a view that Chomsky 1993 called Greed). The moving element was, as it were, unhappy in situ, and initiated movement as an attempt to get out of that situation.

This was exactly the intuition, first formulated by Myriam Uribe-Etxebarria (p.c. to Juan Uriagereka) that Lasnik et al. (2005:ch. 7) exploited to sketch a motivation for successive cyclic movement. According to them, unchecked features on an element, such as an unchecked wh-feature on a wh-phrase, are like selectional features in need of an appropriate selector. In the absence of such a selector in its local domain, the element moves to a position where its unchecked feature would typically be satisfied (say, an A-bar position like SpecCP). If the CP it lands in doesn't have the right value (if it is not an interrogative CP), the element finds itself again in a misplaced position, and keeps moving until it finds an appropriate landing site.

Franks and Lavine (to appear) develop an account along similar lines, although their system is phase based. They claim that if an element fails to have all its features checked by the time the phase that contains it is completed, the element will move, in an agnostic fashion, in the hope that this movement will make it accessible to an appropriate Probe. As in Lasnik et al.'s account, no feature other than the unchecked feature on the moving element is driving movement.

Finally, Bošković (2006) proposes that successive cyclicity follows from properties of the PF-component; more precisely, from the fact that linearization takes place in a cyclic or phase-based fashion (in the spirit of Fox and Pesetsky 2004). Specifically, Bošković argues that the key function of phases is to establish information relevant to the PF-component, so that linearization can take place in a cyclic fashion. Bošković notes that if we adopt Fox and Pesetsky's idea that PF does not tolerate any contradictory linearization information (i.e., the linear "slot" of an element is determined only once), it follows that an element will have to move out of the complement of a phase if it is to be pronounced in a displaced position. By targeting the phase edge, phase by phase, the element will move successive cyclically.

To implement this idea technically, Bošković adopts from Chomsky's work the idea that complements of phases are the structural units sent to Spell-Out. He also adopts the idea that an unchecked/unvalued feature on an element signals that movement is required. These two assumptions combined yield the fact that if an element with an unchecked/unvalued feature finds itself in the complement domain of a phase, the element will initiate movement

to prevent what one might call "premature linearization" (linearization that would have to be altered once movement has taken place), to avoid being trapped in the complement of a phase, which would give the wrong linearization instruction to PF. Movement will target a position which precedes the element it ends up preceding in the final PF-representation, i.e., at the left edge of each unit sent to PF for linearization – the phase edge. Intermediate steps are thus required for purposes of pronunciation, and are therefore crucially not required of operations that have no effect on the PF-output, such as Agree.

Bošković's approach is much more detailed than the other two proposals reviewed in this section, but common to all three approaches is the idea that an unchecked feature on the moving element is enough to license movement. Such a claim, however, cannot be maintained in a framework that allows for an Agree operation (which the three approaches adopt, most explicitly in the case of Bošković). Under an Agree framework, checking of features is divorced from movement. Although I agree with all the approaches that an unchecked feature on the moving element is necessary for movement to take place, it cannot be considered a sufficient condition.

In addition to this conceptual problem, all three approaches in their own way suffer from the absence of a principled theory of landing site. In Lasnik et al.'s approach, it is not clear how to capture the A/A-bar distinction that determines the landing sites of the various movements, matching the identity of the unchecked feature on the moving element (a problem this approach shares with Takahashi's 1994 analysis). For Franks and Levine, the problem is that they assume a phase-based system, which makes landing sites arbitrary, as I have argued at length in chapter 3.

Although Bošković's approach crucially relies on the existence of the Phase Impenetrability Condition, Bošković insists that his PIC is a PF-condition, not a syntactic condition, and as such differs from Chomsky's. Bošković further notes that his notion of phase need not coincide with Chomsky's, and that his system is equally compatible with the idea that all projections are phases (a claim also made by Epstein and Seely 2006), in a way very similar to the (quasi-)uniform path approach I have argued for in chapter 3. But closer examination reveals that Bošković's position that every projection is a phase cannot be maintained. If every projection were

a phase, the complement of the first phase would invariably be trapped in situ, as movement of that element to the first phase edge would be forbidden on grounds of anti-locality. If it is not the case that every projection is a phase, we are left with no non-arbitrary way of determining which projections are phases.

An additional problem for Bošković is that there is good evidence that covert operations like Quantifier Raising, covert wh-movement, etc. apply in a successive cyclic fashion (see Fox 2000; Nissenbaum 2000). This is completely unpredicted under Bošković's approach, which restricts successive cyclic movement to elements that must be linearized.

Finally, saying that intermediate steps of movement are PF driven by treating the PIC as a PF-condition raises non-trivial questions of look-ahead and touches on the very autonomy of syntax. For these reasons I find Bošković's approach unsatisfactory, despite the fact that it comes closest to the view on successive cyclicity defended in this book.

6.5   Conclusion

In this chapter I have reviewed alternative approaches to successive cyclic movement in the minimalist program. Many of them were domain based (Bošković's, Franks and Lavine's, Grohmann's); others analyzed successive cyclic movement as feature driven (the agreement-based approach of Rackowski and Richards 2005). Yet other approaches claimed that successive cyclic movement is an illusion, the result of iterative clausal mergers underneath the root of the tree (Castillo and Uriagereka's, and Drury's). I have found all these alternatives unsatisfactory for a variety of reasons. Recurring problems appear to be (i) a problematic notion of possible landing site (too few of them, too arbitrarily determined, etc.); (ii) a conflict between feature checking by movement and Agree (checking without movement); (iii) stipulative locality principles based on phases (PIC) or prolific domains; and (iv) failure to address why-questions (why would phi-features drive movement, why does pronunciation influence syntactic operations, etc.). I hope to have avoided these unsatisfactory features in my analysis of successive cyclicity, but I encourage the reader to be as critical of my approach as I have been when it comes to alternatives.

Chapter 7

Successive Cyclicity and Other Aspects of Locality

The present book has developed an analysis of successive cyclicity that guarantees Takahashi's (1994) desired result that chain links be kept as short as possible. It does so by requiring that movement, once initiated, target each possible landing site along the movement path. Since intermediate steps of movement are not taken to be feature driven, I have not found it necessary to characterize the notion of "possible intermediate landing site" in featural terms (say, A- vs. A-bar, or finer-grained distinctions within these). This is an advantage of the present proposal, as I don't know of any non-arbitrary way of achieving the right featural characterizations.

I have also found it unnecessary to appeal to some notion of domain or bounding node (equivalently, barrier, phase, or "prolific domain"), beyond the notion of phrase (necessary to define antilocality properly). This should again be seen as a welcome consequence of the present analysis, given that no non-arbitrary definition of domain exists in the literature, despite the extensive use of the notion of domain since at least Chomsky (1973) in the context of locality.

The only notion of locality I have adopted, i.e., strict subjacency, arguably falls within what Chomsky would call "virtual conceptual necessity." It is nothing more than adjacency defined "vertically" on a syntactic tree, as illustrated in the figure on the next page.
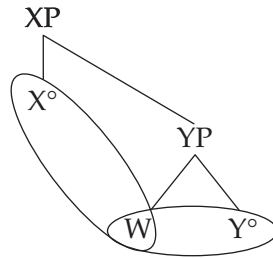
At this point the reader may well wonder whether this impoverished view of locality will be enough once we move beyond the domain of successive cyclicity and address issues pertaining to islandhood. This is clearly a topic that I cannot address in this chapter. It requires an independent study of the type undertaken in Boeckx (2006b). What I would like to do in this chapter is highlight a few recent findings in the domain of locality that ought to be taken into account once we move beyond successive cyclicity and tackle other aspects of locality. To put it differently, what follows are modest remarks that may well serve as starting points in the study of locality at large.

7.1   The Standard View on Islands

Much theoretical activity leading to the minimalist program concentrated on the characterization of the Empty Category Principle (ECP), which regulated the occurrences of gaps (i.e., traces) (and other empty categories). It is not my intention to review this vast and fascinating literature in this chapter. This task has already been carried out, successfully, by Manzini (1992), Hornstein and Weinberg (1995), and Szabolcsi (2006). Suffice it to say that locality conditions are generally assumed to fall into two broad classes: weak or selective islands (WI) and strong or absolute (SI) islands. WIs (wh-island, factive island, negative island, etc.) are domains which prohibit extraction of certain types of elements (say, adjuncts), but not others (say, arguments). Witness the contrast in (1).

(1) a. ?Which of the two books do you wonder whether Mary read?
     b. *For which reason do you wonder whether Mary read *Moby-Dick*?

To a very large extent, the selectiveness of WIs is characterizable in terms of Rizzi's (1990) relativized minimality. In a situation like (2) β acts as an intervener (i.e., creates an island), blocking any relation between α and γ, *unless* β and (α,γ) are of distinct types.

(2) α . . . β . . . γ

Obviously, the task of the theorist is to find the adequate types of elements functioning in syntax and entering into (2) (for relevant discussion, see Rizzi 2004b; Starke 2001; Boeckx and Jeong 2004; Bejar 2003). At present, it seems quite clear that this will necessitate an organization of features in terms of class, sub-class (dimensions), and possibly values, a feature hierarchy/geometry not unlike that developed in phonological theory.

Details aside, the type of solution offered by relativized minimality to the phenomenon of WIs is very elegant, and minimalist in character. If tenable, analyses relying on relativized minimality reach the conclusion that search terminates as soon as the very first element of the relevant type on the search path is found. Syntax does not have the ability to look beyond this first element even if a more appropriate or more "needy" element of the relevant type could be found if the search were allowed to continue.

SIs have so far received a less satisfactory treatment in the minimalist literature. Most approaches adopt and seek to refine Huang's (1982) fundamental insight that any extraction out of non-governed domains is barred (CED), where non-governed domains are adjuncts and subjects (/specifiers occupied by moved elements); that is, roughly, non-complements. Typical examples appear in (3).

(3) a. Who did you see [pictures of *t*]?
     b. *Who did [pictures of *t*] cause Bill to cry?
     c. *Who did John cry [after Bill kissed *t*]?

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. The type of solution that various researchers have provided, however, is uniform: domains out of which extraction is barred emerge derivationally, from the computational dynamics and

resources of narrow syntax (see Uriagereka 1999a, 1999b, 2003; Ochi 1999a, 1999b; Stepanov 2001; Nunes 2004; Nunes and Uriagereka 2000; N. Richards 2001; Chomsky 2000).[1] There are no non-transparent elements *per se*; non-transparency/opacity is a property that emerges once certain elements appear in certain structural configurations (the notorious "domains"). Why opacity emerges at all, and why only in some configurations but not others, are still matters of debate. I don't think we have found a satisfactory solution yet.

7.2   Puzzles for the Standard View

I think that the failure to adequately characterize SIs is due to various factors.

7.2.1   *Movement, freezing, and escape hatch*

First, the standard intuition behind at least part of the CED (the subject part, not the adjunct part) results from the fact that displaced constituents become islands for sub-extraction (see Wexler and Culicover's 1980 Freezing Principle; see also Takahashi 1994 and Ormazabal et al. 1994). This is often related to the fact that moved elements freeze in place (cf. chapter 5). This intuition often conflicts with another intuition, present since Chomsky (1973): in order to extract successfully, an element must move. In *Barriers*, complements had to adjoin to VP to void the barrierhood of the latter (a prerequisite to successful wh-extraction). Much more recently, under the phase system, elements must move through the edge of phases.

The "move away" from complements is to be found in other approaches. For example, in Larson's (1988) VP-shell treatment of ditransitives, innermost complements are adjuncts, and traditional complements are generated in specifier positions. A similar state of affairs obtains in the various approaches stemming from Kayne's (1994) anti-symmetry framework, where, in practice, complements vacate their base-Merge positions. In fact, this practice was elevated to a requirement in Chomsky (1995:338) that "every right-branching

structure must end in a trace," and in Kayne (2000) that every element must move at least once to be visible to linearization.

This trend has the side effect that it becomes very difficult to formulate the generalization, seemingly at the heart of the CED, that extraction procceds most easily from unmoved constituents. Of course, one can always try to distinguish among various types of movement, some being opaque to extraction, others not. But I know of no non-arbitrary way of doing that.

7.2.2 Island by default?

Second, as Postal points out (1997), "while there is far from agreement on what principles separate islands from non-islands, the many partially diverse approaches share a key property . . . : constituents are in effect taken to be *non*-islands *by default*. The problem for theory construction then reduces to a search for principles assigning *some* constituents to the island category." Postal also points out that taking elements to be non-islands by default may just be the wrong perspective if one aims at a natural characterization of islandhood. Perhaps it may be more insightful to follow Cinque's (1978) proposal that constituents are taken to be *islands by default*, and that a special clause needs to be added to let extraction take place in a minimal set of cases.

7.2.3 Island-obviation

Third, and in my view most important, is the fact that perhaps the premise that there are absolute islands is false (the following remarks reiterate and amplify some of the points I made concerning the nature of islands in Boeckx 2003a, 2006b, to which the reader is referred for more technical discussion). As we will see momentarily, there appears to be evidence that even strong islands are not completely opaque in all circumstances. In others words, perhaps all islands are selective. Therefore, it is perhaps wrong to try to characterize SIs in terms of computational dynamics, or "build 'islandhood' in the definition of Move" (Aoun and Li 2003). If even SIs allow for some instances of extraction, the system must have enough computational resources to allow at least these.

This should come as no surprise to those familiar with the minimalist claim that movement is not a primitive operation (see Hornstein 2001; Nunes 2004; Chomsky 1995). If movement is decomposable into various sub-operations (say, Agree, Copy, and remerge), we may find enough conceptual room to allow for some kind of movement across any kind of island. That is to say, maybe the notion of island is to be relativized to the sub-operations that, once combined, yield "movement." This in turn would mean that islands (specifically, SIs) should no longer be taken to be diagnostics for movement (contrary to Chomsky 1977, who took the presence of SIs to indicate absence of movement).

A similar point is expressed very clearly in the following quotation from Adger and Ramchand (2005:162):

> Given recent approaches to syntax [e.g., Chomsky 2000, 2001], locality effects can no longer be assumed to be a diagnostic of movement. This is because, in theories like [Chomsky's], the abstract operation Agree, which applies between features of heads in a structure, must itself be constrained by some theory of locality. The syntactic operation of movement is parasitic on Agree, so it is not possible to use locality as a diagnostic for whether movement has taken place.

To put it another way, the presence of a canonical SI need not be equated with the impossibility of movement, just as we saw in chapter 3 that lack of reconstruction does not entail absence of movement.

Let me expand on these last remarks, and illustrate them by means of a few salient examples.

7.2.3.1   *Sluicing*

Perhaps the clearest and easiest-to-demonstrate piece of evidence for the selectivity of all islands, including SIs, comes from sluicing. Sluicing refers to examples like those in (4), which is ellipsis of the sentential complement to an interrogative complementizer hosting a wh-phrase:

(4) a.   Jack bought something, but I don't know what.
     b.   A: Someone called. B: Really? Who?
     c.   Beth was there, but you'll never guess who else.

d. Jack called, but I don't know {when/how/why/where from}.
e. Sally's out hunting – guess what!
f. A car is parked on the lawn – find out whose.

Ross's (1969) take on sluicing was to assume that the wh-phrase has been moved from its usual position to the beginning of the clause. That movement operation is then followed by phonetic deletion of the rest of the clause (including the position from which wh-movement originated).

(5)
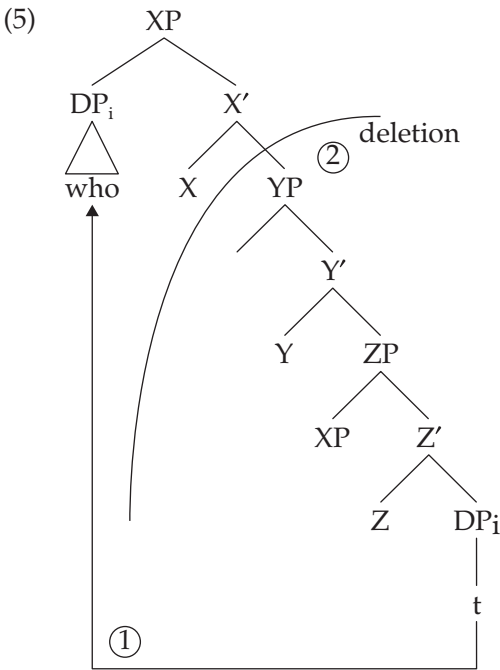
Recently, Lasnik (2001a, 2005) and especially Merchant (2001) have offered a reappraisal of Ross's analysis in a minimalist setting.

I will assume that the movement + deletion analysis of sluicing is fundamentally correct. (Below, we'll see some good evidence that it must be correct.) I will here focus on a property of sluicing that has attracted a lot of attention in recent years.

As originally noted by Ross (1969) (see also Chomsky 1972; Lasnik 2001a, 2005; Merchant 2001), sluicing appears to "rescue" island violations, as in the following examples. (Strikethrough indicates elision.)

Strong island (subject condition):

(6) a. *Which Marx brother did she say that [a biography of
         t] is going to be published?
     b. A biography of one of the Marx brothers is going to be
         published next week, but I don't know which ~~she said
         that a biography of *t* is going to be published~~.

Weak island (wh-island):

(7) a. ?*Which book did every journalist go out today to find
         out who was selling *t*?
     b. Every journalist went out today to find out who was sell-
         ing a certain book, but I don't know which (book) ~~every
         journalist went out today to find out who was selling *t*~~.

If sluicing is analyzed as movement + deletion, examples like (6b)
and (7b) strongly suggest that islands are not the result of com-
putational properties of narrow syntax that block the application
of rules like wh-movement, for wh-movement does take place
across an "island," even a strong island in (6).

Interestingly, although sluicing appears to be able to repair all
kinds of island violations (see Lasnik 2005 for arguments in favor
of this strongest conclusion, *contra* Merchant 2001), it is not cap-
able of repairing all kinds of locality violations. For example, Boeckx
and Lasnik (2006) argue on the basis of examples from Serbo-
Croatian that sluicing cannot repair superiority effects. The gist of
their argument is as follows. (Boeckx and Lasnik independently show
that Serbo-Croatian is like English in displaying island repair
under sluicing.) Serbo-Croatian exhibits superiority effects (in
most contexts, including indirect questions; see Bošković 2002 for
discussion).[2]

(8)   Ivan i   Marko ne znaju . . .
         Ivan and Marko neg know
     a. ko   je šta   kupio
         who is what bought
         'Who is buying what?'
     b. *šta je ko kupio
         'Ivan and Marko don't know who bought what'

Interestingly, as Stjepanović (1999, 2003) discusses, in sluicing contexts superiority effects are exhibited. (9) shows this for embedded questions (M. Petrović, M. Simić, p.c.).

(9) A: Somebody bought something, but
     B: a. Ivan i Marko ne znaju ko šta
             Ivan and Marko neg know who what
         b. *Ivan i Marko ne znaju šta ko
             'but Ivan and Marko don't know who what.'

In a similar vein, Boeckx (2006a) has argued that sluicing cannot repair anti-locality violations. The crucial piece of evidence here comes from preposition-stranding data. 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-sentence in examples 10), we also find the possibility 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-sentence in (10). (Examples taken from Merchant 2001.)

(10) a. Peter was talking with someone, but I don't know
          (with) who.
      b. Who was he talking with?

In other Germanic languages, which generally do not allow preposition stranding under wh-movement, retention of the preposition under sluicing is obligatory. (11) is taken from German:

(11) a. Anna hat mit jemandem gesprochen, aber ich weiß
          Anna has with someone   spoken   but I   know
          nicht, *(mit) wem.
          not   with who
          'Anna spoke with someone, but I don't know who.'
      b. *Wem hat sie mit gesprochen?

As Merchant (2001) has extensively documented, this correlation holds true of a large number of languages: a language will exhibit preposition stranding under sluicing only if it exhibits preposition stranding in non-ellipsis contexts.

This correlation finds its most natural explanation in the theory of sluicing that takes it to consist of the usual operation of wh-movement, subject to the usual language-particular constraints, followed by deletion of the portion of the clause out of which extraction has taken place. Indeed, the preposition-stranding generalization uncovered by Merchant constitutes one of the strongest pieces of evidence for a movement + ellipsis account of sluicing.

Now, we can say that whatever prevents preposition stranding in the languages just mentioned cannot be repaired by sluicing. A natural way of making sense of this is to take the ban on preposition stranding as constraining syntactic derivations: derivations that would violate the ban on preposition stranding cannot even take place, hence cannot be repaired by ellipsis. (A similar reasoning would hold for superiority.) Abels (2003) has proposed that the ban on preposition stranding is a reflex of anti-locality. For present purposes, I will simplify Abels's account, and simply posit that in languages banning preposition stranding, Prepositional Phrases (PPs) have a special requirement that forces any extraction out of them to proceed through SpecPP. This ban conflicts with the requirement that movement must at least cross a full phrasal category, as it would require the complement of P to move through SpecPP. The conflict is solved if no preposition stranding takes place, and the preposition is pied-piped under wh-movement. By contrast, in languages like English, which allow preposition stranding, prepositions are complex expressions that consist of two projections: PP, and pP, say, as schematized in (12).

(12) $[_{pP}\ p°\ [_{PP}\ P°]]$

Suppose that in those languages the requirement that forces any extraction out of "prepositional phrases" to proceed through a specifier position actually holds of pP, not PP. This time, the movement-through-spec requirement won't conflict with anti-locality, since the targeted specifier position does not belong to the same projection from which movement originates. Hence, preposition stranding is possible.

If the discussion is on the right track, the sluicing data reviewed here offer rather compelling evidence for the claim that conditions like anti-locality and superiority constrain computations in narrow syntax. The sluicing data show that consequences of economy

conditions like "form the shortest chain possible" (anti-locality and superiority) are so deeply embedded within the computational system of human language that operations like ellipsis cannot alleviate their effects on linguistic representations. At the same time, the sluicing facts offer a challenge for attempts to unify all locality conditions. They seem to suggest that the grammar needs to encode both derivational and representational conditions, with only representational conditions repairable by ellipsis.[3]

7.2.3.2   *Resumption*

An argument virtually identical to the one made on the basis of sluicing can be made on the basis of resumption, which, as is well known, "rescues" island effects (see Boeckx and Lasnik 2006). The data I will use to illustrate my point come from Lebanese Arabic. Aoun and Li (2003) show that Lebanese Arabic, like English, forms questions by fronting interrogative words.

(13)   Miin ʃəft?
        Who saw.2sg
        'Who did you see?'

As in English, this fronting process is sensitive to islands.

(14)   *Miin btaʕrfo   l-mara   yalli ʃeefit
        Who know.2pl the-woman that saw.3sgfem
        bə-l-maTʕam?
        in-the-restaurant
        'Who do you know the woman that saw in the restaurant?'

And like English, the language can obviate island effects by using a resumptive pronoun.

(15)   Miin btaʕrfo   l-mara   yalli ʃeefit-o
        who know.2pl the-woman that saw.3sgfem-him
        bə-l-maTʕam?
        in-the-restaurant
        'Who do you know the woman that saw him in the restaurant?'

The traditional view is that the resumptive pronoun in such cases indicates that fronting has not taken place. The resumptive pronoun basically acts as a signal that if there had been no island the interrogative word would have been displaced from there. The fascinating finding in Aoun and Li's (2003) discussion is that even in those island contexts, superiority effects – standardly taken to indicate shortest *move* – still obtain, as the following examples show.

(16)   a.   Miin ʔənbasatto laʔinno saami ʕarraf-o
            who pleased.2pl because Sami introduced-him
            ʕa-miin?
            to-whom
            'Who were you pleased because Sami introduced (him) to whom?'
       b.   *Miin ʔenbasatto laʔinno saami ʕarraf   miin
            who pleased.2pl because Sami introduced whom
            ʕəl-e?
            to-him
            'Who were you pleased because Sami introduced who to him?'

Boeckx (2003a) argues at length that the critical cases of resumptive structures seemingly "violating" islands involve movement across such "islands," as in the sluicing cases. Hence the superiority effect in (16). Boeckx proposes a derivation for resumptive structure according to which the resumptive pronoun and its antecedent start off as a big DP constituent, which is split in the course of the derivation (in a way very reminiscent of Sportiche's 1988 influential Quantifier-Float stranding analysis). This is schematized in (17).

(17)   [Wh [C^0 [. . . [. . . [$_{DP}$ RP [<Wh>]] . . .]]]]

If this derivation of resumptive structures is correct, islands cannot constrain movement. To put this differently, the sluicing facts alluded to above and the resumption facts in Boeckx (2003a) indicate that no domain is an *absolute* island. All islands appear to be selective.

7.3 Ross's View

This is not far removed from the position entertained by Ross (1967) (although it is rarely represented in all its subtlety). Contrary to standard characterization, Ross did not treat islands as impenetrable domains. For Ross, movement was unbounded. Crossing an island in and of itself did not suffice to yield a deviant output. Rather, only certain *types* of rules were sensitive to islands. Ross identified two such types: chopping rules and feature-changing rules. For feature-changing/chopping rules, islands constitute impenetrable domains. By contrast, copying rules, distinct from chopping/feature-changing rules, were said to be insensitive to islandhood. Although notions like chopping/feature-changing vs. copying rules are no longer available in current frameworks, Ross's intuition that islands are sensitive to types of processes is, I think, on the right track.

7.4 Agreement and Islandhood

I have argued in Boeckx (2003a) that (strong) agreement (i.e., agreement in all phi-features) has a major effect on island formation. In the course of arguing for a derivation like (17) for resumption, I discovered that in order for this "big DP" splitting to take place, the resumptive pronoun and the wh-phrase cannot be in an agreement relation (if they match in features, this matching must be seen as accidental). This correctly predicts the pervasive instances of anti-identity one finds in language after language in the context of resumption.[4] Witness the following cases ((18)–(19) are taken from Irish, (20)–(21) from Scots Gaelic).
 Anti-Person Agreement:

(18) a   Alec, tusa a   bhfuil an Béarla   aige
        hey Alec you aN is   the English at-him
        'hey Alec you that know(s) English'

Anti-number agreement:

(19)   na daoine a chuirfeadh   isteach ar an phost sin
        the men   C put-cond-3sg in   for the job   that
        'the men that would apply for that job'

Anti-gender agreement:

(20)   Dè   a'mhàileid   a chuir thu am peann ann?
       which the.bag-Fem C put   you the pen   in-3-Masc
       'Which bag did you put the pen in?'

Anti-case agreement:

(21)   a.   Bha   thu a'geàrradh na craoibhe.
            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?'

On the basis of such facts Boeckx (2003a) hypothesizes that lack of agreement is a precondition on successful extraction (lack of opacity/islandhood).

There is independent evidence that movement is closely related to what is known in the literature as "anti-agreement" effects. Consider the following examples from Northern Italian dialects.

(22)   a.   La Maria l'   è venuta.
            the Maria she is come
            'Maria came.'
       b.   Gli è venuto la   Maria.
            it is come   the Maria
            'Maria came.'
       c.   Quante   ragazze gli è venuto con te?
            how.many girls   it is come   with you
            'How many girls came with you?'
       d.   *Quante   ragazze le   sono venute con te?
            how.many girls   they are   come   with you
       e.   Chi hai detto che __ e partito?
            who has said   that is   left
            'Who did he say that left?'

(22a) shows that preverbal subjects relate to a ϕ-feature matching clitic (3rd fem. sg). By contrast, postverbal subjects don't (22b); the clitic bears default morphology. (22c) and (22d) show that subject

extraction requires the use of a non-agreeing clitic. On the basis of (22c), we can conclude (as did Rizzi 1982) that in standard Italian examples like (22e), subject extraction takes place from a postverbal position related to a silent non-agreeing clitic (*pro*), which obviates the [*that*-trace] effect.

The anti-agreement effects under successive cyclic movement in Selayarese discussed in the previous two chapters point in the same direction. In a similar vein, the fact that objects are islands for standard wh-extraction in languages with object agreement, like Basque (23), suggests that agreement tends to turn a transparent element into an island.

(23) *Nori buruzko sortu zitusten aurreko asteko istiluek
      Who about-of create scandals last   week   scandals
      zurrumurruak?
      rumors
      'Who have last week's scandals caused [rumors about]?'

Why should agreement have this effect? Boeckx (2003a) argues that chains can contain at most "strong" positions, where "strong position" can be equated with [+wh]-checking position or with "strong agreement." (See Boeckx 2003a for a more precise characterization; for a similar intuition, see N. Richards 1997, 2001; Rizzi 2006; Rizzi and Shlonsky 2005.) As a result of this ban on "chains that are too strong," elements that normally agree ("A"-type agreement, or what Chomsky 2001 calls "complete φ-feature agreement") must disagree (i.e., "anti-agree") in order for them to successfully enter into a checking relation with an A-bar target (wh-/A-bar-feature checking). The intuition behind Boeckx's analysis is Last Resort. Once a chain contains a strong position (of any kind), it cannot contain another, equally strong position (of any kind).

Needless to say, this is far from a comprehensive analysis of islandhood. I have shown in Boeckx (2003a) that this view helps us understand resumptive structures, and the lack of island effects in such contexts. (Boeckx claims that the role of a resumptive pronoun is to take care of "A-style" agreement, freeing the wh-phrase from establishing a successful relationship with an A-bar target.) I believe the approach captures the subject-part of the CED, and it may extend to some of the repair-under-ellipsis cases, but adjunct islands remain more problematic.[5]

7.5   Conclusion

To sum up this chapter, I want to stress once again that I have not tried to develop a comprehensive theory of islandhood here. But it seems to me that a few conclusions can already be reached at this point. The above remarks certainly strengthen the case against domains I have made in previous chapters on the basis of successive cyclicity. It is also clear that phenomena like island, movement, etc. will need to be decomposed into more basic operations and relations, in a way similar to what I have done with successive cyclicity.

   These are among the few issues that seem to me to be a good place to start. I would hope that all the points I have raised here will play a role in the development of a minimalist, explanatory theory of islands.

Notes

1   An important exception to this trend is Takahashi (1994), where SIs are characterized in terms of a uniformity conditions of chains, i.e., representationally. (See also Ormazabal et al. 1994.) Path-based approaches to locality in the GB era, such as Kayne (1984), also constitute exceptions to the dominant derivation-based view based on Chomsky's work. For a reappraisal of representational characterizations of islandhood, see Boeckx (2006b).

2   I ask the reader to disregard the exact position of *je*, which appears to break up the wh-cluster by being second-position clitic. For a treatment of *je* consistent with a treatment of (8) where both wh-phrases are in CP-specifiers, see Bošković (2001).

3   In other words, the data just reviewed argue in favor of a "mixed" theory.

   Recently, the status of mixed theories has been questioned at both ends of the representational–derivational spectrum, with Samuel D. Epstein arguing in favor of a "strictly" derivational system (see Epstein 1999; Epstein et al. 1998; Epstein and Seely 2002, 2006), and Michael Brody arguing in favor of a "strictly" representational one (see Brody 1995, 2003). The shape of the argument from either side is as follows: derivations and representations duplicate each other; a parsimonious theory of syntax of the kind favored by the minimalist program should dispense either with representations or with derivations.

It is indeed very hard to disagree with Brody's (2002:20) clear state-
ment that "Having both [representations and derivations duplicating
each other] would weaken the theory in the sense of increasing the
analytic options available . . . , hence very strong arguments would be
needed to maintain that both concept-sets are part of the competence
theory of syntax." But Brody's argument (or Epstein's for that matter)
is valid only if indeed derivational and representational properties
duplicate one another. For evidence that they don't, see Boeckx (2003c,
2006b).

4   It also accounts for Merchant's (2001:146) generalization that "No
resumptive-binding operator can be case-marked" once case-marking
is taken to be a reflex of ("A-") agreement (see Chomsky 2000, 2001).

5   Boeckx (2003a) argues that adjuncts are islands because they cannot enter
into any Agree relation, like displaced arguments. While descriptively
correct, this generalization is in need of an explanation.

Chapter 8

Concluding Remarks

The present book has focused on an empirically well-established fact about human languages: long-distance syntactic dependencies exhibit signs that strongly suggest that they are the result of a conjunction of small dependencies embedded in one another – a phenomenon known as successive cyclicity. After reviewing the evidence that has accumulated over the past thirty years or so, I broke down the phenomenon into three major questions. The first question is the more general one, and arises in virtually all syntactic frameworks. It concerns the length of the small dependencies that together result in long-distance relations. Another way of phrasing the question is this: what are the landing sites for movement? I argued that the standard answer within generative grammar (that only some positions are available for movement) begs questions and raises problems that have not received any satisfactory solution. I therefore argued that all positions on the path of movement are potentially available as intermediate landing sites, irrespective of the movement type (A- or A-bar movement).

The second question I addressed is narrower, as it arises in frameworks that resort to a derivational idiom. The question concerns the timing of movement. Are intermediate movement steps taken early, before the final landing site is introduced into the derivation, or late (as part of the final movement step)? I provided empirical evidence in favor of the "early" position, which many would regard as conceptually more appealing. I examined empirical arguments favoring the opposite conclusion, and showed that they aren't strong enough to question my conclusion.

The third question is more decidedly minimalist in spirit, as it takes as its background assumption that movement is a Last Resort operation. The question concerns the driving force (if any) of intermediate movement steps. I argued that intermediate movement steps are not driven by feature-checking requirements; nor should they be seen as required for economy reasons, locality reasons, or as side-effects of linearization concerns. Rather, they take place because in so doing the moving element maximizes its derivational options. Put differently, movement happens because it does no harm to the moving element, and it may turn out to be beneficial in the long run. In such an agnostic state, movement is favored by the system, *locally*. Although this answer constitutes a departure from earlier minimalist assumptions, especially concerning Last Resort, I nevertheless argued that my proposal does not jeopardize the Last Resort tenet of minimalism, for there is evidence that not any movement can happen. In particular, I argued that intermediate steps of movement are subject to an anti-locality condition (which reduces to a ban on gratuitous movement), which receives empirical support from a variety of contexts. Such a condition would make no sense if Last Resort weren't assumed.

The last two chapters focused on alternative proposals (which I showed to be inferior to the approach I defended here), and other aspects of locality, where my arguments against a domain-based view of locality and in favor of a decomposition of movement into more primitive operations were strengthened.

The emerging picture is richer, deeper, and more detailed than if successive cyclicity were taken to be axiomatic. This is an immediate result of subjecting the phenomenon to minimalist inquiry. Whatever the final answer turns out to be, there is no doubt that more is learned about properties of the language faculty by daring to ask minimalist questions.

References

Abels, K. 2003. Successive-cyclicity, anti-locality, and adposition stranding. Doctoral dissertation, University of Connecticut.

Adger, D. 2003. *Core syntax*. Oxford: Oxford University Press.

Adger, D. and G. Ramchand. 2005. Merge and move: *wh*-dependencies revisited. *Linguistic Inquiry* 36, 161–93.

Alexiadou, A. and G. Fanselow. 2000. On the correlation between morphology and syntax. Ms., ZAS and Potsdam University.

Anagnostopoulou, E. 2003. *The syntax of ditransitives*. Berlin: Mouton/de Gruyter.

Aoun, J. and Y.-H. A. Li. 2003. *Essays on the representational and derivational nature of grammar: the diversity of wh-constructions*. Cambridge, MA: MIT Press.

Baker, M. 1985. The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16, 373–416.

Baker, M. 1988. *Incorporation*. Chicago, IL: University of Chicago Press.

Baker, M. 1997. Thematic roles and syntactic structure. In *Elements of grammar*, ed. L. Haegeman, 73–137. Dordrecht: Kluwer.

Baker, M. 2001. *The atoms of language: the mind's hidden rules of grammar*. New York: Basic Books.

Baker, M. 2004. Applicatives. Ms., Rutgers University.

Baker, M. 2005. Mapping the terrain of language learning. *Language Learning and Language Development* 1, 93–124.

Baker, M., K. Johnson, and I. Roberts. 1989. Passive arguments raised. *Linguistic Inquiry* 20, 219–51.

Baković, E. 1998. Optimality and inversion in Spanish. In *Is the best good enough? Optimality and competition in syntax*, eds. P. Barbosa, D. Fox, M. McGinnis, P. Hagstrom, and D. Pesetsky. Cambridge, MA: MIT Press and MITWPL.

Barss, A. 1986. Chains and anaphoric dependence. Doctoral dissertation, MIT.

Bejar, S. 2003. Phi-syntax. Doctoral dissertation, University of Toronto.

Bhatt, R. 2002. The raising analysis of relative clauses: evidence from adjectival modification. *Natural Language Semantics* 10, 43–90.

Bhatt, R. and Y. Sharvit. 2006. A note on intensional superlatives. In *Proceedings of SALT XV*. Ithaca, NY: Linguistics Department, Cornell University.

Bissell-Doggett, T. 2004. All things being unequal: locality in movement. Doctoral dissertation, MIT.

Bobaljik, J. D. 1998. Floated quantifiers: handle with care. *Glot International* 3.6, 3–10.

Bobaljik, J. D. 2000. The rich agreement hypothesis in review. Ms., McGill University.

Bobaljik, J. D. 2002. A-chains at the PF-interface: copies and "covert" movement. *Natural Language and Linguistic Theory* 20, 197–267.

Bobaljik, J. D. and S. Wurmbrand. 2005. The domain of agreement. *Natural Language and Linguistic Theory* 23, 809–65.

Boeckx, C. 1999. Conflicting C-command requirements. *Studia Linguistica* 53, 227–50.

Boeckx, C. 2000a. Quirky agreement. *Studia Linguistica* 54, 354–80.

Boeckx, C. 2000b. EPP eliminated. Ms., University of Connecticut.

Boeckx, C. 2000c. A note on contraction. *Linguistic Inquiry* 31, 357–66.

Boeckx, C. 2001a. Mechanisms of chain formation. Doctoral dissertation, University of Connecticut.

Boeckx, C. 2001b. Scope reconstruction and A-movement. *Natural Language and Linguistic Theory* 19, 503–48.

Boeckx, C. 2003a. *Islands and chains*. Amsterdam: John Benjamins.

Boeckx, C. 2003b. Intricacies of agreement in Icelandic. Ms., University of Maryland and Harvard University.

Boeckx, C. 2003c. Symmetries and asymmetries in multiple checking. In *Multiple wh-fronting*, eds. C. Boeckx and K. K. Grohmann, 17–26. Amsterdam: John Benjamins.

Boeckx, C. 2004. Long-distance agreement in Hindi: some theoretical implications. *Studia Linguistica* 58, 23–36.

Boeckx, C. 2006a. *Linguistic minimalism: origins, methods, concepts, and aims*. Oxford: Oxford University Press.

Boeckx, C. 2006b. Bare syntax. Ms., Harvard University.

Boeckx, C. 2006c. Honorification as agreement. *Natural Language and Linguistic Theory* 24, 385–98.

Boeckx, C. 2007. Case and agreement. Ms., Harvard University.

Boeckx, C. and K. K. Grohmann. In press. Putting phases into perspective. *Syntax*.

Boeckx, C. and N. Hornstein. To appear. Superiority, reconstruction, and islands. In *Fundamental issues in linguistic theory*, eds. R. Freidin, C. Otero, and M.-L. Zubizarreta. Cambridge, MA: MIT Press.

Boeckx, C. and Y. Jeong. 2004. The fine structure of intervention in syntax. In *Issues in current linguistic theory: a festschrift for Hong Bae Lee*, eds. Chungja Kwon and Wonbin Lee, 83–116. Seoul: Kyungchin.

Boeckx, C. and H. Lasnik. 2006. Intervention and repair. *Linguistic Inquiry* 37, 150–5.

Boeckx, C. and F. Niinuma. 2004. Agreement constraints in Japanese. *Natural Language and Linguistic Theory* 22, 453–80.

Boeckx, C. and S. Stjepanović. 2001. Head-ing toward PF. *Linguistic Inquiry* 32, 345–55.

Borer, H. 2005. *Structuring sense*. Oxford: Oxford University Press.

Bošković, Ž. 1994. D-structure, θ-criterion, and movement into θ-positions. *Linguistic Analysis* 24, 247–86.

Bošković, Ž. 1997a. *The syntax of non-finite complementation: an economy approach*. Cambridge, MA: MIT Press.

Bošković, Ž. 1997b. On certain violations of the superiority condition, AgrO, and economy of derivation. *Journal of Linguistics* 33, 227–54.

Bošković, Ž. 1998. Multiple wh-fronting and economy of derivation. In *Proceedings of WCCFL* 16, 49–63. Stanford University, Stanford, CA.

Bošković, Ž. 1999. On multiple feature checking: multiple wh-fronting and multiple head-movement. In *Working minimalism*, eds. S. D. Epstein and N. Hornstein, 159–87. Cambridge, MA: MIT Press.

Bošković, Ž. 2001. *On the nature of the syntax–phonology interface: cliticization and related phenomena*. London: Elsevier.

Bošković, Ž. 2002. A-movement and the EPP. *Syntax* 5, 167–218.

Bošković, Ž. 2004. Be careful where you float your quantifiers. *Natural Language and Linguistic Theory* 22, 681–742.

Bošković, Ž. 2005. On the operator freezing effect. Presented at the 6th European Conference on Formal Description of Slavic Languages, University of Potsdam, December 2005.

Bošković, Ž. 2006. On the locality of move and agree. Ms., University of Connecticut.

Bošković, Ž. and H. Lasnik. 1999. How strict is the cycle? *Linguistic Inquiry* 30, 689–97.

Bošković, Ž. and H. Lasnik 2006. *Minimalist syntax: the essential readings*. Oxford: Blackwell.

Bouchard, D. 2002. *Adjectives, number, and interfaces: why languages vary*. Oxford: Elsevier.

Bresnan, J. and S. Mchombo 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63, 741–82.

Bresnan, J. and L. Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21, 147–85.

Brody, M. 1995. *Lexico-logical form*. Cambridge, MA: MIT Press.

Brody, M. 1997. Perfect chains. In *Elements of grammar: handbook of generative syntax*, ed. L. Haegeman, 139–67. Dordrecht: Kluwer.

Brody, M. 2002. On the status of derivations and representations. In *Derivation and explanation in the minimalist program*, eds. S. D. Epstein and T. D. Seely, 19–41. Oxford: Blackwell.

Brody, M. 2003. *Towards an elegant syntax*. London: Routledge.

Broekhuis, H. 2006. Extraction from subjects: some remarks on Chomsky's "On phases." In *Organizing grammar*, eds. H. Broekhuis, N. Corver, and R. Huybregts, 59–68. Berlin: Mouton/de Gruyter.

Bruening, B. 2001. Syntax at the edge: cross-clausal phenomena and the syntax of Passamaquoddy. Doctoral dissertation, MIT.

Bruening, B. 2006. Differences between the wh-scope-marking and wh-copy constructions in Passamaquoddy. *Linguistic Inquiry* 37, 25–49.

Carnie, A. 2002. *Syntax*: *a generative introduction*. Oxford: Blackwell.

Carstens, V. 2001. Multiple agreement and case-deletion: against phi-completeness. *Syntax* 4, 147–63.

Castillo, J. C. and J. Uriagereka. 2002. A note on successive cyclicity. In *Derivations*, J. Uriagereka, 136–46. London: Routledge.

Castillo, J. C., J. Drury, and K. K. Grohmann. 1998. Merge-over-move and the EPP. In *University of Maryland Working Papers in Linguistics* 8, 63–103. University of Maryland, College Park. [Published in part in K. K. Grohmann, J. C. Castillo, and J. Drury. 1999. No more EPP. In *Proceedings of WCCFL* 19, 153–66. Somerville: Cascadilla Press.]

Ceplova, M. 2001. Minimalist islands. Ms., MIT.

Cheng, L. 2006. Decomposing Bantu relatives. Ms., Leiden University.

Chomsky, N. 1955. The logical structure of linguistic theory. Ms., Harvard/MIT. [Published in part, 1975, New York: Plenum.]

Chomsky, N. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, N. 1964. *Current issues in linguistic theory*. The Hague: Mouton.

Chomsky, N. 1970. Remarks on nominalization. In *Readings in English Transformational Gramma*r, eds. R. Jacobs and P. Rosenbaum, 184–221. Waltham, MA: Blaisdell.

Chomsky, N. 1972. Some empirical issues in the theory of transformational grammar. In *The goals of linguistic theory*, ed. S. Peters, 63–130. Prentice-Hall: Englewood Cliffs, NJ.

Chomsky, N. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, eds. S. Anderson and P. Kiparsky, 232–86. New York: Holt, Rinehart, and Winston.

Chomsky, N. 1977. On wh-movement. In *Formal syntax*, eds. P. Culicover, T. Wasow, and A. Akmajian, 71–132. New York: Academic Press.

Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. 1986a. *Knowledge of language*. New York: Praeger.

Chomsky, N. 1986b. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, N. 1993. A minimalist program for linguistic theory. In *The view from Building 20*, eds. K. Hale and S. J. Keyser, 1–52. Cambridge, MA: MIT Press.

Chomsky, N. 1994. Bare phrase structure. MIT Occasional Papers in Linguistics 5. Cambridge, MIT, MIT Working Papers in Linguistics. [Published in G. Webelhuth, ed. 1995. *Government and binding and the minimalist program*, 385–439. Oxford: Blackwell.]

Chomsky, N. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. 2000. Minimalist inquiries: the framework. In *Step by Step*, eds. R. Martin, D. Michaels, and J. Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, N. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. M. Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Chomsky, N. 2002. *On nature and language*. Cambridge: Cambridge University Press.

Chomsky, N. 2004a. Beyond explanatory adequacy. In *Structures and beyond*, ed. A. Belletti, 104–31. Oxford: Oxford University Press.

Chomsky, N. 2004b. *The generative enterprise reconsidered*. Berlin: Mouton/ de Gruyter.

Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1–22.

Chomsky, N. 2006. Approaching UG from below. Ms., MIT.

Chomsky, N. To appear. On phases. In *Foundational issues in linguistic theory*, eds. R. Freidin, C. Otero, and M.-L. Zubizarreta. Cambridge, MA: MIT Press.

Chomsky, N. and H. Lasnik. 1993. Principles and parameters theory. In *Syntax: an international handbook of contemporary research*, eds. J. Jacobs, A. von Stechow, W. Sternefeld, and T. Vennemann, 506–69. Berlin: de Gruyter.

Chung, S. 1982. Unbounded dependencies in Chamorro grammar. *Linguistic Inquiry* 13, 39–77.

Chung, S. 1994. Wh-agreement and "referentiality" in Chamorro. *Linguistic Inquiry* 25, 1–44.

Chung, S. 1998. *The design of agreement*. Chicago: University of Chicago Press.

Chung, S. and C. Georgopoulos. 1988. Agreement with gaps in Chamorro and Palauan. In *Agreement in natural language: approaches, theories, and description*, eds. M. Barlow and C. A. Ferguson, 251–67. Stanford, CA: CSLI.

Cinque, G. 1978. Towards a unified treatment of island constraints. In *Proceedings of the 12th International Congress of Linguistics*, Inssbrucker Beitrage zur Sprachwissenschaft, eds. W. Dressler and W. Meid, 344–8.

Cinque, G. 1999. *Adverbs and functional heads*. Oxford: Oxford University Press.

Cinque, G., ed. 2002. *Functional structure in DP and IP*. Oxford: Oxford University Press.

Cinque, G. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36, 315–32.

Cinque, G. and R. S. Kayne, eds. 2005. *The Oxford handbook of comparative syntax*. Oxford: Oxford University Press.

Clements, G. N. 1984. Binding domains in Kikuyu. *Studies in the Linguistic Sciences* 14, 37–56.

Clements, G., J. McCloskey, J. Maling, and A. Zaenen. 1983. String-vacuous rule application. *Linguistic Inquiry* 14, 1–17.

Cole, P. and G. Hermon. 1998. The typology of wh-words: wh-questions in Malay. *Syntax* 1, 221–58.

Collins, C. 1993. Topics in Ewe syntax. Doctoral dissertation, MIT.

Collins, C. 1994. Economy of derivation and the generalized proper binding condition. *Linguistic Inquiry* 25, 45–61.

Collins, C. 1997. *Local economy*. Cambridge, MA: MIT Press.

Culicover, P. and R. Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.

Davies, W. 2003. Extreme locality in Madurese wh-questions. *Syntax* 6, 237–59.

Den Dikken, Marcel. 2006. Vacuous movement in focus. Ms., CUNY.

Drury, J. 2005. Alternative directions for minimalist inquiry: expanding and contracting phases of derivation. Doctoral dissertation, University of Maryland.

Epstein, S. D. 1999. Un-principled syntax: the derivation of syntactic relations. In *Working minimalism*, eds. S. D. Epstein and N. Hornstein, 317–45. Cambridge, MA: MIT Press.

Epstein, S. D. To appear. On inter-modular, I-functional explanation. *Linguistic Analysis*.

Epstein, S. D. and T. D. Seely. 1999. Against the GF-notion "subject": eliminating the EPP and successive cyclic A-movement. Ms., Michigan University and Michigan State University.

Epstein, S. D. and T. D. Seely. 2002. Rule application as cycles in a level-free syntax. In *Explanation and derivation in the minimalist program*, eds. S. D. Epstein and T. D. Seely, 65–89. Oxford: Blackwell.

Epstein, S. D. and T. D. Seely. 2006. *Derivations in minimalism*. Cambridge: Cambridge University Press.

Epstein, S. D., E. Groat, R. Kawashima, and H. Kitahara. 1998. *The derivation of syntactic relations*. Oxford: Oxford University Press.

Fanselow, G. and A. Mahajan. 2000. Towards a minimalist theory of wh-expletives, wh-copying, and successive cyclicity. In *Wh-scope marking*, eds. U. Lutz, G. Müller, and A. von Stechow, 195–230. Amsterdam and Philadelphia: John Benjamins.

Felser, C. 2003. *Wh*-copying, phases, and successive cyclicity. *Lingua* 114, 543–74.

Feynman, R. 1965. *The Feynman lectures in physics*. Reading, MA: Addison-Wesley.

Finer, D. L. 1997. Contrasting A-bar dependencies in Selayarese. *Natural Language and Linguistic Theory* 15, 677–728.

Fitch, T., M. Hauser, and N. Chomsky. 2004. The evolution of the language faculty: clarifications and implications. Available at http://wjh.harvard.edu/~mnkylab/publications. [Abridged version published 2005 in *Cognition* 97, 179–210.]

Fox, D. 2000. *Economy and semantic interpretation*. Cambridge, MA: MIT Press.

Fox, D. and H. Lasnik 2003. Successive-cyclic movement and island repair: the difference between sluicing and VP-ellipsis. *Linguistic Inquiry* 34, 143–54.

Fox, D. and D. Pesetsky. 2004. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31.1–2, 1–46.

Frampton, J. and S. Gutmann. 1999. Cyclic computation, a computationally efficient minimalist syntax. *Syntax* 2, 1–27.

Frampton, J. and S. Gutmann. 2002. Crash-proof syntax. In *Derivation and explanation in the minimalist program*, eds. S. D. Epstein and T. D. Seely, 90–105. Oxford: Blackwell.

Frank, R. 2002. *Phrase structure composition and syntactic dependencies*. Cambridge, MA: MIT Press.

Franks, S. and J. Lavine. To appear. Case and word order in Lithuanian. *Journal of Linguistics*.

Freidin, R. 1999. Cyclicity and minimalism. In *Working minimalism*, eds. S. D. Epstein and N. Hornstein, 95–126. Cambridge, MA: MIT Press.

Fujii, T. 2006. Copy-raising. Ms., University of Maryland.

Fukui, N. 2001. Phrase structure. In *The handbook of contemporary syntactic theory*, eds. M. Baltin and C. Collins, 374–406. Oxford: Blackwell.

Gallego, Á. J. 2006. Phase sliding. Ms., Universitat Autònoma de Barcelona and University of Maryland, College Park.

Georgopoulos, C. 1985. Variables in Palauan syntax. *Natural Language and Linguistic Theory* 3, 59–94.

Georgopoulos, C. 1991. *Syntactic variables: resumptive pronouns and binding in Palauan*. Dordrecht: Kluwer.

Groat, E. 1999. Raising the case of expletives. In *Working minimalism*, eds. S. D. Epstein and N. Hornstein, 27–43. Cambridge, MA: MIT Press.

Groat, E. and J. O'Neil. 1996. Spell-out at the LF interface. In *Minimal ideas*, eds. W. Abraham, S. D. Epstein, H. Thráinsson, and C. J.-W. Zwart, 113–39. Amsterdam: John Benjamins.

Grohmann, K. K. 2000. Prolific peripheries: a radical view from the left. Doctoral dissertation, University of Maryland.

Grohmann, K. K. 2003a. *Prolific domains*. Amsterdam: John Benjamins.

Grohmann, K. K. 2003b. Successive cyclicity under (anti-)local considerations. *Syntax* 6, 260–312.

Hagstrom, P. 1998. Decomposing questions. Doctoral dissertation, MIT.

Haïk, I. 1990. Anaphoric, pronominal, and referential INFL. *Natural Language and Linguistic Theory* 8, 347–74.

Haïk, I., H. Koopman, and D. Sportiche. 1985. INFL en mooré et le liage dans le système A'. *Rapport de recherches du groupe de linguistique africaniste* année 1985–1986.

Halle, M. and A. Marantz. 1993. Distributed morphology and the pieces of inflection. In *The view from Building 20*, eds. K. Hale and S. J. Keyser, 111–76. Cambridge, MA: MIT Press.

Harley, H. 2004. Merge, conflation, and head movement: the First Sister Principle revisited. In *Proceedings of NELS 34*. University of Massachusetts, Amherst: GLSA.

Henderson, B. 2006. The syntax and typology of Bantu relative clauses. Doctoral dissertation, University of Illinois, Urbana-Champaign.

Henry, A. 1995. *Belfast English and Standard English: dialect variation and parameter setting*. Oxford: Oxford University Press

Heycock, C. 2005. On the interaction of adjectival modifiers and relative clauses. *Natural Language Semantics* 13, 359–82.

Hinzen, W. 2006. *Mind design and minimal syntax*. Oxford: Oxford University Press.

Hiraiwa, K. 2001. Multiple agree. Ms., MIT.

Hiraiwa, K. 2003. Cyclic locality. Ms., MIT.

Hiraiwa, K. 2005. Dimensions of symmetry in syntax: agreement and clausal architecture. Doctoral dissertation, MIT.

Hornstein, N. 2001. *Move! A minimalist theory of construal*. Oxford: Blackwell.

Hornstein, N. 2005. Deriving C-command. Ms., University of Maryland.

Hornstein, N. and A. Weinberg. 1995. The Empty Category Principle. In *Government and binding and the minimalist program*, ed. G. Webelhuth, 241–96. Oxford: Blackwell.

Hornstein, N., J. Nunes, and K. K. Grohmann. 2006. *Understanding minimalism*. Cambridge: Cambridge University Press.

Huang, C.-T. J. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation, MIT.

Huang, C.-T. J. 1993. Reconstruction and the structure of VP: some theoretical consequences. *Linguistic Inquiry* 24, 103–38.

Hulsey, S. and U. Sauerland. To appear. Sorting out relative clauses. *Natural Language Semantics*.

Ishii, T. 1999. Cyclic spell-out and the *that-t* effects. In *Proceedings of WCCFL* 18, 220–31. Somerville Cascadilla Press.

Jeong, Y. 2004. Multiple questions in Basque. Ms., University of Maryland.

Jeong, Y. 2006. The landscape of applicatives. Doctoral dissertation, University of Maryland.

Johnson, K. 2002. Towards an etiology of adjunct islands. Ms., University of Massachusetts, Amherst.

Kayne, R. S. 1984. *Connectedness and binary branching*. Dordrecht: Foris.

Kayne, R. S. 1989. Facets of Romance past participle agreement. In *Dialect variation and the theory of grammar*, ed. P. Benincà, 85–103. Dordrecht: Foris.

Kayne, R. S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kayne, R. S. 2000. Thoughts on antisymmetry. Presented at NELS 30, Georgetown University.

Kayne, R. S. 2002. Pronouns and their antecedents. In *Derivation and explanation in the minimalist program*, eds. S. D. Epstein and T. D. Seely, 133–66. Oxford: Blackwell.

Kayne, R. S. 2004. Antisymmetry and Japanese. In *Variation and universals of biolinguistics*, ed. L. Jenkins, 3–35. Oxford: Elsevier.

Kayne, R. S. 2005. Some notes on comparative syntax, with special reference to English and French. In *The Oxford handbook of comparative syntax*, eds. G. Cinque and R. S. Kayne, 3–69. Oxford: Oxford University Press.

Kayne, R. S. and J.-Y. Pollock. 1978. Stylistic inversion, successive cyclicity, and Move NP in French. *Linguistic Inquiry* 9, 595–621.

Kimenyi, A. 1980. *A relational grammar of Kinyarwanda*. Berkeley: University of California Press.

Kishimoto, H. 2005. Wh-in-situ and movement in Sinhala questions. *Natural Language and Linguistic Theory* 23, 1–51.

Kitahara, H. 1997. *Elementary operations and optimal derivations*. Cambridge, MA: MIT Press.

Ko, H. 2005. Syntactic edges and linearization. Doctoral dissertation, MIT.

Koopman, H. 2005. Korean (and Japanese) morphology from a syntactic perspective. *Linguistic Inquiry* 36, 601–33.

Koopman, H. and D. Sportiche. 1982. Variables and the bijection principle. *Linguistic Review* 2, 139–60.

van de Koot, H. 1989. The vacuous movement hypothesis, superiority and the ECP. In *Formal parameters of generative grammar IV*, eds. G. de Haan and W. Zonneveld, 41–51. Utrecht: OTS Yearbook 1988.

Koster, J. 2003. The configurational matrix. Ms., University of Groningen.

Kroch, A. and A. Joshi. 1985. The linguistic relevance of Tree Adjoining Grammar. Ms., University of Pennsylvania.

Larson, R. 1988. On the double object constructions. *Linguistic Inquiry* 19, 335–91.

Lasnik, H. 1999a. *Minimalist analysis*. Oxford: Blackwell.

Lasnik, H. 1999b. Chains of arguments. In *Working minimalism*, eds. S. D. Epstein and N. Hornstein, 189–215. Cambridge, MA: MIT Press.

Lasnik, H. 1999c. On the locality of movement: formalist syntax position paper. In *Functionalism and formalism in linguistics*, eds. M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, and K. Wheatley, 33–54. Amsterdam: John Benjamins.

Lasnik, H. 2000. *Syntactic structures revisited*. Cambridge, MA: MIT Press.

Lasnik, H. 2001a. When can you save a structure by destroying it? In *Proceedings of NELS* 31, 301–20. University of Massachusetts, Amherst: GLSA.

Lasnik, H. 2001b. A note on the EPP. *Linguistic Inquiry* 32, 356–62.

Lasnik, H. 2003. On the Extended Projection Principle. *Studies in Modern Grammar* 31, 1–23.

Lasnik, H. 2005. Review of Jason Merchant, *The syntax of silence*. *Language* 81, 259–65.

Lasnik, H. 2006. Conceptions of the cycle. In *Wh-movement: moving on*, eds. L. Cheng and N. Corver, 197–216. Cambridge, MA: MIT Press.

Lasnik, H. and M. Saito. 1992. *Move α*. Cambridge, MA: MIT Press.

Lasnik, H. and J. Uriagereka, with C. Boeckx. 2005. *A course in minimalist syntax*. Oxford: Blackwell.

Lebeaux, D. 1991. Relative clauses, licensing, and the nature of the derivation. In *Syntax and semantics 25: perspective on phrase structure: heads and licensing*, ed. S. Rothstein, 209–39. San Diego, CA: Academic Press.

Lebeaux, D. 1998. Where does the binding theory apply (II)? Ms., NEC Research Institute.

Lee, J.-E. 2004. Ditransitive structures and (anti-)locality. Doctoral dissertation, Harvard University.

Lee-Schoenfeld, V. 2005. Beyond coherence: the syntax of opacity in German. PhD dissertation, University of California, Santa Cruz.

Legate, J. 2003. Some interface properties of the phase. *Linguistic Inquiry* 34, 506–16.

Lightfoot, D. W. and A. Weinberg. 1988. Review of N. Chomsky, *Barriers*. *Language* 64, 366–83.

Lobeck, A. 1995. *Ellipsis*. Oxford: Oxford University Press.

Lutz, U., G. Müller, and A. von Stechow, eds. 2000. *Wh-scope marking*. Amsterdam and Philadelphia: John Benjamins.

Manzini, M. R. 1992. *Locality*. Cambridge, MA: MIT Press.

Manzini, M. R. 1994. Locality, minimality, and parasitic gaps. *Linguistic Inquiry* 25, 481–508.

Manzini, M. R. and A. Roussou. 2000. A minimalist theory of A-movement and control. *Lingua* 110, 409–47.

Marantz, A. 1984. *On the nature of grammatical relations*. Cambridge, MA: MIT Press.

Marantz, A. 1993. Implications of asymmetries in double object constructions. In *Theoretical aspects of Bantu grammar*, ed. S. Mchombo, 113–50. Stanford, CA: CSLI.

Marantz, A. 2000. Words. Ms., MIT.

Marušič, F. 2005. On non-simultaneous phases. PhD dissertation, State University of New York, Stony Brook.

Matushansky, O. 2005. Going through a phase. In *Proceedings of the MIT Workshop on EPP and Phases*, eds. M. McGinnis and N. Richards. Cambridge, MA: MITWPL.

May, R. 1985. *Logical form: its structure and derivation*. Cambridge, MA: MIT Press.

Mayr, C. 2006. Deriving T from (anti-)locality. Ms., Harvard University.

Mayr, C. and M. Reitbauer. 2005. Left-dislocation, agreement, and the notion of anti-locality. Ms., University of Vienna.

McCloskey, J. 1979. *Transformational syntax and model-theoretic semantics: a case study in Modern Irish*. Dordrecht: Reidel.

McCloskey, J. 2000. Quantifier float and wh-movement in an Irish English. *Linguistic Inquiry* 31, 57–84.

McCloskey, J. 2002. Resumption, successive cyclicity, and the locality of operations. In *Derivation and explanation in the minimalist program*, eds. S. D. Epstein and T. D. Seely, 184–226. Oxford: Blackwell.

McGinnis, M. 1998. Locality in A-movement. Doctoral dissertation, MIT.

McGinnis, M. 2001. Variation in the phase structure of applicatives. In *Linguistics Variation Yearbook* 1, 105–46.

Merchant, J. 2001. *The syntax of silence*. Oxford: Oxford University Press.

Müller, G. 2004. Phase impenetrability and wh-intervention. In *Minimality effects in syntax*, eds. A. Stepanov, G. Fanselow, and R. Vogel, 289–325. Berlin: Mouton de Gruyter.

Murasugi, K. and M. Saito. 1995. Adjunction and cyclicity. In *Proceedings of WCCFL* 13, 302–17. Stanford, CA: CSLI.

Neeleman, A. and H. van de Koot. 2002. The configurational matrix. *Linguistic Inquiry* 33, 529–74.

Nevins, A. and P. Anand. 2003. Some AGREEment matters: a cross-linguistic generalization on reconstruction and agreement. Paper presented at WCCFL 22.

Nissenbaum, J. 2000. Explorations in covert phrase movement. Doctoral dissertation, MIT.

Nunes, J. 1995. The copy theory of movement and linearization of chains in the minimalist program. Doctoral dissertation, University of Maryland.

Nunes, J. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press.

Nunes, J. and J. Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3, 20–43.

Obenauer, H.-G. 1994. Aspects de la syntaxe A-barre. Thèse de doctorat d'état, Université de Paris VIII.

Ochi, M. 1999a. Constraints on feature checking. Doctoral dissertation, University of Connecticut.

Ochi, M. 1999b. Some consequences of Attact-F. *Lingua* 109, 81–109.

Ormazabal, J., J. Uriagereka, and M. Uribe-Etxebarria. 1994. Word order and wh-movement: towards a parametric account. Paper presented at GLOW 17, Vienna, April 1994.

Ortiz de Urbina, J. 1989. *Parameters in the grammar of Basque*. Dordrecht: Foris.

Ott, D. 2005. Phase syntax: a critical view on derivational cyclicity. Ms., University of Cologne.

Pearson, M. 2005. The Malagasy subject/topic as an A-bar element. *Natural Language and Linguistic Theory* 23, 381–457.

Pesetsky, D. 1995. *Zero syntax*. Cambridge, MA: MIT Press.

Pesetsky, D. 2000. *Phrasal movement and its kin*. Cambridge, MA: MIT Press.

Pesetsky, D. and E. Torrego. 2001. T-to-C movement: causes and consequences. In *Ken Hale: a life in language*, ed. M. Kenstowicz, 355–426. Cambridge, MA: MIT Press.

Pesetsky, D. and E. Torrego. 2004. Tense, Case, and the nature of syntactic categories. In *The syntax of tense*, eds. J. Gueron and J. Lecarmem, 495–537. Cambridge, MA: MIT Press.

Phillips, C. To appear. The real-time status of island constraints. *Language*.

Platzack, C. 2001. Multiple interfaces. In *Cognitive interfaces: constraints on linking cognitive information*, eds. U. Nikanne and E. van der Zee, 21–53. Oxford: Oxford University Press.

du Plessis, H. 1977. Wh-movement in Afrikaans. *Linguistic Inquiry* 8, 723–6.

Postal, P. 1972. On some rules that are not successive cyclic. *Linguistic Inquiry* 3, 211–22.

Postal, P. 1997. Islands. Ms., NYU.

Putnam, M. 2006. Prolific scrambling. Doctoral dissertation, University of Kansas.

Pylkkänen, L. 2002. Introducing arguments. Doctoral dissertation, MIT.

Rackowski, A. 2002. The structure of Tagalog: specificity, voice, and the distribution of arguments. Doctoral dissertation, MIT.

Rackowski, A. and N. Richards. 2005. Phase edge and extraction: a Tagalog case study. *Linguistic Inquiry* 36, 565–99.

Reintges, C., P. LeSourd, and S. Chung. 2006. Movement, wh-agreement, and apparent wh-in-situ. In *Wh-movement: moving on*, eds. L. Cheng and N. Corver, 165–94. Cambridge, MA: MIT Press.

Rezac, M. 2003. The fine structure of cyclic Agree. *Syntax* 6, 156–82.

Richards, M. 2004. Object shift and scrambling in North and West Germanic: a case study in symmetrical syntax. Doctoral dissertation, Cambridge University.

Richards, N. 1997. What moves where when in which language? Doctoral dissertation, MIT.

Richards, N. 1998. The Principle of Minimal Compliance. *Linguistic Inquiry* 29, 599–629.

Richards, N. 1999. Featural cyclicity and the ordering of multiple specifiers. In *Working minimalism*, eds. S. D. Epstein and N. Hornstein, 127–58. Cambridge, MA: MIT Press.

Richards, N. 2001. *Movement in language*. Oxford: Oxford University.

Richards, N. 2002. Very local A-bar movement in a root-first derivation. In *Derivation and explanation in the minimalist program*, eds. S. D. Epstein and T. D. Seely, 227–48. Oxford: Blackwell.

Richards, N. 2004. Against ban on lowering. *Linguistic Inquiry* 35, 453–63.

van Riemsdijk, H. 1978. *A case study in syntactic markedness: the binding nature of prepositional phrases*. Dordrecht: Foris.

Rizzi, L. 1982. *Issues in Italian syntax*. Dordrecht: Foris.

Rizzi, L. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.

Rizzi, L. 1996. Residual verb-second and the wh-criterion. In *Parameters and functional heads*, eds. A. Belletti and L. Rizzi, 63–90. Oxford: Oxford University Press.

Rizzi, L. 1997. The fine structure of the left periphery. In *Elements of grammar*, ed. L. Haegeman, 281–337. Dordrecht: Kluwer.

Rizzi, L., ed. 2004a. *The structure of IP and CP*. Oxford: Oxford University Press.

Rizzi, L. 2004b. Locality and left periphery. In *Structures and beyond*, ed. A. Belletti, 223–51. Oxford: Oxford University Press.

Rizzi, L. 2006. On the form of chains. In *Wh-movement: moving on*, eds. L. Cheng and N. Corver, 97–133. Cambridge, MA: MIT Press.

Rizzi, L. and U. Shlonsky. 2005. Strategies of subject extraction. Ms., University of Siena and University of Geneva.

Rooryck, J. and J. Costa. 1996. On pseudo-raising. In *Actes du second colloque Langues & Grammaire, Paris VIII*, eds. L. Nash, G. Tsoulas and A. Zribi-Hertz, 48–58.

Ross, J. R. 1967. Constraints on variables in syntax. Doctoral dissertation, MIT. [Published 1986 as *Infinite Syntax!* Norwood, NJ: Ablex.]

Ross, J. R. 1969. Guess who? In *Papers from the 5th regional meeting of the Chicago Linguistic Society*, 252–86. Chicago, IL: Chicago Linguistic Society.

Rudin, C. 1988. On multiple questions and multiple *wh*-fronting. *Natural Language and Linguistic Theory* 6, 445–501.

Sabel, J. 1998. Principles and parameters of wh-movement. Habilitation thesis, University of Frankfurt/Main.

Sauerland, U. 2003. Intermediate adjunction with A-movement. *Linguistic Inquiry* 34, 308–14.

Saito, M. and K. Murasugi. 1990. N'-deletion in Japanese. In *UConn Working Papers in Linguistics* 3, 86–103.

Schneider-Zioga, P. 1995. Specifier/head agreement in Kinande. *Cahiers Linguistiques D'Ottawa* 23, 67–93.

Schneider-Zioga, P. 2000. Anti-agreement and the fine structure of the left periphery. *University of California Irvine Working Papers in Linguistics* 6.

Schneider-Zioga, P. 2002. The case of anti-agreement. In *Proceedings of MITWPL AFLA* 8. Cambridge, MA: MITWPL.

Schneider-Zioga, P. 2004. Anti-agreement, shortest chain and minimality: the syntax of dislocated subjects. Ms., UC Irvine.

Schneider-Zioga, P. 2006. Wh-agreement is not sufficient evidence for successive cyclic movement. Ms., UC Irvine.

Schneider-Zioga, P. To appear. Anti-agreement, anti-locality, and minimality: the syntax of dislocated subjects. *Natural Language and Linguistic Theory*.

Sells, P. 1984. Syntax and semantics of resumptive pronouns. Doctoral dissertation, University of Massachusetts, Amherst.

Sener, S. 2006. (Non-)P-stranding in Turkish and anti-locality. Ms., University of Connecticut.

Seuren, P. 2004. *Chomsky's minimalist program*. Oxford: Oxford University Press.

Sportiche, D. 1988. A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19, 425–49.

Sportiche, D. 1989. Le mouvement syntaxique: contraintes et parameters. *Langages* 95, 35–80.

Sportiche, D. 2005. Division of labor between Merge and Move. Ms., UCLA.

Stateva, P. and Stepanov, A. In press. Successive cyclicity as residual wh-scope marking. *Lingua*.

Starke, M. 2001. Move dissolves into merge: a theory of locality. Doctoral dissertation, University of Geneva.

Stepanov, A. 2001. Cyclic domains in syntactic theory. Doctoral dissertation, University of Connecticut.

Stjepanović, S. 1999. What do scrambling, second position cliticization, and multiple wh-fronting have in common? Doctoral dissertation, University of Connecticut.

Stjepanović, S. 2003. Multiple wh-fronting in Serbo-Croatian matrix questions and the matrix sluicing construction. In *Multiple wh-fronting*, eds. C. Boeckx and K. K. Grohmann, 255–84. Amsterdam: John Benjamins.

Svenonius, P. 2004. On the edge. In *Peripheries: syntactic edges and their effects*, eds. D. Adger, C. de Cat, and G. Tsoulas, 259–87. Dordrecht: Kluwer.

Szabolcsi, A. 2006. Strong vs. weak islands. In *The Blackwell companion to syntax*, eds. M. Everaert and H. van Riemsdijk, IV, 479–531. Oxford: Blackwell.

Takahashi, D. 1994. Minimality of movement. Doctoral dissertation, University of Connecticut, Storrs.

Ticio, E. 2003. On the structure of DPs. Doctoral dissertation, University of Connecticut.

Ticio, E. 2005. Locality and Anti-locality in Spanish DPs. *Syntax* 8, 229–86.

Torrego, E. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15, 103–29.

Tuller, L. 1985. Tense features and operators in Hausa. *Rapport de recherches du groupe de linguistique africaniste* année 1985–1986, 493–516.

Ura, H. 1996. Multiple feature-checking. Doctoral dissertation, MIT.

Uriagereka, J. 1988. On government. Doctoral dissertation, University of Connecticut, Storrs.

Uriagereka, J. 1998. *Rhyme and reason*. Cambridge, MA: MIT Press.

Uriagereka, J. 1999a. Multiple spell out. In *Working minimalism*, eds. S. D. Epstein and N. Hornstein, 251–82. Cambridge, MA: MIT Press.

Uriagereka, J. 1999b. Minimal restrictions on Basque movements. *Natural Language and Linguistic Theory* 17, 403–44.

Uriagereka, J. 2003. Spell-out consequences. Ms., University of Maryland, College Park.

Uriagereka, J. Forthcoming. *Syntactic anchors*. Cambridge: Cambridge University Press.

Wagner, M. 2005. Prosody and recursion. PhD dissertation, MIT.

Watanabe, A. 2002. Loss of overt wh-movement in Old Japanese. In *Syntactic effects of morphological change*, ed. D. W. Lightfoot, 179–95. Oxford: Oxford University Press.

Wexler, K. and P. Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Williams, E. 1983. Against small clauses. *Linguistic Inquiry* 14, 287–308.

Williams, E. 2003. *Representation theory*. Cambridge, MA: MIT Press.

Wurmbrand, S. 2004. No TP-fronting meets nearly headless Nick. Ms., University of Connecticut.

Yang, C. 2006. *The infinite gift*. New York: Schuster.

Zwart, C. J.-W. 1996. Shortest move vs. fewest steps. In *Minimal ideas*, eds. W. Abraham, S. D. Epstein, H. Thrainsson, and C. J.-W. Zwart, 305–27. Amsterdam: John Benjamins.

Index