Phasing in Full Interpretation

A dissertation presented

by

Hiroki Narita

to

The Department of Linguistics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Linguistics

Harvard University

Cambridge, Massachusetts

May 2011

Thesis advisor   Author

Professor C.-T. James Huang   **Hiroki Narita**

Phasing in Full Interpretation

Abstract

Contrary to the conventionally assumed system of labeled phrase structure, this thesis will argue for the hypothesis that the theory of bare phrase structure (Chomsky 1994, 1995 *et seq.*) should assume no mechanism of labeling/projection. It will be argued that the Inclusiveness Condition and the No-tampering Condition necessitate, among other things, structure-generation without labeling/projection and cyclic derivation by phase. It will be further claimed that this restrictive model of syntax always generates, phase by phase, syntactic objects (SOs) of the form {H, α}, where H is a lexical item (LI) and α an SO. This restriction, referred to as the H-α schema, will be shown to derive a number of empirically grounded locality constraints in tandem with the Phase-Impenetrability Condition (Chomsky 2000, 2004, 2008), a desirable result both in terms of breadth of empirical coverage and theory-internal simplicity. Another important consequence is that the notion of endocentricity/headedness can no longer be attributed to (X-bar theoretic) projection. It will be claimed that each SO {H, α} can be efficiently assigned endocentric interpretation by the Conceptual-Intentional system (CI) and the Sensorimotor system (SM) via minimal search of LIs—H is always the head of {H, α}, primarily configuring the interpretation of the SO. Full and consistent satisfaction of Full Interpretation is thereby achieved, maximally corroborating the desideratum of the Strong Minimalist Thesis.

Table of Contents

Acknowledgments

At various stages of developing this thesis, I received many helpful comments and suggestions from a number of people, to whom I am really grateful. I am sure I will fail to complete the list of the people who have kindly been part of the process in organizing my thoughts on the relevant topics in my PhD study at Harvard University.

At least I know where to start, but not quite how. The thing is, I cannot express how deeply I am honored to have had the dissertation committee that I did: C.-T. James Huang, Noam Chomsky, Naoki Fukui and Wesley Jacobsen. First and foremost, I want to express my deep gratitude to my advisor, C.-T. James Huang. From the earliest days of my graduate study at Harvard, he has unfailingly provided generous advice and encouragement concerning not only my research on linguistics proper but also my career establishment in academia. I am deeply thankful to all the opportunities and support that Jim has kindly offered to me throughout my graduate study. Without Jim's unfailing support and faith in me for the past four years, I could never have gotten this far. Specifically for linguistics, Jim never fails to keep reminding me of the classic ideas from the earlier GB framework (ECP, government, CED, etc.), and I agree with him that not a few of the problems that had motivated them are still left unanswered today. Since complex problems by no means grant complication of the theory (say a dozen different notions of 'government'), I have been struggling to find the right balance between the complexity of data and the simplicity of the theory. The present thesis is what I managed to provide through my attempt at reconciling the conflict, but I don't know if Jim found it fully convincing.

As is the norm for linguists of my age, I have entered into the circle of contemporary linguistics by taking in all the legends and anecdotes about Noam Chomsky and his work, and when I came to Cambridge I was under the impression that he was literally an unreachable being or the like. But he was not, and I was rather surprised at how generous and friendly he has been to me since our first meeting. Despite his busy schedule, he has been extremely generous in finding time for appointments and e-mail discussions, and I was able to meet him every other week or so for almost two years. I rather felt like he was

a friendly uncle of mine, with whom I was able to talk about and share my personal life and worries. His frankness was of course often equated with his being thoroughly critical on every corner of my work, and I am sincerely thankful for all the valuable discussions that I had with him, which have led to substantial revisions of my past proposals. Readers will never fail to notice the tremendous influence that Noam has had on me, although the conclusions that I reached through this study rather significantly depart from Noam's ideas in some respects.

Since my previous master's study at Sophia University in Tokyo (or even before that), Naoki Fukui has been a major source of insights into a number of issues, not only in theoretical comparative syntax but also in various points of intersection of linguistics and other fields, like mathematics, philosophy and physics. Through enlightening discussions with him, I became fully convinced that a serious comparative syntactician must be at the same time an earnest minimalist and vice versa, and that the pursuit of theoretical simplicity is the unifying force of macroparametric syntax and biolinguistic minimalism. I have been thrilled and honored that my former mentor back in Japan could also be part of my dissertation committee, and I am really indebted to him for his comments and suggestions at various stages of my career in linguistics.

I am also deeply grateful to Wesley Jacobsen. I was really lucky to have him as a teacher and also as a friend for the four years of my PhD study at Harvard. He helped me to explore the syntax and semantics of Japanese and other languages from various points of view, and also critically pointed out a number of theory-laden but unwarranted assumptions and shortcomings hidden in the earlier versions of this work, for which I am really thankful. The dissertation committee that I had was the best of the best on any imaginable scale, and I am tremendously grateful for the opportunity to work with them.

Moreover, this work would never have been possible without the support from quite a lot of people in addition to the committee members mentioned above. First of all, I would also like to express my special gratitude to Cedric Boeckx, who has been a

major source of enlightening discussions on a number of issues, not only in syntax but also on semantics and philosophy of language. I am very grateful to the consistent and warm support for my research that he kindly provided since the very beginning of my admission to Harvard, and which has lasted even after he left this university in 2009. I am also sincerely grateful to those people who have been at Harvard and have provided great sources of joyful experience and enlightening discussions with me: Sun-hee Bae, Gennaro Chierchia, Lauren Eby, Hannes Fellner, Laura Grestenberger, Hiroshi Hasegawa, Hidehito Hoshi, Toru Ishii, Peter Jenks, Soo-Yeon Jeong, Yongmi Jeong, Li Julie Jiang, Beste Kamali, Jenny So-Yeon Lee, Hsu-Chen Daphne Liao, Chi-Ming Louis Liu, Clemens Mayr, Andreea Nicolae, Alexander Nikolaev, Dennis Ott, Paul Pietroski, Hazel Pearson, Maria Polinsky, Patrick Rich, Tom Roeper, Jacopo Romoli, Bridget Samuels, Greg Scontras, Emma Shoucair, Sverre Johnsen, Hiroyuki Tanaka, Yasutada Sudo, Ulutas Suleyman, Cheng-Yu Edwin Tsai, Yimei Xiang, Yoko Yamazaki, and Kyoko Yamakoshi. Many thanks are also to those who provided generous administrative support for me to finish my degree, especially Cheryl Murphy, Kate Pilson, Helen Lewis, and Bev Stohl. I cannot imagine what my life would be without all the support and encouragement that I got from them.

Outside of the Harvard department of linguistics, I am grateful to Leah Bauke, Jonathan David Bobaljik, Željko Bošković, Samuel Epstein, Koji Fujita, Ángel Gallego, Erich Groat, Hironobu Kasai, Takaomi Kato, Hisatsugu Kitahara, Yukino Kobayashi, Masakazu Kuno, Terje Lohndal, Marc Richards, Norvin Richards, Daniel Seely, Michelle Sheehan, Masanobu Sorida, Juan Uriagereka, Masaya Yoshida, the audience of the 32nd GLOW Colloquium (Nantes University, April 15th-18th, 2009) and the Conference on Minimalist Approaches to Syntactic Locality (the Research Institute for Linguistics of the Hungarian Academy of Sciences, Budapest, Hungary, 26-28 August 2009), and the two anonymous reviewers for Narita (forthcoming) for their valuable comments and suggestions on earlier versions of this work. I would like to also thank Yuko Asada, Michael Mitcho Yoshitaka Erlewine, Gen Fujita, Kazuki Iijima, Toshikazu Ikuta, Ako Imaoka, Yusuke Imanishi, Koji

Kawahara, Junya Nomura, Miki Obata, Ayaka Sugawara, Yu Tanaka, Masairo Yamada, and Hideto Yoshida for their encouragement and support on various occasions.

I would also like to mention that my strong inclination toward a strictly syntactic, internalist approach to the syntax-semantics interface is the underlying theme of the present thesis. We know very little about semantics/pragmatics, or more generally the creative use of linguistic knowledge, and I became convinced that what little we can do to investigate this ill-understood domain is to first explore what the well-defined empirical study of syntax can tell us about it, not to start with conventional presuppositions as to what semantics ought to be. This research attitude was fostered and polished through my experience in participating in discussions with Cedric Boeckx, Noam Chomsky, Koji Fujita, Naoki Fukui, Wolfram Hinzen, Norbert Hornstein, Yu Izumi, Terje Lohndal, James McGilvray, Dennis Ott, Paul Pietroski, Marc Richards, Bridget Samuels, Wataru Uegaki, and Juan Uriagereka. I am thankful for all the opportunities, as well as puzzlements, that I managed to expose myself to through participating in discussions with them.

When I got an admission to Harvard among other schools, Naoki convinced me that Cambridge was a place where arguably the best of the best set of prominent researchers gather from all corners of the world from all branches of science, and that it would be the most important and invaluable experience to get acquainted with these people through my stay at Harvard, given the interdisciplinary nature of linguistic research. Naoki was absolutely right, and I am really honored to make friends with all the brilliant nonlinguist people in Cambridge, whose majors and interests crosscut everything I am hoping to study more in the future. They are too many to list, but I would like to especially thank Takeru Furuya, Michinao Hashimoto, Yaling Hsiao, Yuichiro Kamada, Takuya Kitagawa, Kei Kobayashi, Tatsuo Okubo, Masahiro Ono, Yusuke Tsuno, Yuichi Yamamoto, Chien Yan-Ting, and Teruyoshi Yoshida for their lasting friendship, encouragement and support.

Last but not least, I would like to express my sincere gratitude to my parents Shinichi Narita and the late Makiko Narita, and my grandmother, Etsu Narita. I know it

hasn't been easy on my family to have me abroad for the past four years. Thanks are also to my late dog Pero, who had been with me for more than seventeen years until he died early this year, 2011. I cannot express my gratitude enough for their love and support, and for making my life possible.

I dedicate this dissertation to all the people who believed in me: friends, teachers, and family.

Chapter 1

Preface

1.1   The Goal of Linguistics

One of the distinct characteristics of human beings is that each individual can acquire one or more grammars of natural language from its first three to four years of experience. The grammar of natural language (*I-language*) is in essence a computational system that generates an infinite array of structured expressions that can be assigned 'sound' (or 'signs') and 'meaning'. Any human infant, apart from serious pathology, somehow manages to acquire such a system of discrete infinity from a finite set of primary linguistic data, with remarkable speed and uniformity. This course of biological growth is uniquely human and shared with no other species, even the genetically closest primates, not to mention kittens and songbirds, despite intensive training. This fact about the human species suggests that the human brain is biologically endowed with a substantive innate mechanism that constitutes the basis of the acquisition/growth as well as the use of the linguistic capacity. This biological endowment, the *faculty of language* (FL) as we may call it, is reasonably assumed to be a distinct module of the human mind/brain, with its own internal rules and constraints.

Chomsky (2005, 2007a) reminds us of a virtual truism that the design of FL, or of any biological system for that matter, should be attributed to three factors:

(1)   Three factors in the language design (Chomsky 2005 and others):

   (i)   genetic endowment

   (ii) external stimulus/environment

   (iii) physical and mathematical ('computational' in particular, in the case of digital computational systems like language) principles that are not specific to FL

Under the current formulation, Universal Grammar (UG), a theory of the initial state of FL, meets the condition of *explanatory adequacy* if it provides an encompassing description of factor (i), constructed in such a way that it maps linguistic data (factor (ii)) to acquired I-languages under the effect of factor (iii) (Chomsky 1965, 2008). The early generative enterprise regarded explanatory adequacy as the ultimate goal for linguistic theory, and, as earlier descriptive studies revealed more and more complicated facts about attested grammars, it went in the direction of enriching UG under the pressure of *descriptive adequacy*, a criterion of encompassing description of observable I-languages. The 80s' and 90s' practice of theoretical comparative syntax in the so-called *Principles-and-Parameters* (P&P) framework provides an explosion of unforseen descriptive studies of various I-languages, a tremendous success in the thousands of years of linguistic inquiry. Specifically, this line of approach has pinned down various language-specific principles, articulated syntactic structures with hierarchically assembled lexical and functional categories, loci of parametric variation distributed over various components of FL, and so on. However, once the P&P approach provided a sufficiently rich first approximation of the character of UG, linguists started to seek a yet deeper level of explanation of human language, beyond explanatory adequacy, by asking *why* FL is designed the way it is, not many other imaginable ways. This is essentially a question of the ontogeny and phylogeny of FL, and the answer to this

why-question should be formulated in biophysical (i.e., not language-specific) terms, with the hope for eventual unification of biolinguistics with other 'core' natural sciences such as biology and physics. Then, let us say that a theory of FL meets the condition of *biological adequacy* if it provides an explanation for how the emergence of FL was/is biologically possible in the evolution/development of human beings.[1]

One of the prominent research programs that seek a simple and hence a (more) biologically adequate theory of FL is called the *minimalist program for (bio)linguistic theory* (MP). The research agenda of the MP is, in essence, to attribute the effects of what has been postulated as constituting factor (i) of the language design (human-specific genetic endowment of FL that is left unexplained) to factor (iii), laws of nature whose explanatory forces are much beyond the confinement of human language and can be further investigated under the rubric of general physics (Chomsky 2005). That is to say, the less we can show UG (the model of factor (i)) contains, the closer we get to the goal of biological adequacy. This enterprise naturally seeks the simplest possible characterization of UG by continuously refining the postulated model of FL through eliminating unnecessary redundancies and unwarranted stipulations, as well as the nature of the third factor principles that enter into the design of human language.

The minimalist quest for a simpler, and hence more explanatory, theory of FL is guided by a leading hypothesis called the *strong minimalist thesis* (SMT), as formulated in (2):

(2)   ***Strong Minimalist Thesis (SMT)*** :[2]

FL is an optimal solution to the conditions imposed by CI and SM.

---

[1]See Narita (2010) for relevant discussion. We may alternatively call this desideratum *evolutionary adequacy*, a term suggested independently by Longobardi (2003) and Fujita (2007, 2009) with more or less the same import. See also Boeckx and Uriagerekas (Boeckx and Uriagereka 2007) discussion of 'natural adequacy'.

[2]See Chomsky (2007a), Narita (2009b) for some clarification of the SMT.

The SMT holds that FL is a perfect computational system that contains nothing more than what is minimally necessary for the efficient linkage of the performance systems that it interfaces with (Chomsky 2000, 2007a). Underlying the SMT is the desideratum that the MP will provide, in due course, a way to relate the properties of FL to simplicity and elegance that constitute a fundamental property of nature, as the modern natural sciences have revealed. Thus, the SMT is intended to serve not only as a heuristic similar to Ockham's razor, guiding us towards the elimination of redundancies in theory-formation, but also as an expression of the fundamental empirical conjecture that the MP seeks to demonstrate (or evaluate).

Under appropriate idealization, FL reduces to a computational system that (A) assembles a finite number of mental representations into structured linguistic expressions, and (B) hands them to the outside systems it interfaces with. To adopt a familiar term, let us call this computational aspect of FL *syntax*. One of the core (and irreducible) operations in FL to fulfill function (A) is the operation *Merge*, which maps two syntactic objects (SOs) to a set of these SOs:

(3)   **Merge**$(\alpha, \beta) = \{\alpha, \beta\}$.

Optimally, the structure-building operation of FL (A) is entirely reduced to Merge. The theory of FL that assumes only Merge to capture (A) is called the framework of *bare phrase structure* (Chomsky 1994, 1995 et seq.), a hypothesis that I will explore in the present study.

Given that language is a system that relates 'sounds' and 'meanings', it is natural to suppose that there are at least two performance systems that FL interfaces with, which are referred to as the *Conceptual-Intentional system* (CI) and the *sensorimotor system* (SM), respectively. The former utilizes structured linguistic expressions for human thoughts, and the latter assigns either 'sounds' or 'signs' to linguistic expressions for communication and other purposes (see, e.g., Petitto 2005 and many others for the modality-independence of human language). We further call linguistic representations that are generated by FL

and interface with CI and SM as *SEM* and *PHON*, respectively. SEM and PHON should be structured in such a way that they can be used by CI and SM, satisfying whatever conditions of usability are imposed by them. As for the interfacing function (B), then, let us call the operations that hand SOs generated by syntax to SEM and PHON *Interpret* and *Spell-Out*, respectively (cf. Chomsky 1995, Lasnik et al. 2005).

(4)   ***Interpret*** maps an SO to SEM.

(5)   ***Spell-Out*** maps an SO to PHON.

These operations, Merge, Interpret and Spell-Out, are necessary for any theory of UG, as long as FL is characterized as a human-unique computational system that generates an infinite number of linguistic expressions interpretable by CI and SM. The MP then guides us to an endeavor that attempts to explain/derive as many properties of FL as possible by means of this very limited set of apparatus available in UG. Any UG stipulations beyond these operations would have to meet a heavy burden of empirical justification.

The question then is how these three operations interact in syntactic derivation. So long as the SMT holds, these three operations should apply in a computationally efficient fashion, presumably due to the effect of factor (iii). The content as well as the effect of the third factor principles of computational efficiency is an empirical matter, and any hypotheses about this should be tested against the facts about human language. Needless to say, we only have a limited understanding of such principles, and the hope is that we will eventually learn more about the third factor principles by studying how the architectural properties of FL contribute to the satisfaction of the conditions they impose (cf. Chomsky 2008:135-136). The investigation of the effect of the third factor is the central topic of the present study.

1.2   Towards Fuller Exploration of Bare Phrase Structure

The history of generative biolinguistics can be characterized in significant part by the development of the theory of phrase structure. In the earliest tradition of transformational generative grammar initiated by Chomsky (1955/1975, 1957), it was assumed that the bifurcated system of *phrase structure rules* (PSRs) and *transformational rules* (*transformations*) is necessary to capture basic properties of phrase structure.[3] Specifically, it was assumed therein that the skeletal structure of a sentence (the kernel of the language (Chomsky 1955/1975, 1957), later called the 'd(eep)-structure' (Chomsky 1965)) is generated by a finite set of PSRs, each of which takes the following form, where A is a single symbol, W a non-null string of symbols, and X and Y possibly null strings of symbols that indicate the 'context' in which rewriting of A to W can apply:

(6)   XAY → XWY

Applying to nonterminal symbols (starting with a designated initial symbol S′), PSRs generate phrase-markers that express the basic structural facts of phrase structure, in particular the following three kinds of information:

(7)   a.   Compositional structure: the hierarchical and combinatorial organization of linguistic constituents

      b.   Labeling: the 'type' of each constituent

      c.   Precedence: the left-to-right order of the constituents

For example, the PSRs in (8) generate phrase-markers like the one in (9).

(8)   a.   S′ → COMP S

      b.   S → NP Infl VP

---

[3]See van Riemsdijk and Williams (1986), Fukui (2001), Lasnik et al. (2000), Carnie (2002, 2008) among others for overview. For detailed discussion on the nature of PSRs and their relation to phrase-markers, see Chomsky (1955/1975, 1957, 1959), McCawley (1968), Lasnik et al. (2000).

    c.   Infl → Present, Past, will, . . .

    d.   VP → V NP

    e.   NP → (D) N

    f.   D → the, a, . . .

    g.   N → boy, mother, student, apple, leaf, . . .

    h.   V → see, eat, hit, make, open, touch, . . .

(9)

```
                    S'
          ┌─────────┴─────────┐
        COMP                  S
                    ┌─────────┼─────────┐
                   NP        Infl       VP
               ┌────┴───┐     │     ┌────┴────┐
               D        N    will   V         NP
               │        │          │      ┌───┴───┐
              the      boy        eat     D       N
                                          │       │
                                         the    apple
```
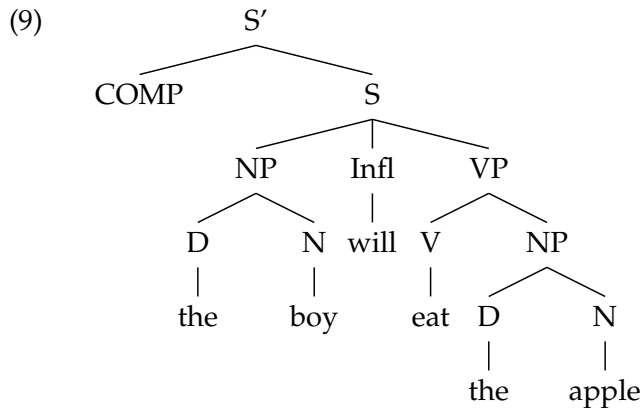
The phrase-marker (9) indicates, for example, that the largest constituent is labeled by an intial symbol S; that S is made up of a constituent NP (Noun Phrase), Infl(ection), and VP (Verb Phrase); that NP precedes Infl, which in turn precedes VP; that the NP consists of two constituents, D(eterminer) and N(oun), in this order; that the phrase consisting of the sequence V—NP is labeled by VP, and so on. Each of the PSRs in (8) encapsulates information about compositional structure, labeling and left-to-right ordering of constituents in phrase-markers.[4] Phrase-markers generated by PSRs are then mapped to the corresponding 'transformational structures' (s(urface)-structures, LF, and so on) by means of transformations, that is, mappings from a phrase-marker to another phrase-marker. It was further assumed that the space of possible transformational structures was also predeter-

---

[4]Note that the system of PSRs does not determine the actual temporal order of rule-applications in a psychological setting. Specifically, each nonterminal symbol can be seen as either the 'input' to, or the 'output' of, an application of a PSR: thus, one may see S as serving as the input to rule (8b) that generates the sequence NP—Infl—VP as the output, or one may also regard the sequence NP—Infl—VP as the input to the same rule that generates (or 'projects') S as the output.

mined by the same system of PSRs (see Emonds 1970, 1976 among others).

The development of the theory of transformational grammar observed the crystallization of the *X-bar theory* initiated by Chomsky (1970). Historically, this hypothesis was put forward as a way to remedy one of the fundamental inadequacies of PSRs pointed out by Lyons (1968), which is that PSRs alone cannot capture the basic fact that phrase structure in human language is *endocentric* (*headed*).

(10) ***Endocentricity (headedness) of phrase structure*** :

The interpretive properties of a phrase Σ are determined largely by the features of a prominent lexical item within Σ (referred to as the *head* of Σ).

In general, there is a certain central element, called the *head*, for each phrase that determines the essential properties of the phrase, thus a verb phrase is predicative due to the property of the head verb, referential properties of a noun phrase is determined by the head noun, and so on. Lyons correctly points out that the system of PSRs is inadequate or insufficient in that it fails to capture the fact that XP always dominates X (NP dominates N, VP dominates V, ...), and therefore, no formalism of PSRs excludes rules of the following sort, unattested and presumed to be impossible in natural language but formally comparable to (8).

(11)   a.   NP → VP PP

b.   PP → D S Infl V

c.   AP → COMP NP

:

The X-bar theory was put forward by Chomsky (1970), Jackendoff (1977) and much subsequent work essentially to overcome this inadequacy of the theory of PSRs. It holds that the class of possible PSRs can be radically reduced to the following two general schemata, where a lexical item X is necessarily dominated by an intermediate category X′, which in turn is necessarily dominated by the maximal category X″ (Chomsky 1970; see also

Jackendoff's 1977 tripartite X-bar structure).

(12)   the X-bar schemata

    a.   X′ → X (Y″) or (Y″) X

    b.   X″ → (Z″) X′

For example, the X-bar schemata generates the following phrase-marker from the same set of terminal elements as those in (9), assuming the DP-analysis of nominals (Brame 1981, 1982, Fukui and Speas 1986, Fukui 1986/1995, Abney 1987) and the head-initial linear order for English:

(13)

```
                    C″
                    |
                    C′
           _____/_____
          C                  I″
                     _____/_____
                    D″                 I′
                    |           _____/_____
                    D′          I             V″
              _____/\_____      |             |
             D          N″     will           V′
             |          |              _____/_____
            the         N′            V             D″
                        |             |             |
                        N            eat            D′
                        |                     _____/\_____
                       boy                   D          N″
                                             |          |
                                            the         N′
                                                        |
                                                        N
                                                        |
                                                      apple
```
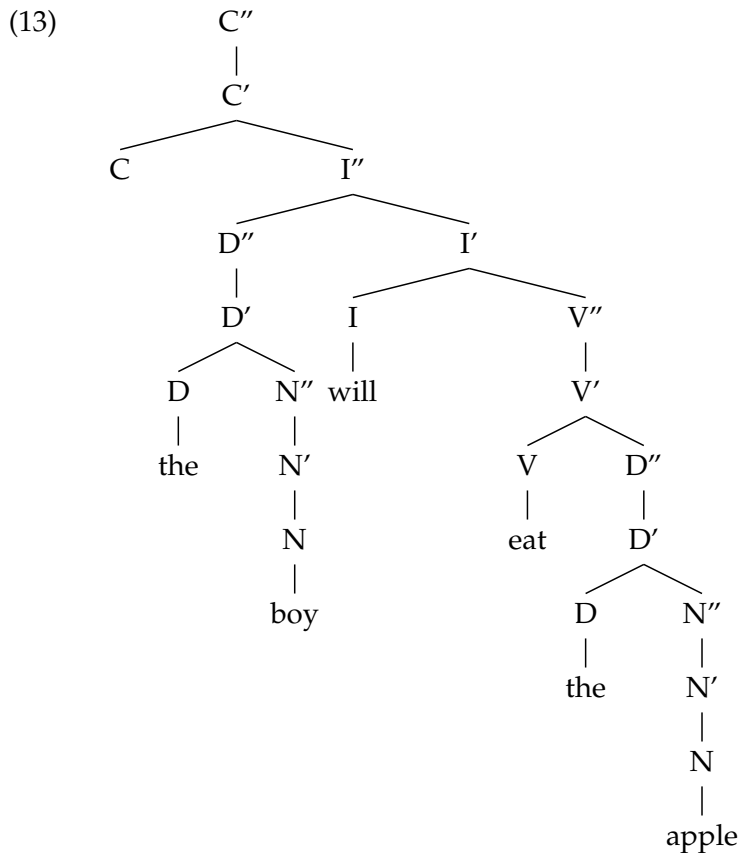
The X-bar theory further holds that featural properties of the head X are 'projected' up to the categories of X′ and X″. According to this conception of phrase-markers, labeling of nonterminal nodes is essentially achieved by the combination of the lexical features of X

and bar-level indices. In this theory, there is a strong sense in which nonterminals like X′ and X″ are 'projections' of X: N′ and N″ are projections of N, V′ and V″ are projections of V, and so on. We may refer to this aspect of the X-bar-theoretic hypotheses as the *labeling-by-projection hypothesis*. The hypothesized projection of lexical features of the head via bar-level projections rather straightforwardly captures the endocentricity/headedness of phrase structure (cf. Lyons 1968).[5],[6]

The X-bar theory is so strong a generalization over the possible form of PSRs that symbol-specific PSRs of the sort exemplified in (8) can be entirely eliminated, a highly desirable result acknowledged by Stowell (1981), Chomsky (1986a) among others. Nonetheless, it should be noted that the X-bar schemata in (12) are still formulated in the form of PSRs. Correspondingly, just like PSRs, the X-bar schemata still make direct claims about the three kinds of information in (7), namely compositional structure, labeling and left-to-right ordering. Thus, X-bar-theoretic phrase-markers directly encode linear ordering of the relevant constituents, though the space of variation is significantly restricted by the 'directionality parameter' as formulated in (12a), and also by the hypothesized universal of 'specifier-left' ordering as in (12b) (see Chomsky 1981; see also Kayne 1994 and much subsequent work). Moreover, the X-bar theory still assumes a system of 'labeled' phrase-markers, whereby each phrasal node in a phrase-marker is associated with a distinct nonterminal symbol, a label, each of which is though radically reduced to combinations of features of head items and bar-level indices, in accord with the labeling-by-projection hypothesis.

The compounding of these three categories of information into a single rule-schema is in itself an empirical claim to be tested and justified against facts about human language. Later scrutiny suggests, however, that the system of PSRs (even if supplemented

---

[5]See Fukui and Speas (1986), Fukui (1986/1995, 1988) for the later development of relativized X-bar theory, whereby the presence of the 'double-bar' projection X″ is parametrized across functional categories. See also Kuroda (1988).

[6]Moreover, the discovery of 'single-bar-level' constituents such as V′, N′ and I′ constitutes a real step toward descriptive adequacy as well, meeting a lot of empirical support as amply reported by Jackendoff (1977) among many others.

with the X-bar schemata) encapsulates several unjustifiable complications. Among other things, it was quickly pointed out that the encoding of precedence in terms of PSRs (or schemata) is highly redundant, and that left-to-right ordering of the relevant constituents is largely predictable by some independent principle of UG, probably applying at the phonological component.[7] Furthermore, evidence accumulated in the literature suggests that linear order actually plays little role in syntactic computation of SEM: Reinhart (1976, 1981, 1983) among others points out that purely hierarchically determined relations like c-command are sufficient to determine the conditions on binding. This strongly suggests that precedence may not be a core property of phrase-markers that persists throughout the derivation, as the system of PSRs predicts, but rather may be assigned relatively 'late' in linguistic computation, probably post-syntactically: the less relevant linear order is shown to be to syntactic computation, the less plausible it becomes to encapsulate precedence into the core rule-schemata of structure-generation.[8]

Importantly, Berwick and Chomsky (2008) further point to the stronger conclusion that there *must not* be any information about linear order available in syntax, so long as we seek to provide a principled explanation for the *structure dependence* of grammatical rules. The relevant generalization, arguably one of the central discoveries made in the generative linguistics tradition (see Chomsky 1957, 1965 *et seq.*), is that all the grammatical rules available in linguistic computation are structure-dependent, *i.e.*, formulable only in terms of hierarchical organization (compositional structure) of linguistic constituents.[9]

---

[7]Be it some sort of directionality parameter (Chomsky 1981; see also Richards 2004, 2007a), some version of Kayne's (1994) Linear Correspondence Axiom (LCA; cf. Chomsky 1995, Moro 2000, Uriagereka 1999 and Sheehan in press), Fukui and Takano's (1998, 2000) Symmetry Principle, or still some other means yet to be discovered. See Chapter 4 for further discussion.

[8]But see also Kayne (1994, 2004a, 2010), Fukui (1993) and Saito and Fukui (1998) for the view that linear order plays role in grammar.

[9]One of the most used examples to support structure dependence (or the poverty of stimulus argument for it) concerns auxiliary fronting in English interrogative-formation (Chomsky 1968). Interrogative sentences in English are formed by fronting the correct auxiliary of semantically related declarative sentences. For example, the interrogative related to the declarative form *Young eagles that can fly can swim.* is formed by fronting the main clause auxiliary as in (ia), but fronting the linearly closer subordinate clause auxiliary produces an ungrammatical sentence as in (ib) (Chomsky 1965).

The non-existence of structure-*in*dependent rules seems very well founded empirically, and we are correspondingly interested in asking why structure dependence is such a central and exceptionless property of linguistic computation. Can we make sense of the ubiquity of structure dependence on principled grounds? Berwick and Chomsky suggest one plausible line of approach to this problem, namely to hypothesize that linear order really *isn't* part of syntax. If linear order is part of syntax, as PSRs are bound to predict, it becomes mysterious why syntactic rules cannot make any recourse to linear order, or why there cannot be any linear-dependent/structure-independent rules learnable by the child. Berwick and Chomsky are essentially urging us to find a way out of this dilemma, and as they suggest, the most straightforward solution to this problem is to construct a theory of syntax that excludes linear order from the core architecture of linguistic computation as a matter of principle.
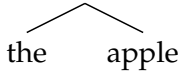
The advent of *bare phrase structure* (Chomsky 1994, 1995 *et seq.*) can be understood as a step forward to overcome the above-mentioned empirical inadequacy of PSRs. The theory of bare phrase structure holds that phrase-markers are generated by recursive application of *Merge*, in itself a quite elementary operation that simply combines two syntactic objects and forms a set of them.

(14)   Merge(α, β) = {α, β}.
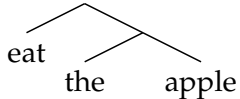
If applied to two lexical items *the* and *apple*, for example, Merge generates a set of them:

(i)   a.   Can$_i$ young eagles that can fly t_i swim?
       b.   *Can$_i$ young eagles that t_i fly can swim?

It is imaginable in principle that children can generate two types of rules, a linear-dependent (structure-independent) rule, by which the first *can* is to be moved, or the correct structure-dependent rule, by which only the movement of the *can* from the main clause is allowed. Crucially, children do not appear to go through a period when they erroneously entertain the hypothesis of linear-dependence, and they reach the correct structure-dependent rule without being ever exposed to the relevant evidence for inferring correct auxiliary fronting (Chomsky 1980b, Crain and Nakayama 1987).

(15)   Merge(the, apple)   =   a.   {the, apple}

b.   (order irrelevant)
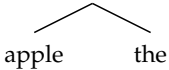
Importantly, Merge generates phrase-markers without specifying precedence. For example, (15a) is just a set, and hence it does not make any claim about the precedence-relation between *the* and *apple*. For expository convenience, we will sometimes represent Merge-based phrase-markers in the form of tree diagrams, as in (15b), but it should be understood throughout the present thesis that such representations make no claims about precedence.[10] Provided that Merge can apply recursively, the output of Merge can constitute another input to the same operation together with some other element, say *eat*, yielding another phrase-marker in (16).

(16)   Merge(eat, {the, apple})   =   a.   {eat, {the, apple}}

b.   (order irrelevant)

Phrase-markers generated by recursive Merge can provide minimal but sufficient information about compositional structure: the phrase-marker in (16) is a phrase that contains (or 'dominates' in earlier terms) three lexical items *eat*, *the* and *apple*; *apple* and *boy* form a subconstituent to the exclusion of *eat*; *eat* is hierarchically the most prominent element that c-commands all the other lexical items in the phrase-marker, and so on. Linear order is not implied in either input or output of Merge, a desirable result.

Incidentally, note that phrase-markers generated by recursive Merge as in (15) and (16) are 'label-less'. The earlier X-bar-theoretic approach would state that *the* (or *apple*)

---

[10]Understood as such, (i) is as precise a tree-graphic representation of (15a) as (15b) is.

(i)

'projects' in (15), or that *eat* 'projects' in (16), but nothing in the set-theoretic representations in (15)–(16) encodes such effects of labeling-by-projection. Phrase-markers generated by Merge such as (15) and (16) are by definition just sets, and no known condition of set theory requires that they are associated with any labels (distinct nonterminal symbols). Only further stipulations can guarantee labeling, and therefore we must ask if such enrichment of UG, should there be any, is really justifiable in terms of the SMT.

In this relation, it should be noted that nonterminal symbols (labels) used to constitute necessary input to and/or output of PSRs in the system of earlier transformational grammar: consider, e.g., S in S → NP Infl VP. This holds regardless of whether the symbols themselves are radically reduced to the combination of projected features and bar-level indices, as claimed by the X-bar theory. However, once we are provided with Merge as the alternative generative device for structure generation, Merge simply serves for compositional structuring of constituents, to the exclusion of labeling as well as precedence. This way, bare phrase structure departs from the earlier hypothesis that the means of structure generation (PSRs or schemata thereof) encompasses three kinds of information (compositional structure, labeling and linear order). To the extent that this move is empirically motivated with respect to linear order, a similar question arises as to labeling as well. Therefore, we have to ask if there is any strong empirical evidence that demands the theory of minimal syntax to be supplemented with a distinct mechanism of labeling.

In order to tackle this problem, it is instructive to recall that the labeling-by-projection hypothesis was originally advanced by Chomsky (1970) to capture the endocentricity of phrase structure (10). One may argue that the empirically well-established generalization (10) constitutes strong enough motivation for labeling even in bare phrase structure. However, such an argument would go through only if it is further shown that labeling by projection is *the* necessary device of commendable theory-internal simplicity to encode endocentricity. This was arguably the case in the earlier PSR-based syntax, where

labels are anyway generated as a necessary component of compositional structuring. But this is not true any more in bare phrase structure, where reference to labels may become even a worrisome departure from the SMT, as pointed out by Chomsky (2007a:23) (see also Collins (2002), Seely (2006) and Narita 2009a). Indeed, the theory of Merge-based syntax opens up a novel approach to the account of endocentricity that makes little to no recourse to labeling by projection, as envisaged by Chomsky (2000, 2008), Collins (2002), Narita (2009a) and as will be further explored in the present thesis.

Building on the prospect of bare phrase structure, the present thesis will aim at making a step toward constructing a theory of syntax without labeling/projection as well as linear order (precedence). That is, we will explore to a fuller extent the resolution of compositional structuring, left-to-right ordering and endocentricity of phrase-markers, as envisaged by the framework of bare phrase structure. Specifically, it will be argued that the proper treatment of the three kinds of information in (7) that were once encapsulated into PSR-schemata should be fully modularized into different components of FL: compositional structuring is fully taken care of by Merge applying in a computationally efficient fashion, maximally respecting bare phrase structure; the mechanism of linearization is properly relegated to the post-syntactic phonological component of FL; there is no mechanism of labeling/projection, and determination of endocentricity is reduced to some sort of minimal search principle, and hence presumably to the category of the third factor of language design along the lines suggested by Chomsky (p.c., lectures at MIT in fall 2010).

1.3   Previewing the *Why*-questions

The present thesis will explore how the theory of bare phrase structure can be advanced to meet various considerations imposed by the SMT. As a preview, I would like to list a set of several *why*-questions that the present inquiry will attempt to address. I do not claim that any of these questions are new, but the specific combination of these

considerations will constitute the major underlying themes of the present study.

The first *why*-question was already mentioned above.

(17)   **Why is linguistic computation always structure-dependent? (Why are there no structure-independent (e.g., linear-dependent) rules?)**

I argue in favor of the answer envisaged by the theory of bare phrase structure, namely that syntactic computation is always structure-dependent since syntax only generates compositional structure via Merge. No notion of linear order or precedence is thus available in syntax, and hence it admits no linear-dependent computation.

Chapter 2 will then pose the following question:

(18)   **Why are applications of insertion and tampering severely restricted in the syntax of natural language?**

I will argue that the *Inclusiveness Condition* (IC; Chomsky 1995, 2000), the *No-tampering Condition* (NTC), and the principle of *Full Interpretation* (FI; Chomsky 1995 *et seq.*) arise as specific manifestations of more general principles of computational efficiency.

(19)   **Inclusiveness Condition (IC)** :

No new features are introduced in the course of linguistic derivation.

(20)   **No-Tampering Condition (NTC)** :

No elements introduced by syntax are deleted or modified in the course of linguistic derivation.

(21)   **Full Interpretation (FI)** :

Every constituent of SEM and PHON contributes to interpretation.

I will further maintain that the IC and the NTC should be regarded as articulating the best answers to the question in (17): insertion and tampering are disallowed because the IC and

the NTC allow no such operations in syntax as a matter of principle.

I will then claim in Chapter 2 that the interplay of the IC, the NTC and FI derives the theory of cyclic derivation by phase of the sort advanced by Chomsky (2007a, 2008), which will constitute an answer to the third *why*-question:

(22)   ***Why does syntactic computation obey cyclic derivation by phase, and enforce locality constraints in this way?***

The theory of phases developed in Chapter 2 will constitute the crucial basis for the proposals articulated in the subsequent chapters. In particular, it will be claimed that Transfer, the composite of Interpret and Spell-Out, applies cyclically phase-by-phase, and it eliminates the interior of the phase to which it applies from the derivational workspace. I will argue in Chapter 3 that such elimination upon interpretation by Transfer constitutes the rationale for what is usually called the *Phase Impenetrability Condition* (PIC).

(23)   ***Phase-Impenetrability Condition (PIC)*** :

After Transfer applies to a phase Σ, the interior of Σ becomes inaccessible to further syntactic computation.

The PIC derives a number of empirically grounded locality constraints, like those traditionally subsumed under Huang's (1982) Condition on Extraction Domains (CED) among others.

As noted above and as will be discussed in detail in Chapter 3, the framework of bare phrase structure paves the way for exploring projection-free syntax. I will maintain that no mechanism of labeling/projection or feature-percolation is formulable in bare phrase sturcture, and hence that these should be eliminated from the theory of syntax. I will then point out that elimination of labeling/projection effectively imposes a strict constraint on the possible mode of Merge-application. Specifically, the following constraint on Merge, to be referred to as the *H-α schema*, will be proposed.

(24)   ***The H-α schema***: Merge(H, α) → {H, α}

Merge must take at least one LI as its input.

The H-α schema maintains that there can be no merger of two phrases. It will be observed that this constraint derives a number of empirically established generalizations regarding the possible 'base-structures' and 'displacement transformations' in human language. Specifically, it will be shown to derive (i) an empirically well-established typology of ph(r)asal-movement (or 'pied-piping'); and (ii) a uniform account of the CED and other locality constraints in tandem with the PIC. Several speculations on the nature of the H-α schema are also touched upon in this chapter, addressing the question of why the H-α schema holds at all.

(25)   ***Why is there no XP-XP merger?***

I will argue that the H-α schema is sustainable only in the framework of phase theory developed in Chapter 2.

Moreover, I will argue that the H-α schema, by virtue of being a strong constraint on the possible forms of syntactic structures, provides an important analytical tool for the acquisition of LIs. The H-α schema will be claimed to also provide a partial answer to the question in (26).

(26)   ***Why is the acquisition of LIs so uniformly rapid and successful?***

Various derivational constraints similar to the H-α schema have been already proposed in the literature, in particular to make sense of the *antisymmetry* of linear ordering in natural languages, an alleged linguistic universal put forward by Kayne (1994) and much subsequent work. Specifically, Chapter 4 will discuss various antisymmetry-based 'H-α schemata' in the literature, including Uriagereka's (1999, 2009) version of the Linear Correspondence Axiom (LCA) (Kayne 1994), Sheehan's label-based resuscitation of

Uriagereka's LCA, and Kayne's (2010) LCA-free alternative account of antisymmetry, in comparison with the H-α schema (24). In Chapter 4, I will put forward various arguments against Uriagereka (1999) and Sheehan's (2009, 2010, in press) attempts at accounting for the CED in terms of the LCA, and point out that the H-α schema-based account of the CED effect is sustainable only if the theory of syntax is dissociated from the LCA. I will draw the even stronger conclusion that the LCA is a departure from the SMT, having no place in the theory of bare phrase structure. I will also discuss Kayne's (2010) p-merge-based theory of antisymmetry in this chapter. While I cannot provide a definite proof that Kayne's p-merge theory is unsustainable, it will be pointed out that this theory inherits several problematic aspects of his earlier LCA-based approach to antisymmetry. The overall discussion in Chapter 4 will then provide a partial answer to the following question:

(27)   *Why are there word order variations in natural languages? Why is there no (strong) antisymmetry, i.e., no universal word order template?*

The answer is simply that linear order is not part of narrowly syntactic computation, and syntax is too simple and minimal to give rise to one and the same solution to every problem of externalization. Indeed, if precedence is an integral part of syntax, as Kayne and others claim, it becomes really mysterious why it does not figure in linguistic computation for most of the time, and also why 'optimal' utilization of such precedence-relations does not lead to a single uniform answer to each and every linearization problem. I will then discuss an alternative approach to the problem of phonological linearization, while still keeping to the hypothesis that syntax assumes no mechanism of labeling/projection.

Chapter 5 will then discuss further consequences of the elimination of labeling/projection from the theory of syntax. I will advance the hypothesis that the interpretive effect of endocentricity (headedness), which is no longer characterizable in terms of labeling by projection, can be reduced to minimal search of the most prominent LI in each constituent.

(28)   ***Minimal head detection (MHD)*** :

        The head of an SO Σ is the most prominent LI within Σ.

The effect of MHD is essentially that for any SO {H, α}, where H is an LI and α an SO, H is the head of {H, α} (cf. the first clause of Chomsky's 2008 labeling algorithm; see also Piattelli-Palmarini et al. 2009:52ff). I will discuss in this chapter how this minimal mechanism of head-detection can provide a principled answer to the *why*-question in (29):

(29)   ***Why does the effect of endocentricity arise so prominently in linguistic computation of SEM and PHON?***

The present thesis will be devoted to making sense of the following answer: endocentricity arises because it is the simplest possible form of compositional interpretation. No extraneous representational tricks such as labeling and projection are available in the minimal theory of bare phrase structure, and thus CI just seeks to determine compositional interpretation of each SO by minimal search. The theory of phase cyclicity and the eradication of projection and feature-percolation pave the way for consistent and efficient satisfaction of this CI-requirement, which amounts to incorporating the derivational constraint of the H-α schema.

Chapter 2

Phasing in Inclusiveness and No-tampering

2.1 Introduction

As we saw in the previous chapter, it is an essential task for the study of FL to uncover the complex interaction of the three factors of language design (1):

(1)   Three factors in the language design (Chomsky 2005 and others):

    (i)   genetic endowment

    (ii) external stimulus/environment

    (iii) physical and mathematical ('computational' in particular, in the case of digital computational systems like language) principles that are not specific to FL

The minimalist program for linguistic theory, the MP for short, is a research program that seeks to find ways to attribute the effect of the constituents previously stipulated in factor

---

*I am indebted to Masanobu Sorida for his generous and enlightening discussion that has helped me to articulate the hypotheses presented in this chapter. The idea that cyclic derivation by phase (particularly the sort hypothesized by Chomsky 2007a, 2008) is a device that strongly contributes to the inclusiveness condition is independently explored in different ways in Sorida (2011, in progress).

(i) to the overarching laws of nature (iii), whose explanatory force may well go beyond the confinement of FL, crosscutting the mental and physical aspects of the world. The less we can show UG contains, the closer we get to the goal of biological adequacy. Under the assumption that the third factor subsumes the principle of computational efficiency, the MP can be practically pursued by formulating the strong minimalist thesis (SMT) as an empirical conjecture and heuristic working hypothesis:

(2)   *The Strong Minimalist Thesis (SMT)* :

        FL is an optimal solution to conditions imposed by CI and SM.

The SMT guides us to seek the effect of computational efficiency in the design of FL.

        It is hardly necessary to add that our understanding of the third factor is quite limited at the current stage of investigation. Nonetheless, we can reasonably conjecture that some general features of minimal computation enter into the principles of the third factor: less is better than more, redundancy is disfavored, minimal search is more preferred than deeper search, and so on. According to such considerations of computational efficiency, the following principles are hardly controversial:[1]

(3)   *Full Interpretation (FI)* :

        Every constituent of SEM and PHON contributes to interpretation.

(4)   *Inclusiveness Condition (IC)* :

        No new features are introduced in the course of linguistic derivation.

---

[1]The original NTC in Chomsky (2008:138), given in (i), was formulated specifically in terms of Merge:

(i)   Merge of α and β leaves the two SOs unchanged (cf. Chomsky 2008).

Indeed, if we restrict our attention to the simplest formulation of Merge, namely an elementary binary set-formation, it is just an operation that does nothing more than combining SOs, and therefore it cannot modify the elements that constitute the input to this operation. The NTC as formulated in (5) is thus trivially satisfied. Here I am reformulating the NTC as a more general ban on tampering with elements introduced by syntax.

(5)   *No-Tampering Condition (NTC)* :

No elements introduced by syntax are deleted or modified in the course of linguistic derivation.

FL in its essence is a system of computation that generates infinite pairs of SEM and PHON from the Lexicon. So long as the SMT requires that the computation is efficient, then, it is naturally expected that SEMs and PHONs generated by syntax should not contain any redundant or extraneous elements that receive no interpretation. This empirical conjecture corresponds to what Chomsky calls the principle of Full Interpretation (FI) (cf. Chomsky 1995:194). It is also expected that the number of derivational steps in constructing such fully interpretable SEMs and PHONs is minimized per derivation. Specifically, syntactic computation should be 'inclusive', and no extraneous entities that are absent in the input from the Lexicon should be inserted in the course of derivation, respecting the IC (4) (cf. Chomsky 1995:ch.4; Chomsky 2000:113, (30c)). Further, since no redundancy in constructing SEM and PHON is expected, no elements that are once introduced in a given derivation should be deleted or replaced with something else in syntactic computation. The strongest formulation of this constraint is the NTC, a term adapted from Chomsky (2008). In a nutshell, the IC bans *insertion* of new features, the NTC *deletion* and *modification* of already introduced elements, and FI uninterpreted elements at SEM and PHON. While it is hardly controversial that these constraints would naturally comply with the SMT, how much the actual computation of FL satisfies these considerations is of course an open empirical question. The hope for the MP is to show that deviations from them are kept minimal.

It is not widely acknowledged that the questions of the IC and the NTC are rooted in much earlier concerns of descriptive and explanatory adequacy. It has been an important research guideline in the earlier transformational generative grammar that the generative power of grammars should be severely restricted in order to approach

descriptive and explanatory adequacy (see in particular Chomsky 1965 and also Lasnik 1990). Specifically, once it is allowed to insert or delete elements in the course of derivation, the generative capacity may become overpowerful. To present the concern with an extreme example, a grammar with unrestricted insertion will fail to exclude the possibility that it generates *Bill thinks that John loves Mary* or *John never loves Mary* from the same initial input for *John loves Mary*, and a grammar with unrestricted deletion will fail to exclude the converse transformations. This is clearly an undesirable result, so everybody agrees that rules of deletion and insertion should be restricted rather severely in their applicability. The question is to what extent. Various proposals have been made in the literature to address this question, pursuing a more restrictive theory of transformational grammar. As for deletion, the so-called 'recoverability condition on deletion' is one of the major proposals to restrict its generative power: see Chomsky (1965:Ch.3-4) and Peters and Ritchie (1973) among others. As for the restriction on insertion, see, e.g., Chomsky (1970, 1973, 1986a, 1995), Emonds (1970, 1976), Stowell (1981), Lasnik (1990) among many others. The question of the extent to which the restrictive theory of syntax satisfies the IC and the NTC can be regarded as a modern restatement of the same old concern, rooted in the earlier considerations of descriptive and explanatory adequacy and now further regarded as grounded in computational efficiency, closely tied to the overarching desideratum of FI (3).

It is instructive to recall that violations of the IC and the NTC were practically inevitable in the earlier conception of transformational grammar. It was assumed therein that the skeletal structure of a sentence was generated by the system of phrase structure rules (PSRs), and then mapped to the transformational structure (s(urface)-structure, LF, etc.) by means of *transformational rules* (*transformations*). Transformations are mappings from a phrase-marker to another phrase-marker, and they prototypically involve massive insertion and tampering. For example, it was assumed in the earlier transformational

grammar of English that the transformation in (6), called WH-movement, maps, e.g., the phrase-marker in (7a) to the one in (7b). In the course of this operation, the wh-phrase is inserted at COMP (cf. Bresnan 1970, 1972), and the original occurrence of a wh-phrase is deleted and replaced with a newly inserted trace, probably accompanied by another insertion of identical indices to the wh-phrase and the trace (cf. Chomsky 1973, 1975 *et seq.*).

(6)   WH-movement:

  structural description (SD): X — COMP — Y — wh-phrase — Z
  structural change (SC):   X — wh-phrase$_i$ — Y — t_i — Z

(7)   (Guess) [*what* the boy will eat]
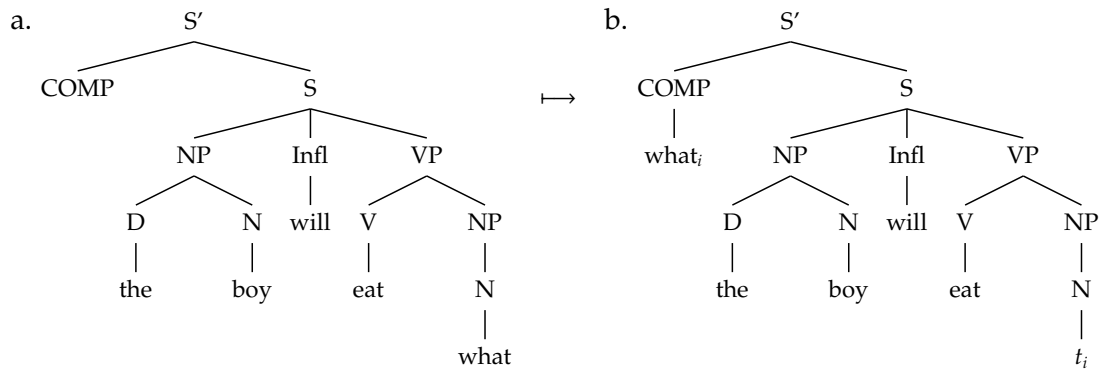
As representatively shown by this single example, transformations are by their very nature rules of insertion and tampering, and every application of a transformation violates in some way or the other the IC and the NTC. Correspondingly, it was practically impossible to even formulate the minimalist desiderata of the IC and the NTC in the framework of transformational grammar without being self-destructive. This partially explains why the formulations of the IC and the NTC had to await four to five decades after Chomsky (1955/1975, 1957) put forward transformational generative grammar, though the concerns have arguably persisted throughout the history of generative grammar. Another point to mention in this regard is the historical fact that the advent of bare phrase structure

(Chomsky 1994, 1995) constituted a crucial basis for approaching the IC and the NTC from a minimalist perspective, as we will see below.

The present chapter will attempt at approaching the problems posed by the IC (4), the NTC (5) and FI (3) from the perspective of the modern minimalist framework. Specifically, I will put forward the hypothesis that these derivatives of computational efficiency constitute a strong rationale for the system of derivation by phase (2000 *et seq.*; see also Uriagereka 1999). The theory of phases presented in this chapter will provide the grounds for the proposals put forward in the following chapters.

2.2   Merge, Inclusiveness, and the NTC

So long as syntax is a computational system that generates structured representations and maps them to SEM and PHON, it would seem to require (i) its own workspace in which it can assemble elements into structures before sending them to CI and SM, and also (ii) some operation that generates such linguistic structures. Specifically for (ii), syntax must minimally assume an operation that takes two linguistic objects already formed as input, and constructs from them a new linguistic object: we call the operation *Merge* (Chomsky 1995 *et seq.*).

(8)   Merge$(\alpha, \beta) = \gamma$.

In principle, we can imagine many other candidates for the formulation of Merge, in particular the characterization of γ in (8). However, Chomsky (2010a) argues that Merge should meet the principle of minimal computation and that, therefore, α and β are not to be modified in the course of this very operation, so that γ equals just a set of these two linguistic objects. Under this conception, the possible mode of Merge is restricted to its simplest form, namely set-formation: thus, Merge maps α and β to the set $\{\alpha, \beta\}$ (9a), without modifying the internal constitution of α and β. Just as an informal but visually

convenient notation, we will also sometimes represent the output of Merge in terms of tree-diagrams as in (9b), but speaking repeatedly, it should be understood throughout the present thesis that trees such as the one in (9b) make no claims about left-to-right ordering of linguistic constituents.

(9)   Merge(α, β)   =   a. $\{\alpha, \beta\}$

                               b.   (order irrelevant)

$$\alpha \quad\quad \beta$$

It is a natural hypothesis that some of the elementary properties of Merge will capture the basic fact about human language that FL can generate an infinite array of linguistic expressions. Any human brain is a finite entity with limited spatial extensions, and thus any mental capacity that is buried in this finite object, such as FL, cannot in itself be infinite. Thus, FL *per se* should be a finite system of computation that is nevertheless designed in such a way that it can support a *recursive*, i.e. infinitely repeatable, use (cf. Wilhelm von Humboldt's description of language as a system which "makes infinite use of finite means;" see Chomsky 1966/2002/2009). Then, arguably the simplest hypothesis is that Merge is this recursively applicable operation. That is, the linguistic object $\{\alpha, \beta\}$, the output of Merge, can constitute further input to recursive Merge, forming a yet larger set-theoretic object, say $\{\delta, \{\alpha, \beta\}\}$.

(10)   Merge($\delta, \{\alpha, \beta\}$)   =   a. $\{\delta, \{\alpha, \beta\}\}$

                               b.   (order irrelevant)

$$\delta \quad\quad \alpha \quad\quad \beta$$

Optimally, the recursive generative procedure of FL is entirely reduced to Merge. As mentioned above, the theory of FL whereby linguistic expressions are generated solely by recursive application of Merge is called the theory of bare phrase structure (Chomsky 1994, 1995 *et seq.*).

An important consequence of bare phrase structure is pointed out by Chomsky (2004, 2007a, 2008), which is that the simplest formulation of Merge in (9) provides, without further stipulation, both structure-building and movement for free, which is an improvement over how these operations were once captured by independent mechanisms of phrase structure rules (PSRs) and transformations, respectively. Suppose that two linguistic constituents α and β are subjected to Merge, forming $\{\alpha, \beta\}$. By logic, it must be either that α and β are two independent objects, as in (11), or that one of them (say α) is part of of the other (β), as in (12) ($\longmapsto$: 'is mapped to').

(11) $\alpha, \beta \longmapsto \{\alpha, \beta\}$
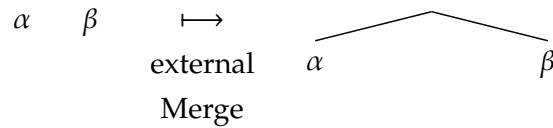
(12) $[_\beta \ldots \alpha \ldots] \longmapsto \{\alpha, [_\beta \ldots \alpha \ldots]\}$
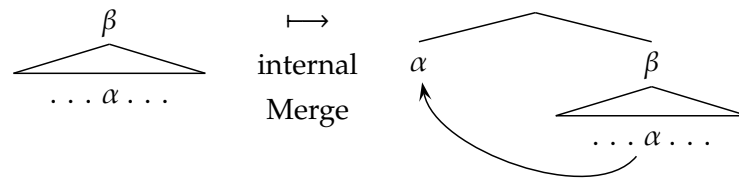
The first case in (11) is referred to as *external Merge* (EM), and the second *internal Merge* (IM). On the one hand, EM yields a rather familiar mode of compositional structuring. For example, EM can combine *the* and *apple*, forming {the, apple}, and this set can be further merged externally with *eat*, forming {eat, {the, apple}}, and so on. On the other hand, unless stipulated otherwise, Merge should be able to take two linguistic objects one of which is part of the other, yielding IM. For example, if there is a sentential structure of the form (13a), nothing should prevent Merge from applying to (13a) and *what* contained in (13a),

yielding IM of *what*.[2]

(13)   a. $\longmapsto$ b.

```
        C                          Merge              what
          we                                              C
            Infl                                             we
              see    what                                       Infl
                                                                  see    what
```
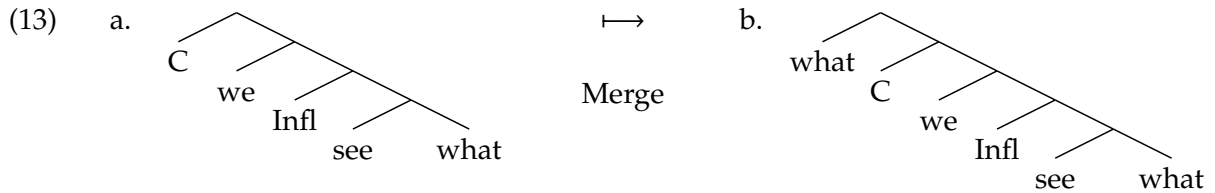
As shown, application of IM yields two *copies* of α in $\{\alpha, \beta\}$, one external to β, one within β. Assuming that the lower copy is to be deleted at the phonological component of linguistic computation, IM yields familiar displacement of linguistic constituents, a ubiquitous phenomenon observed in one way or another in every natural language. Moreover, given the simplest formation of Merge in (9), IM derives what has been called the 'copy theory of movement' (Chomsky and Lasnik 1993, Chomsky 1993), whose empirical force has been verified by various facts about the so-called 'reconstruction' effect (see Chomsky 1977, Freidin and Lasnik 1981 and much subsequent work). This way, the single hypothesis of Merge elegantly unifies the accounts of two very different aspects of syntactic computation, structure-building and movement, which were treated very differently by PSRs and transformations in earlier linguistic theories. This already provides a very strong motivation for the minimalist pursuit of the bare phrase structure framework. Therefore, without adding any stipulation to the contrary, we will assume that both kinds of Merge, EM and IM, are available for the syntax of FL.

It should be noted that for any sort of Merge-based computation to work at all, there must exist some finite set of 'atomic elements' that can serve as input to, but are not themselves constructed by, Merge-based compositional structuring. We may call these 'atoms' of syntactic computation *lexical items* (LIs), and we may further refer to the collection of LIs stored in the human mind/brain as the *Lexicon*, to adopt a familiar terminology. We really don't know the internal composition of these LIs, but to say the very least, for each

---

[2]The syntax of clausal structures will be elaborated in due course, but the simplified structure in (13) is sufficient for the present exposition.

LI X, X should be by definition associated with some property that allows it to be subjected to Merge, a feature that is called the *edge-feature* (EF).[3] X may further contain, in addition to the EF, some formal features that differentiate one LI from the other in their functions and distributions in syntactic derivation, as well as some other intrinsic features that can contribute to interpretation either at CI or SM or both (semantic and phonetic features, respectively).

Then, let us define the notion of *syntactic object* (SO) as follows:[4]

(14) α is a *syntactic object* (SO) iff

      a. α is an LI, or

      b. α is a set $\{\beta, \gamma\}$, where β and γ are SOs.

According to this definition, SOs are linguistic representations that are either LIs or set-theoretic objects constructed from a finite number of LIs by means of recursive Merge. We can further define a set of primitive relations in the theory of SOs. First, applying to two SOs α and β, Merge establishes the relation *constitute* for $\{\alpha, \beta\}$ as defined in (15).

(15) α and β *constitute* γ (or conversely, γ *consists of* α and β) iff γ is the output of Merge(α, β).

Constitute (or *consists-of*) is an irreflexive, asymmetric, nontransitive, and nonassociative relation.[5] Moreover, the *term-of* relation can be defined as follows (cf. Chomsky 1995:247,

---

[3]See Chomsky (2007a, 2008), Fukui (2008, forthcoming), Boeckx (2010b, 2009) for relevant discussion. We will explore the properties of EFs in Chapter 3.

[4]Guimarães (2000), Kayne (2009) and Fukui (forthcoming) argue that Merge should also be able to create singleton-sets of the form $\{\alpha\}$. If there really exist such SOs, then we need to modify the definition of SOs in (14). I will refrain from this issue in what follows, for the lack of strong evidence in favor of such singleton-set structures in human language.

[5]The *Constitutes* relation is reminiscent of the "is-a" relation in Chomsky (1955/1975, 1957) (see also Lasnik et al. 2000). Since "is-a" is defined as a relation between a linear sequence of terminal symbols and a nonterminal symbol in phrase-markers, it is unformulable in bare phrase structure where no notion of linear order is available in syntax. Nor are nonterminal symbols or labels, as I will argue in Chapter 3. Incidentally, being nontransitive, the constitute-relation is distinct from the traditional notion of 'dominance'.

et seq.), assuming composition of relations:

(16)　　For any SO K,

　　　　a.　K is a term of K;

　　　　b.　If K is a term of L and K = {α, β}, then α and β are terms of L.

Term-of is a reflexive and transitive relation as in (16a). I further assume that the *con-stitute* and *term-of* relations are established derivationally and step by step, by means of recursive application of Merge. That is to say, I adopt Epstein et al. (1998) and Epstein's (1999) derivationalist hypothesis that syntactic relations are not unexplained (stipulated) definitions defined on already built-up syntactic representations, but rather are properties of independently motivated elementary operations themselves.

　　　　Note that so long as SOs respect the definition in (14), they will satisfy the formal condition of *Inclusiveness* (cf. Chomsky 1995:ch.4; Chomsky 2000:113, (30c)), which holds that no new features are added to linguistic representations in the course of syntactic derivation:

(4)　　***Inclusiveness Condition (IC)*** :

　　　　No new features are introduced in the course of linguistic derivation.

Further, the theory of SOs as defined in (14) also satisfies the *No-Tampering Condition* (NTC), as defined in (5).

(5)　　***No-Tampering Condition (NTC)*** :

　　　　No elements introduced by syntax are deleted or modified in the course of linguistic derivation.

Insofar as the generative power of syntax is properly restricted by these strict conditions, its computation essentially reduces to just formal rearrangement of the elements and features

provided by the Lexicon. These conditions strongly respect efficient computation, minimizing tampering with SOs. The question is, then, how much of the actual computation of FL can be shown to satisfy these conditions. Put in other words, it is an open empirical question how closely the classes of SEMs and PHONs correspond to the class of SOs as defined in (14), satisfying the IC and the NTC. It may turn out that actual SEMs and PHONs subsume more complicated structures than pure SOs, containing additional features that are absent in the initial array of LIs, and/or undergoing deletion of some elements that are once introduced to the derivation. If this is what we find as a matter of fact, the theory of syntax must be supplemented with some additional devices other than Merge to insert the newly added features and/or delete the introduced elements. Hypotheses about these newly added formal devices, be they insertion or deletion operations, would bear a heavy burden of empirical justification, and would also pose a serious question as to how such devices could emerge in the course of evolution/development of human language.

In reality, departure from the IC and the NTC seem plainly unquestionable as for PHON, and most of the contemporary theories of formal semantics hold that SEM involves massive violations of the IC and the NTC, too, inevitably posing questions of biological adequacy. Note that complication of PHON may be justifiable to a certain extent, given that externalization is in itself a very complex phenomenon involving multiple factors for articulation, perception, and communication. Most of these factors are not operative in the mapping to CI, and cannot be employed for the explanation of IC and the NTC violations in SEM, should there really be any.[6]

---

[6]It is possible to further stipulate that CI is configured in such a way that it necessitates certain violations of the IC and the NTC. An interesting question in this line of approach is whether such features of CI are shared with other animals or evolved only in the human species (see Hurford (2007) for some relevant discussion).

2.3   Transfer: Compounding Interpret and Spell-Out

Another crucial aspect of linguistic computation is that the hierarchically orga-nized structures generated by syntax are fed to interpretation by the 'thought' system (called the Conceptual-Intentional system, CI) and to externalization (via sounds or signs) by the sensorimotor system (SM). Then, at some point in syntactic derivation, certain SOs already formed in the workspace are handed to the interfaces with CI and SM. We call the operations that hand SOs to CI and SM *Interpret* and *Spell-Out*, respectively (cf. Chomsky 1995, Lasnik et al. 2005).

(17)   *Interpret* maps an SO to SEM.

(18)   *Spell-Out* maps an SO to PHON.

Keeping to the SMT, we refrain from adding any further stipulations to the formulation of these two interfacing operations. Then, in principle, it may be that Interpret and Spell-Out apply only once per derivation, as hypothesized by Chomsky (1995), or more than once, as proposed on different grounds by Uriagereka (1999) and Chomsky (2000) and much subsequent work. Moreover, (17) and (18) do not specify whether the timing of Interpret and Spell-Out always coincides (Chomsky 2004 and others) or is differentiable (see, e.g., Chomsky 1995, Nissenbaum 2000, Marušič 2009), or whether application of these interfacing operations render the relevant SOs inaccessible for further computation (Chomsky 2000 *et seq.*) or not (see, e.g., Fox and Pesetsky 2005).

However, it is naturally expected that the third factor principles of computational efficiency severely restrict the possible modes of application of Interpret and Spell-Out. Chomsky (2004:107) argues that, "in the best case," presumably determined by the third factor, Interpret and Spell-Out always apply at the same point in the derivation, and further, the domain subjected to them will be literally 'forgotten', i.e., eliminated from the derivational workspace. Following Chomsky, we will call this operation *Transfer*:

(19)   *Transfer* (preliminary version):

Transfer subjects an SO Σ to Interpret and Spell-Out, and eliminates Σ from the derivational workspace.

Indeed, there are several grounds for the belief that Transfer meets criteria of efficient computation, which we will discuss in what follows.

One of the most important contributions of Transfer to computational efficiency is its reduction of active workspace. Specifically, I contend along with Chomsky (2000, 2004, 2008) that compounding of Interpret and Spell-Out readily allows syntax to terminate access to the SO subjected to Interpret and Spell-Out for later computation. In order to see this point, let us consider what happens if Interpret and Spell-Out can apply independently of each other (see, e.g., Nissenbaum 2000, Marušič 2009 for proposals to this effect). In this system, it would be necessary for syntax to keep track of which operation has applied to which SO. This would require some mechanism of marking such SOs as Interpreted and/or Spelled-Out. The mechanism would take the form of either (i) assigning features like [±Interpreted] and [±Spelled-Out] or the equivalent to SOs in the derivational workspace, or (ii) deleting part of semantic and phonological features of the relevant structures upon application of Interpret and Spell-Out, tampering with the featural constitution of LIs and SOs.[7] Crucially, note that (i) would violate the IC (4), whereas (ii) would violate the NTC (5), thus departing from the SMT. Moreover, another costly aspect of such independence of Interpret and Spell-Out is that it would also be inevitable to store in the active workspace those SOs which are 'half-processed' by Interpret or Spell-Out for later application of the other. Consequently, syntax would always need to do some inspection of the workspace and/or the derivational history to see which part of the SOs it is 'done' with and it can forget about for later computation. On the other hand, compounding of Interpret and Spell-Out avoids such potential computational overload: if Interpret and Spell-Out always apply to

---

[7](ii) is exactly what Chomsky (1995, 2000) proposes, introducing the notion of 'modified LIs' (MLIs). It is argued in Chomsky (2001) and later work that this notion can and should be dispensed with.

the same SO together at the same time, then, Transfer is arguably the most efficient way to terminate further access to previously Interpreted/Spelled-Out SOs. This consideration already gives us a strong motivation in favor of the hypotheses in (20)–(21):

(20)   Interpret and Spell-Out apply simultaneously in the derivation.

(21)   Further access by syntax to SOs that have been subjected to both Interpret and Spell-Out is terminated.

The composite of (20) and (21) amounts to a deduction of the hypothesis of Transfer in (19). Thus, Transfer arises as the mode of Interpret/Spell-Out in the 'best case' scenario as determined by minimal computation, as I will assume.

That said, it remains to be determined how syntax configures the mode of applications of Transfer. Here, different economy considerations favor different answers. For example, it is not unreasonable to suppose that one aspect of efficient computation favors minimization of the memory load for the active syntactic workspace ("Less memory load is better than more"). That is, all things being equal, it is presumably more efficient to keep the active domain accessible to computation as small as possible. To meet this end, it may be that Transfer should apply as many times as there are applications of syntactic operations (including Merge), exploiting the effect of cyclic reduction of active syntactic domains to its limit, which yields Epstein and Seely's theory of Transfer (Epstein and Seely 2002, 2005). However, this Epstein-Seely approach inevitably clashes with the other economy consideration that favors minimization of derivational steps, which is sometimes referred to as the principle of economy of derivation (Chomsky 1995, Fukui 1996). Again, all things being equal, it is more efficient to keep the number of rule-applications in a given derivation as small as possible: "Less steps are better than more." According to this economy criterion, the number of Transfer-applications should be kept as small as possible. To the limit, it reduces to just one—the single Transfer model of Groat and O'Neil's (1996) sort

explores exactly this scenario, but again with the cost of potential blow-up of memory load. Here, we see a clear conflict between memory minimization and step minimization, both of which are presumably just different aspects of the general principle of computational efficiency. However, we will see in what follows that there are other considerations that enter into the determination of the point of Transfer.

2.4   Copy-identification and Derivational Simultaneity

2.4.1   The Bifurcation of External and Internal Merge

We saw in §2.2 that the principle of minimal computation restricts the form of Merge to simple set-formation, (9) repeated here.

(9) $\text{Merge}(\alpha, \beta)$ =   a. $\{\alpha, \beta\}$

b.   (order irrelevant)
α β

We saw that the simplest formulation of Merge in (9) provides, without further stipulation, both EM and IM for free (Chomsky 2004, 2007a, 2008). If two SOs α and β are merged, constituting $\{\alpha, \beta\}$, it may be either that α and β are two independent SOs, as in (22), or that one of them (say α) is part of of the other (β), as in (23).

(22) $\alpha, \beta \longmapsto \{\alpha, \beta\}$

α β $\longmapsto$

EM α β

(23) $[_{\beta} \ldots \alpha \ldots] \longmapsto \{\alpha, [_{\beta} \ldots \alpha \ldots]\}$
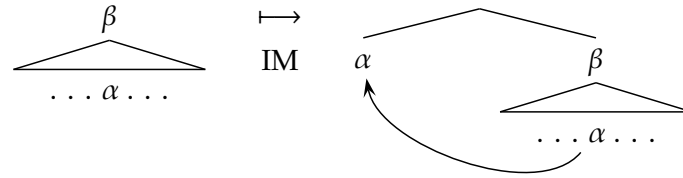
It is quite natural to expect that the two modes of Merge, EM and IM, are exploited in different manners by CI and SM. Chomsky (2008:140) notes,

> If the means of language are fully exploited by the interface systems, in accord with a reasonable interpretation of SMT, then we would expect the two types of Merge to have different effects at the interfaces. At the phonetic interface, they obviously do; IM yields the ubiquitous displacement phenomenon. At the semantic interface, the two types of Merge correlate well with the duality of semantics that has been studied within generative grammar for almost forty years, at first in terms of "deep and surface structure interpretation" (and of course with much earlier roots). To a large extent, EM yields generalized argument structure (θ-roles, the "cartographic" hierarchies, and similar properties); and IM yields discourse-related properties such as old information and specificity, along with scopal effects. The correlation is reasonably close, and perhaps would be found to be perfect if we understood enough—an important research topic.

Indeed, such empirical generalizations as the duality of semantics, the ubiquity of dislocation and the predicted availability of interpretive reconstruction effects seem to be quite well-established. Hence, we will assume that the freely available dichotomy of EM and IM is indeed exploited by CI and SM.

That said, at this point we would like to point out that the inherent computational cost of IM, or, more precisely, that of copy identification at CI and SM, is not trivial. It is true that once syntax has Merge as set-formation (9), IM comes for free, and its application should be "as free as EM" (Chomsky 2008:140). However, at the point of Transfer where a certain SO is handed over to Interpret and Spell-Out, there must be a way to distinguish copies created by IM from independently externally merged items with identical syntactic constitution, since otherwise CI and SM cannot exploit the natural bifurcation of EM and IM for interpretive purposes. To see the problem clearly, let us suppose that an SO α is externally merged with another SO β, which contains a term whose syntactic constitution

is indistinguishable from α. The situation can be schematized as in (24):

(24) $\alpha, [_\beta \ldots \alpha \ldots] \longmapsto \{\alpha, [_\beta \ldots \alpha \ldots]\}$
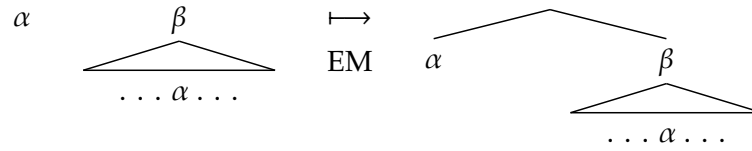
Note that the result of external merger of α and β in (24) is *representationally* identical with that of internal merger of α and β in (23). As far as the syntactic constitution is concerned, then, the representations in (23) and (24) are indistinguishable. The problem that FL faces is hence to store the distinction between IM in (23) and EM in (24) by the time $\{\alpha, \beta\}$ are subjected to Transfer; otherwise CI and SM cannot exploit the two modes of Merge for interpretation.

It has been customary to trivialize this computational problem by introducing such descriptive devices as 'referential indices'[8] in the course of linguistic computation. These hypothetical constructs are introduced to linguistic theory not so much as an expression of theoretical insights into the nature of UG as artificial technology to meet or circumvent descriptive demands. However, under the assumption that FL meets the IC (4), reference to such descriptive apparatus is strictly barred in the MP. Hence, we need an alternative approach to the differentiation of IM from EM that makes no reference to resultant representations.

The problem in question is derivational in nature. The only real difference between IM and EM of α to β is that, *at the point of application of Merge(α, β), α is selected from the terms of β in the former while not in the latter.* This distinction should be made available at Transfer, which amounts to reconstruction of the selectional procedure in the derivational

---

[8]Or much more complicated structures referred to as 'assignment functions', as dominantly presumed in the literature of modern formal semantics. See, e.g., Heim and Kratzer (1998).

history. It is a form of *backtracking*, which induces a large-scale computational load. To respect computational efficiency, backtracking is disfavored, and hence to be avoided as much as possible. Specifically for the purpose of reconstructing IM, then, the timing difference (in the derivational sense) between its application and Transfer of the relevant SO must be minimized. Optimally, the time difference reduces to zero, as is expected from the perspective of computational efficiency. Arguably, then, the best way for syntax to meet this end is to subject a given derivation to Transfer *simultaneously with* IM. Given these considerations, we claim that, just like the case of Interpret and Spell-Out discussed in the previous section, the principle of minimal computation derives derivational simultaneity of IM and Transfer (see Chomsky 2007a, 2008; see also Hiraiwa 2005).

(25) IM applies simultaneously with Transfer.

Note that externally merged SOs need not be identified with any other SOs. Unlike IM, then, EM does not require backtracking. EM is hence not required to apply simultaneously with Transfer, contrasting sharply with IM.

The discussion here partially corroborates the hypothesis established in §2.3 that derivational simultaneity of Interpret and Spell-Out contributes to computational efficiency: it is presumably the case that both CI and SM exploit the bifurcation of EM and IM for different interpretive purposes, and thus the consideration of backtracking minimization applies to both Interpret and Spell-Out. Hence, both Interpret and Spell-Out are required to apply simultaneously with IM.

We are not primarily claiming here that we can technically circumvent the problem of inclusiveness by stipulating that IM applies at the same time with Transfer. Rather, our stronger claim is that it is the very satisfaction of the IC (4) that leads syntax to adopt derivational simultaneity of IM and Transfer as a necessary result of minimal computation.

It remains to be determined what sort of information is subjected to Transfer when it is in concert with IM. We will come back to this problem in §2.5.

2.4.2 Derivational Simultaneity of Value/Agree and Transfer

Since Merge is just a set-formation operation that generates sets of SOs, it cannot rearrange elements internal to already created SOs, due to the NTC (5). Merge is arguably the simplest possible recursive operation for symbol manipulation, so the null hypothesis is that all formal arrangement in syntactic computation is achieved by Merge. However, empirical evidence suggests that something that cannot be naturally expressed by Merge is also at stake in human language: linguistic expressions exhibit certain *covariation* of morphophonolgical or semantic features among multiple elements that possibly range over non-sister relations. For example, in English *there*-expletive constructions as in (26) the main verb exhibits apparently long-distance agreement in number with an associate NP. Similarly, the gender marking on determiners and adjectives in Romance nominals covaries with the inherent gender specification of a head noun, yielding nominal-internal morphological concord as shown in (27) from Spanish. To take another example from Icelandic (28), the past participle agrees in Case, number and gender with the accusative NP in the so-called ECM/Raising-to-Object construction.

(26) a. There *seems* to be likely to be *a boy* in the garden.

　　　b. There *seem* to be likely to be *boys* in the garden.

(27) *Spanish*: (Carstens 2000:322)

　　　a. *la*　　casa　blanc*a*
　　　　　the-FEM house white-FEM
　　　　　'the white house'

　　　b. *el*　　　coche blanc*o*
　　　　　the-MASC car　　white-MASC
　　　　　'the white car'

　　　c. *las*　　　casas　　blanc*as*
　　　　　the-FEM.PL house.PL white-FEM.PL
　　　　　'the white houses'

    d.   *los*   coches blanc*os*
            the-MASC.PL car.PL white-MASC.PL
            'the white cars'

(28)   *Icelandic*: (Frampton and Gutmann 2000, Chomsky 2001, 2004, Hiraiwa 2005)

     Ólafur   hefur   lílega   tali   *einhvern*   hafa ver
     Olaf(NOM) has(3.SG) probably believed someone(ACC.M.SG) have been
     drepinn.
     killed(ACC.M.SG)
     'Olaf has probably believed someone to have been killed.'

Such long-distance featural covariation cannot be readily captured by Merge, since its effect involves (i) elements that go beyond local sister-relations, and arguably (ii) features of LIs that are not themselves susceptible to Merge (see §3.5 for the atomicity of LIs). So, syntax must provide a way (or several ways) to code such (potentially long-distance) featural covariation among multiple LIs.

       Chomsky (2000) and much subsequent work suggest that *Agree* is responsible for capturing (at least some of, optimally all of) such covariation of features among LIs. Agree is an operation that relates two (or more) LIs via a derivational search operation relative to a given feature F. Some LI P with an *unvalued* feature F (henceforth [uF]) acts as a *probe*, and it seeks in a certain search domain a matching F with a lexically specified value (henceforth [vF]) of a *goal* LI G. Once such a probe-goal relation is identified, Agree holds, and the value of F of G is copied onto the F of P.[9] (30) schematically shows how the mechanism works.

---

[9] According to Chomsky (2001), the technical working of Agree is somewhat more complicated. In order to establish Agree(P, G) with respect to some feature F, G must contain some *activating feature f* which renders G eligible as a goal for F, e.g., Case for Agree with respect to φ-features agreement (Chomsky 2000, 2001). If G contains an activating feature for Agree with respect to F, it can successfully ensure the valuation of F on P by G at Transfer. If G doesn't, then Agree between P and G violates the Inactivity Condition on Agree (as stated in (i)), leading to the cancellation (or crash) of the derivation (Chomsky 2001, 2004, 2008).

(i)   Inactivity Condition on Agree:
       If G lacks an activating feature *f* for Agree with respect to some feature F, then Agree(P, G) cannot lead to the valuation of P's F by G's F.

In what follows I will abstract away from this possible complication of the Agree-mechanism. Nothing in what follows will hinge on this decision, as far as I can tell.

(29)   **Agree:**

    Probe:

        An LI P with an unvalued feature F [uF] probes into its domain for the closest

        matching valued feature [vF] of a goal LI G.

    Value:

        The value of [vF] is copied onto [uF].

(30)
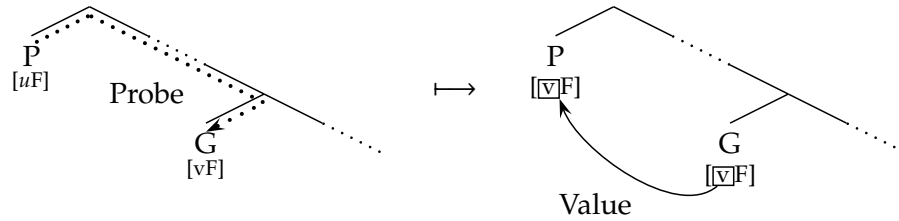
According to Chomsky (2000 *et seq.*), the domain of P is defined as P's sister. Note that (29) further holds that Agree must be established between P and its 'closest' goal. This sort of minimality requirement is naturally expected for any search operation such as Agree, for the principle of computational efficiency requires that the search step that is required to relate P and G be minimized.

      Prima facie, Agree is a costly operation that requires derivational search for a matching feature, but the assumption is that unvalued features receive no interpretation at the interfaces, and that the failure to value them by means of Agree violates the principle of FI (3), repeated here.

(3)   *Full Interpretation (FI)* :

    Every constituent of SEM and PHON contributes to interpretation.

Thus, if syntax introduces any LI with [uF] to the derivational workspace, [uF] necessitates some application of Agree.

Crucially, note further that Agree involves copying just like IM: the only difference between Agree and IM is that IM creates copies of SOs,[10] whereas Agree creates copies of feature-values internal to LIs (recall that LIs are by definition 'atoms' for Merge). I now argue that due to this property, Agree is required to apply in simultaneity with Transfer, much like IM. We saw in §2.4.1 that IM necessarily accompanies simultaneous application of Transfer, so long as CI and SM exploits the 'costless' bifurcation of EM and IM while still keeping to minimal computation. Specifically, IM results in SOs that are representationally indistinguishable from comparable SOs created by EM, and syntax applies IM and Transfer simultaneously in order to deliver the EM/IM distinction to CI and SM. The problem that necessitates simultaneity of IM and Transfer is representational indistinguishability of copied material from that which is externally merged. It should be pointed out in this connection that more or less the same representational indistinguishability problem arises in cases of Agree as well. Given FI (3), [*u*F]'s must be valued via Agree before they reach the interfaces. However, once valued by Agree, they become indistinguishable from inherently interpretable (i.e., lexically valued) features, as far as the resultant representation is concerned. There is evidence that the interfaces need to know the identity of feature value copies created by Agree: thus, for example, even if one or more copies of φ-features are created by Agree (probably applying to multiple goals; cf. Hiraiwa's 2005 theory of Multiple Agree), they are typically interpreted only at certain designated positions determined by CI and SM; valued φ-features at T are to be left uninterpreted at CI, whereas they will surface as morphological agreement at SM. So long as minimal computation avoids backtracking, then, it follows that [*u*F]s must be subjected to Transfer as soon as they are valued. If Value takes place before Transfer, the identity of copied feature-values will be representationally undetectable at the interfaces; if Value takes place after Transfer, then the derivation will crash at both interfaces due to FI. Therefore, it follows that Value must apply in simultaneity with Transfer.

---

[10]Chapter 3 will put forward the hypothesis that IM creates copies only of LIs, not of any phrasal SOs.

(31)   Value and Transfer of [*u*F] must apply simultaneously in the derivation.

This consequence is first effectively pointed out by Epstein and Seely (2002).[11] Note its similarity to the derivational simultaneity of IM and Transfer (25). IM and Agree are both operations that create copies of elements they apply to, and hence they induce the same sort of copy-identification problem at the interfaces. As a result, their applications are synthesized due to the required minimization of backtracking.

2.5   Defining Phases

In §2.3, we reached the conclusion that Interpret and Spell-Out are required to apply simultaneously in the derivation. The compound of Interpret and Spell-Out is called Transfer. Let us refer to SOs to which Transfer applies as *phases*, following the terminology introduced by Chomsky (2004, 2007a, 2008). It remains to be determined under what conditions an SO can constitute a phase, a problem to which we now turn.

2.5.1   The Edge and Interior of Phases

Our initial formulation of Transfer is given in (19), which is repeated here for ease of reference.

(19)   *Transfer* (preliminary version):

Transfer subjects an SO Σ to Interpret and Spell-Out, and eliminates Σ from the derivational workspace.

---

[11]Note that Epstein and Seely (2002) draw from (31) a somewhat different conclusion from ours that Spell-Out and Interpret apply "inside" each application of syntactic operations. However, it should be noted as for their "rule-applications-as-cycles" model that (i) the number of required applications of Interpret and Spell-Out are undesirably magnified according to the number of rule-applications, and also that (ii) Interpret and Spell-Out can no longer apply in a way that cyclically reduces the domain of computation as Transfer defined in (19), provided that these operations are buried inside rules applying to features and hence can affect only relevant features and their values. Given these undesirable consequences, in what follows I will explore the phase theory derived from Transfer as defined in (19).

(19) holds that Transfer applying to Σ will unambiguously terminate further access to Σ for later computation.

If Σ is the final phase that corresponds to the root clause, there will be no further computation beyond Σ, and hence Transfer should be able to interpret/spell out Σ in full, terminating the derivation. However, if Σ is anything but the final phase, Transfer cannot be *required* to terminate access to Σ in its entirety, since this would exclude any continuation of the computation beyond the first phase: see Chomsky (2004:108) for relevant discussion. For example, Transfer of the entire phase would preclude merger of the phase with an LI in the higher phase. Moreover, if Transfer applies to the entire phase, it also terminates further access to even displaced (internally merged) elements, barring interphasal (successive-cyclic) movement. This is clearly an undesirable result in terms of expressive potential, in that it significantly restricts the applicability of IM, which comes for free in bare phrase structure (modulo the simultaneity with Transfer; see (25)). This is also undesirable in terms of computational efficiency, in that it effectively poses a single-Transfer model where the phase must always be the root clause which will not be accessed further. So, for cyclic computation to be meaningful in terms of semantic expressiveness and computational efficiency, Transfer as defined in (19) cannot be forced to apply to an SO in full. Then, each phase must be able to have an 'edge' that is part of the phase but will be left accessible for further computation after Transfer. However, given our hypothesis that Transfer has the effect of reducing the computational load by terminating further access to the phase for later computation, it is naturally expected that such an edge to be left in the derivational workspace should be kept minimal, in order to approach computational efficiency.

In order to embody the relevant edge v.s. non-edge distinction, let us introduce a slight modification to the definition of Transfer. I will specifically introduce the term *interior* to refer to the largest domain of a phase which is subjected to deletion upon interpretation.

(32)   *Transfer* (second version):

Applied to an SO Σ, Transfer subjects the largest possible term Σ' of Σ (the *interior* of Σ) to Interpret and Spell-Out, eliminating it from the derivational workspace.

I contend that at each application of Transfer, the interior of a phase Σ is defined as the largest possible term of Σ whose elimination will not prevent the continuation of computation if there is any (see the definition of 'term-of' in (16)). If Σ is the root clause, then no continuation of computation is necessary. Thus, the interior of Σ can be defined as Σ itself, Transfer eliminating Σ in full, along the line suggested by Chomsky (2004:108) and also investigated in an interesting way by Goto (2010) (note that any SO is a term of itself, according to the definition in (16)). If Σ is not the root clause, then the interior of Σ is the next largest SO Σ' that is properly contained in Σ. The remainder of Σ with Σ' subtracted will be the *edge* of Σ that remains accessible for further computation. For example, if Σ is of the form {X, YP}, where X is an LI and YP is an SO, then YP is the largest possible sub-term of {X, YP} and thus qualifies as its interior. Transfer eliminates YP from the derivational workspace at the completion of the phase, and subjects YP to Interpret and Spell-Out.

(33)   a.   X   YP $\longmapsto$ b.   X
                                Transfer
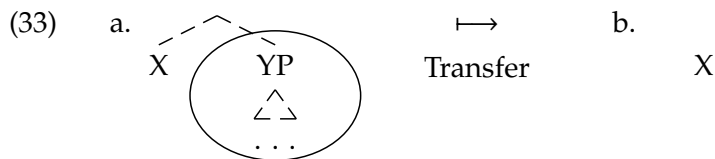
By leaving X for later computation, syntax can continue recursive structure building by Merge. I will assume, along the line with Chomsky (2008:143), that all the structural information related to the phase-interior will be literally 'forgotten', i.e., eliminated from the derivational workspace after Transfer. Specifically, Transfer should be able to ship off to Interpret and Spell-Out not only YP and elements within it but also all the syntactic relations (constitute, term-of, etc.) established with regard to them. For the reason to be explicated in the following chapters, I will specifically assume that the information subjected to Transfer includes the 'constitute'-relation that has been established by Merge

from X and YP to {X, YP} (see (15) for the definition of 'constitute'). Thus, after Transfer, YP as well as the information that YP has constituted {X, YP} with X will be lost, leaving only X for later computation, as I will assume.

In the course of discussion, I will often refer to the LI that is part of the edge of a phase Σ and has once merged with the interior of Σ as the *phase head* of Σ. In (33), X is the phase head of the phase {X, YP}. Incidentally, there are cases where the edge of a phase contains more than a phase-head LI, as I will argue in what follows. For example, it is standardly assumed that the external argument is merged to the specifier of a phase-head LI v^* (v with transitive argument structures). Thus, if the external merger applies before the v^*-phase-level Transfer, then the edge of this phase contains v^* as well as the external argument. More generally, the edge of a phase may subsume not only a phase-head LI X but also one or more specifier of X, say Z, as schematically shown in (34):

(34)   a. $\longmapsto$ b.
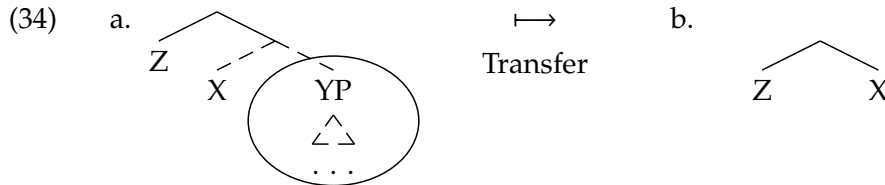
Transfer eliminates the interior YP of X, and thereby {X, YP} is reduced to X just like the case in (33). However, X and Z as well as the 'term-of' relations that X and Z hold to the relevant phase remain accessible for later computation.

Incidentally, it should be noted that the Transfer operation, as formulated in (32), seems to involve a certain form of deletion, applying to the interior SO subjected to Interpret and Spell-Out. Readers might wonder if this would be a violation of the NTC (5). I would like to first point out that elimination from the derivational workspace does not necessarily entail tampering with the relevant interior. Thus, it is perfectly possible to suppose that Transfer leaves the phase unchanged, while terminating further access to the interior SO. Further, I argue that this mode of elimination upon interpretation is of a principled sort, strongly contributing to the desideratum of efficient computation. Note

that the SO subjected to such deletion by interpretation reaches SEM and PHON without any loss, thus this will never increase the generative power of the grammar, meeting the empirical consideration behind the NTC (5). Furthermore, I will claim in Chapter 3 that such deletion by interpretation in the form of Transfer is also empirically corroborated by a number of observations on the locality of syntactic operations. Note that after the elimination of the interior YP from the derivational workspace, no syntactic operation (IM, Agree, etc.) should apply to the elements internal to YP. This is what is often referred to as the *Phase-Impenetrability Condition* (PIC), whose effect will be intensively explored in Chapter 3. I contend that the above-mentioned hypothesis of deletion by interpretation is the best way to characterize the PIC (see Chomsky 2008:143).

2.5.2 Phase Convergence Redux

When Chomsky (2000:107) first introduced the notion of phase, he proposed the following characterization (35):

(35)   Phases are propositional.

According to (35), phases are defined in semantic terms: vP is propositional in that it satisfies the full-fledged argument structure, and CP is also propositional in that it is the minimal construction that includes Tense and event structure and (at the matrix, at least) force. Chomsky (2000) proposed (35) mainly for empirical reasons. Among what he calls 'core functional categories', CP, TP and vP, he wants to distinguish CP and vP as phases, to the exclusion of TP. The former two categories pattern together in contrast to TP in many ways, such as relative phonological and syntactic independence (which can be tested in topicalization, extraposition, pseudoclefting, response fragments, etc.), and so on. For example, the following data from English suggest the clear distinction in syntactic independence between phasal CPs/vPs and non-phasal TPs/VPs.

(36) *English*: Topicalization

    a. (i) I believe that John criticized Mary.

      (ii) [$_{CP}$ that John criticized Mary]$_i$, I believe t_i.

      (iii) *[$_{TP}$ John criticized Mary]$_i$, I believe that t_i.

      (iv) (... and) [$_{vP}$ criticize Mary]$_i$, I believe John (really) did t_i.

    b. (i) I believe John to have criticized Mary.

      (ii) *[$_{VP}$ John [$_{TP}$ to have criticized Mary]]$_i$, I believe t_i.

      (iii) *[$_{TP}$ to have criticized Mary]$_i$, I believe John t_i.

    c. (i) I prefer for John to take care of his mother.

      (ii) [$_{CP}$ for John to take care of his mother]$_i$, I prefer t_i.

      (iii) *[$_{TP}$ John to take care of his mother]$_i$, I prefer for t_i.

(37) *English*: Cleft ((a)-(b) are from Chomsky 2001: fn. 13, attributed to Luigi Rizzi)

    a. It is [$_{CP}$ C PRO to go home (every evening)]$_i$ that John prefers e_i.

    b. *It is [$_{TP}$ t_{John} to$_T$ go home (every evening)]$_i$ that John seems e_i.

    c. *It is [$_{VP}$ Mary to$_T$ go home (every evening)]$_i$ that John believes e_i.

    d. *It is [$_{TP}$ to$_T$ go home (every evening)]$_i$ that John believes [$_{VP}$ Mary e_i].

(38) *English*: Pseudocleft

    a. What John prefers e_i is [$_{CP}$ C PRO to go home (every evening)]$_i$.

    b. *What John$_j$ seems e_i is [$_{TP}$ t_j to$_T$ go home (every evening)]$_i$.

    c. *What John believes e_i is [$_{VP}$ Mary to$_T$ go home (every evening)]$_i$.

    d. *What John believes Mary e_i is [$_{TP}$ to$_T$ go home (every evening)]$_i$.

We will discuss in Chapter 3 how these patterns arise. For now, let us just make the observation that there are certain empirical data that support the differentiation of TP from CP/*v*P. See Samuels (in press) and references cited therein for discussion on phases from a phonological perspective.

Although the characterization of phasehood in terms of propositionality (35) may give rise to the distinction with regard to CP and *v*P versus TP at first glance, it evokes both conceptual and empirical problems. First of all, the crucial notion 'proposition(al)' is left undefined. Given the absence of any 'overt' data in the domain of CI, biolinguistic investigations must proceed to seek understandings of the properties of CI, primarily based on empirical considerations on language. Thus, semantic notions like 'proposition(ality)' need to be clarified in the course of biolinguistic investigations, but are probably not things that can be reasonably assumed to play any explanatory role in such inquiries. This worry becomes more serious in the face of the emerging approach to the 'naturalization of meaning' (Hinzen 2006, 2007, Uriagereka 2002, 2008; see also Chomsky 2007a, 2008, Narita 2009b), whose core claim is that syntax rather strongly configures and determines the properties of the CI- (and SM-) interpretations, not vice versa. Specifically for 'propositionality', Hinzen, for example, argues that such a mysterious notion is just an idiosyncratic stipulation borrowed heavily from 'semantic externalism' in the literature of philosophy, and should be eliminated in favor of a more well-understood theory of syntax. In this approach, the properties of such mysterious thought contents as propositions, if they bear any significance at CI, are to be configured (or 'carved out'[12]) by the role of CP (and *v*P) in syntactic generation. To the extent the reduction is reasonable, we can shift propositions and other semantic entities into the properties of SOs (CP, *v*P, etc.). Importantly, SOs are natural objects "that we can study as such, even though we see them, somewhat miraculously, systematically condition properties of linguistic meaning that we can empirically attest" (Hinzen 2006:235; see also Chomsky 2007a:15). These considerations make it clear that we should not rest on such ill-understood notions like propositionality for the study of syntax. See Hinzen (2006), Uriagereka (2008), Chomsky (2007a), Narita (2009b) for relevant discussion.

---

[12]Uriagereka (1999:275/2002:64), Hinzen (2006:250), Chomsky (2007a:15).

Moreover, even if we tentatively grant that we can make recourse to 'proposition-ality' in the definition of phasehood, there are certain empirical data suggesting that being 'propositional' (35) is not a necessary condition of phasehood, either. For example, there are a certain number of suggestive arguments in favor of the phasehood of nominal phrases (see Matushansky 2003, Svenonius 2004, among many others) and PP (Abels 2001, Gallego 2009, 2010b, among others; see already van Riemsdijk 1978; but see Gallego 2010a), as expected from the common observation that these phrases can move as independent constituents.[13] Indeed, the following chapters will present some additional arguments in favor of the view that PPs as well as noun phrases can define their own phases, even though they are not 'propositional'. Moreover, (35) is not a sufficient condition of phasehood, either, if Chomsky (2000, 2008) is correct in assuming that unaccusative vPs and passive vPs are not phases (at least in the sense that they define independent Transfer domains).[14] *Prima facie*, passive and unaccusative structures seem to be as 'propositional' as transitive ones, especially given that all the relevant θ-roles are discharged within these vPs, too, and thus it remains unclear under (35) why unaccusative and passive vs, as opposed to transitive v (v*, to adopt Chomsky's notation), cannot constitute their own phases. For example, the unaccusative vP in *the plane$_i$ [$_{vP}$ arrived t$_i$]* is presumably as complete as possible, propositionally and θ-theoretically. They don't assume external arguments, but it is not at all clear in most cases how they can be seen 'defective' by any means.

It should be noted that Chomsky (2000:107) briefly touches on the possibility of an alternative characterization of phases (39):

(39)   Phases are convergent.

---

[13]Much recent work on PP suggests that PP should be viewed as equivalent to NP/VP, with p being a potential phase head. See Abels (2001), Hiraiwa (2005), Svenonius (2003, 2010), Cinque and Rizzi (2010) and references therein.

[14]I will refrain from introducing the superfulous terminological distinction between weak phases and strong phases, and keep the term *phase* to unambiguously refer to the domain subjected to Transfer.

According to (39), phases are defined as domains that are convergent at the CI- and SM-interfaces. An SO is said to be *convergent* if it satisfies FI (3) at the interfaces (cf. Chomsky 1995:219-220):

(3)   ***Full Interpretation (FI)*** :

         Every constituent of SEM and PHON contributes to interpretation.

Specifically, we will adopt the conventional assumption that what FI precludes are, among other things, unvalued features (Chomsky 2001, 2007a; see also Richards 2007b, Narita 2009a for further exploration of FI).

(40)   FI resists unvalued features remaining at SEM and PHON.

Under (39) and (40), then, phases can be defined under (39) as SOs that contain no unvalued features.

         (39) was entertained in class lectures at MIT in 1995, but rejected in Chomsky (2000) mainly for empirical reasons. One such consideration comes from, again, the problem of TP:[15, 16] as noted above, the target result therein was to ensure that CPs and vPs are phases, whereas TPs are not. However, it was not at all obvious how anything like (39) would derive the correct cut between CP/vP and TP, especially when Chomsky (2000) was written. It was then widely assumed that the loci of the Nominative and Accusative Case-assignment are located on T and v*, respectively, thus the relevant classification in terms of convergence

---

[15]Chomsky (2000:107) also argues against (39) by pointing out that strictly local determination of phasehood is impossible under (39). However, given that phase cyclicity strictly restricts the workspace for syntactic computation, the detection of convergent domains does not require significant computational load, as I will assume.

[16]Chomsky (2000:107) also discusses the following example.

(i)   *which article$_i$* is there some hope [$_\alpha$ that John will read t_i]

Resting on the assumption that the wh-phrase has an uninterpretable feature [wh], he notes that its presence on t_i should make α nonconvergent and hence a non-phase, which he argues is an undesirable result. As articulated elsewhere (2000:128), he assumes that the wh-phrase has an uninterpretable feature [wh] and an interpretable feature [Q], which matches the uninterpretable probe [Q] of a complementizer in the final stage.

seemed impossible.

However, current recognition of C being the locus of Nominative Case-assignment has changed the picture. Based on the empirically well-established observation that Nominative Case assignment is contingent on the presence of finite C (see Watanabe 1996 among others for earlier attempts to capture the C-T dependency in terms of feature-checking), Chomsky (2008) proposes that C's probe is the 'prime agent' responsible for Nominative Case-assignment, while probing by T for φ-feature agreement is only an ancillary process contingent on C's probing. Thus, it is only after the introduction of C and v^* that the assignment of Nominative and Accusative Case via Agree becomes possible. Note that the relevant interior-domains, TP selected by C and VP selected by v^*, can receive a uniform characterization in this approach, namely that they are SOs that become convergent upon completion of phase-level operations.

Then, it is possible to provide a uniform necessary condition of phases in terms of convergence, as in (41).

(41)   An SO Σ can be a phase only if the interior of Σ is convergent (i.e., containing no unvalued features).

Note that this condition readily distinguishes between transitive v^*P and unaccusative/passive vP: v^*P is a phase since its VP-complement is convergent, for v^*'s probing can ensure the valuation of Case-feature on the object KP; on the other hand, unaccusative/passive v cannot define a phase since its complement VP contains a KP whose Case-feature remains unchecked.[17] Moreover, this definition can be readily extended to K and P, too: assuming that it is K but not its complement that assumes an uninterpretable Case-feature, the struc-

---

[17]See Legate (2003, 2005) and Richards (forthcoming) among others for the different conclusion that unaccusative/passive vPs are also 'strong' phases whose completion triggers immediate Transfer. It is worth noting in this regard that the hypothesis that seems to lie behind their approaches is that dislocation of KPs with [uCase] to the edge of these vPs can somehow make the complement of vPs convergent, an assumption that I will refrain from adopting. Note that the [uCase] located in the 'trace'/copy of the moved KP at the base-position is as unvalued/uninterpretable as the one at the edge of the phase, and therefore, unless stipulated otherwise, its existence keeps making vPs non-convergent.

ture of the form {K, {D, NP}} would readily satisfy (41); P can assign oblique Case to its complement, thus [P XP] also satisfying (41).

It should be noted that (41) is also a null hypothesis straightforwardly deducible from the principle of FI (3). As long as we define phases as SOs whose interiors are handed over CI and SM, they are required to be convergent, since otherwise they cannot receive legitimate interpretations at CI and SM, violating FI. Thus, (41) not only makes the empirically advisable distinction between phases and non-phases, but also arises as a constraint inevitably imposed by FI. Given these considerations, I will adopt (41) as the necessary condition of phasehood. The operation of Transfer is correspondingly reformulated as in (42):

(42)   *Transfer* (third version):

Applied to an SO Σ, Transfer subjects the largest possible convergent term Σ' of Σ (the interior of Σ) to Interpret and Spell-Out, eliminating it from the derivational workspace.

To examine how the mechanism of Transfer as defined in (42) yields the correct demarcation of the relevant phases, let us go through a derivation of a simple sentence, say *...(that) the boy will eat the apple*. First, starting from the object nominal phrase *the apple*, recursive applications of EM construct the SO in (43a). For reasons which we will return to in later chapters, I propose that noun phrases are uniformly headed by a functional category with an unvalued Case-feature [uCase], K(ase) as we may call it (Bittner and Hale 1996a,b, Neeleman and Weerman 1999, Asbury 2008, Caha 2009 and references cited therein; cf. Chomsky's 2007a n^*). I assume that N bears interpretable (valued) number and gender features, and D a person feature, among possibly other things,[18] and further that

---

[18] According to some theories of Distributed Morphology (Halle and Marantz 1993, 1994, Marantz 1996, 1997, 2007, Embick and Marantz 2008), N(oun) is further decomposed into the 'nominalizer' category n and the 'root' category $\sqrt{root}$, an LI whose categorial property is to be specified by a neighboring categorizer like n, v(erb), and a(djective). Thus, for example, a noun 'sight' is decomposed into {n, $\sqrt{see}$}, $\sqrt{see}$ being a category-neutral LI associated with intrisic semantic features (and maybe impoverished bits of phonological features; see, e.g,

K bears, in addition to [uCase], a full set of unvalued φ-features, [uPerson], [uNumber] and [uGender], which is represented as [$u\varphi$] in (43a). By the external merger of K, these unvalued and hence uninterpretable features are introduced to the derivational workspace. Although valuation of [uCase] must be postponed for later computation, [$u\varphi$] can be fully valued by Agree within (43a), taking values from N and D. Thus, K probes N and D and values [$u\varphi$], which applies in simultaneity with Transfer (43b). Given that {the, apple} is the largest convergent term of (43b) (K still assuming [uCase]), Transfer chooses it as the interior of this phase and eliminates it from the derivational workspace, leaving only K for later computation.

(43)   a.

       K
   [uCase,$u\varphi$] the   apple
           [D]   [N]

b.

      K
  [uCase,⊡φ] the   apple
           [D]   [N]

$\longmapsto$ c.
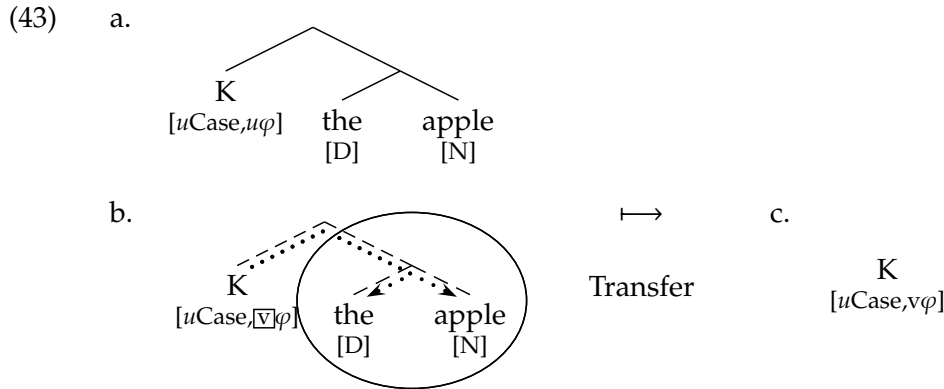
Transfer   K
                [uCase,vφ]

Basically the same analysis can be given to the subject KP, {K, {the, boy}}, which will be reduced in the same way to K with [uCase,vφ] after Transfer.

Next, the *v*P-phase structure in (44a) is constructed by recursive application of EM. Throughout this thesis, I will adopt the so-called predicate-internal subject hypothesis (Koopman and Sportiche 1983, Fukui 1986/1995, Sportiche 1988, Kuroda 1988; see also Marantz 1984, Hale and Keyser 1993, 2002, Hinzen 2006), which holds that the argument structure of a predicate category X is fully 'saturated' strictly locally within the 'projection' of X: stated in terms of bare phrase structure, it says that all nominal arguments of a verbal category are introduced into the SO headed by *v*/*v** by EM. Specifically, the subject is base-

---

Embick and Marantz 2008). This proposal is fully compatible with the analysis proposed in the present thesis.

generated as the 'specifier' of v^* (but see §3.2.2; see also Chapter 5 for further discussion of the notion of 'specifier'). I will also follow the standard analysis in the literature that v^* bears a set of unvalued agreement features $[u\varphi]$ that is responsible for Accusative Case assignment to the object nominal. Thus, after the SO in (44a) is constructed by EM, v^* probes the $[v\varphi]$ of the object K and induces Agree, applying simultaneously with Transfer (44b). Upon valuation of the Case-feature of the object, the interior of the v^*-phase, {eat, K}, becomes convergent, and thus Transfer can eliminate it from the derivational workspace, leaving v^* and the subject K for later computation (44c).

(44) a.

K
$[u\text{Case},v\varphi]$ v^*
$[\text{Acc},u\varphi]$ eat K
$[u\text{Case},v\varphi]$

b. $\longmapsto$ c.

K Transfer K v^*
$[u\text{Case},v\varphi]$ v^* $[u\text{Case},v\varphi]$ $[\text{Acc},v\varphi]$
$[\text{Acc},\boxed{v}\varphi]$ eat K
$[\boxed{\text{Acc}},v\varphi]$
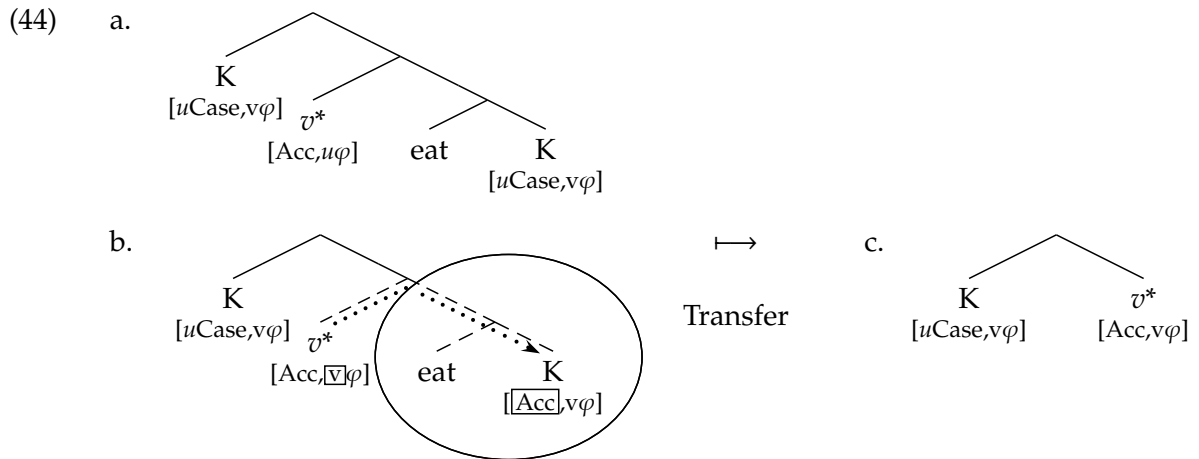
In passing, a question may be raised as to why $\{v^*, \{\text{eat}, K\}\}$ cannot be singled out as the domain to be eliminated by Transfer, even though it contains no unvalued-features and thus it is apparently a convergent domain that is larger than {eat, K}. I have to postpone a fuller discussion of this problem until Chapter 5, but the short answer is that elimination of v^* at this point will prevent v^* from assigning its 'agent' θ-role to the subject K, which will presumably violate the CI requirement that each argument KP should be assigned one and only one θ-role, an aspect of FI traditionally referred to as the 'Θ-criterion' (Chomsky 1981 *et seq.*).

Further, T and C are introduced by EM, forming (45a). I follow the traditional assumption that T bears unvalued agreement features $[u\varphi]$ to be valued via Agree by $[v\varphi]$ of the subject. Further, as argued above, I also adopt the hypothesis that C is primarily

responsible for the Nominative Case assignment to the subject, a property that I will annotate [Nom] in (45) (see Chomsky 2007a, 2008; see also Iatridou 1993, Watanabe 1996, Chomsky 2000, 2001, Hiraiwa 2005, Richards 2007b for various implementations).[19],[20] The consequence of this hypothesis for the present concern is that [uCase] of the subject K remains unchecked until the introduction of C. Thus, EM keeps applying until it reaches the phase-level of C, where Agree and IM apply. Among other things, C and T probe the subject K and establish Agree, valuing [$u\varphi$] of T and [uCase] of K. Moreover, it is known that the subject KP prototypically moves to somewhere preceding T but following C, as the surface word order apparently suggests (...[∅/*that/if/whether*]$_C$ [$_{KP}$ *the boy*] *will*$_T$ *eat the apple*). This is a phenomenon traditionally attributed to the effect of the 'Extended Projection Principle' (EPP), which holds that the specifier of T must be filled by some phrase. Note that the 'T'-node' ({T, {K, v^*}}) is already merged with C by EM, and so, assuming no tampering, the internal merger of K to the relevant node would result in the two partially overlapping SOs in (46). In terms of tree-diagram representations, they correspond to the 'multi-rooted' structure shown in (45b).[21]

---

[19]Richards (2007b) and Chomsky (2007a, 2008) argue that [$u\varphi$] are introduced first at C and then 'inherited' onto T, an interesting proposal that I will largely put aside for the purpose of this thesis. Of course, this operation of 'feature-inheritance' is a prototypical violation of the NTC and the IC, but Richards and Chomsky argue that this departure from the NTC and the IC is nevertheless justifiable on principled grounds.

[20]One of the oft-cited pieces of evidence for the idea that T's [$u\varphi$] originates from C, discussed in the previous footnote, comes from the so-called 'complementizer agreement' phenomenon in a number of varieties of West-Germanic languages, where subjects of finite embedded clauses trigger agreement on the complementizer, as illustrated in (i):

(i)   *South Hollandic Dutch*:
   a.   dat-ə   sə   spel-ə
        that-PL 3-PL play-PL
        '..that they play.'
   b.   dat-(*ə) sə   speel-t
        that-(PL) 3SG.FEM play-3SG
        '...that she plays.'

See Hoekstra and Marácz (1989), Zwart (1993), Watanabe (2000), Carstens (2003) among others for discussions of this phenomenon.

[21]In contrast, Chomsky (2007a, 2008) contends that the EPP-driven movement of the subject to a place above T but below C takes the form of 'tucking-in' (Richards 1997, 1999, 2001), replacing the 'T'-node' ({T, {K, v^*}}) with {K, {T, {K, v^*}}}, an apparent violation of the NTC (5).

(45) a.

b.

$\longmapsto$ c.
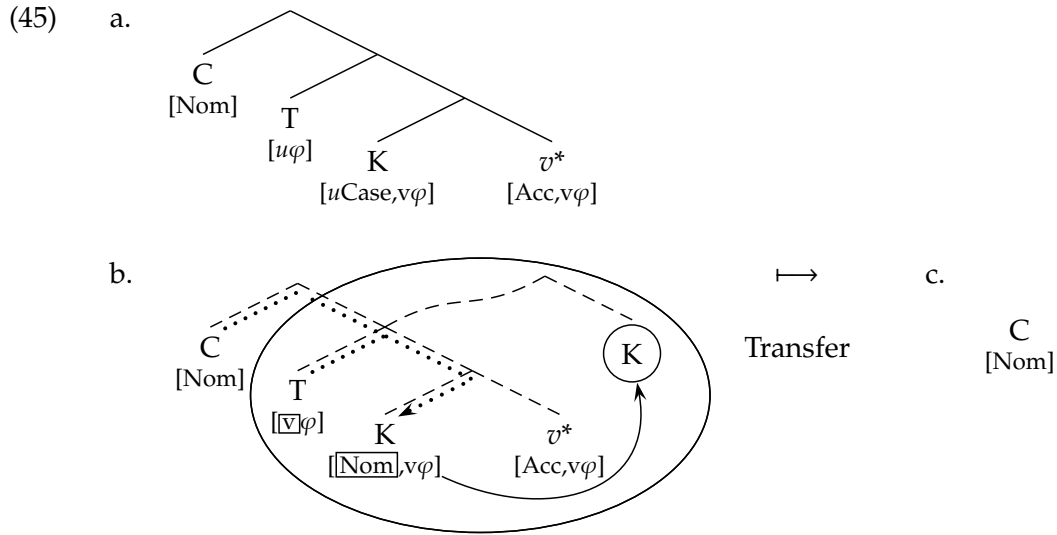
Transfer

(46)   (45b) corresponds to the following two partially overlapping SOs:

a.   {C, {T, {K, v^*}}}

b.   {K, {T, {K, v^*}}}

Now, Transfer should apply in simultaneity with IM, subjecting a certain domain (interior) of the structure to Interpret and Spell-Out. What is the interior of this multi-rooted structure? I maintain that it is the 'second-root' SO, {K, {T, {K, v^*}}} that constitutes the interior of this phase, i.e., the largest convergent term of the phase: [uCase] of K receives Nominative Case from C, [$u\varphi$] of T receives values from the subject K via Agree, and so there is no unvalued feature remaining in it; moreover, it is larger than the sister of C, since the former properly contains the latter. Therefore, the 'second-root' SO, {K, {T, {K, v^*}}} will be immediately subjected to Transfer upon creation by IM, and eliminated from the workspace as in (46c) (see Epstein et al. 2009 for relevant discussion).[22] I will also assume that {T, {K, v^*}} in {C, {T, {K, v^*}}} will also be subjected to Interpret and Spell-Out, by virtue of overlapping with a term of the interior. Therefore, what remains after Transfer is a simplex LI C, to which next-phase operations apply if any. If not, then the phase in (45)

---

[22]I will return to the problem of how such 'multi-rooted' structures are appropriately linearized at PHON in §4.4. See also Chapter 5 for semantic consequences at SEM.

should be the undominated matrix clause, and Transfer just eliminates the whole CP (45b) upon interpretation, concluding the derivation.

The derivation in (43)–(44)–(45) shows that the formulation of Transfer in (41) correctly singles out CP, v*P and KP as phases, while not TP and VP. It thus yields the natural class of phases, whose syntactic independence and coherency are supported by, e.g., the data in (36)–(38). In the following chapters, we will observe further pieces of evidence in favor of this classification of phases v.s. non-phases.

2.5.3 Rationale for Unvalued/uninterpretable Features

Importantly, the discussion of the derivation (43)–(44)–(45) demonstrates the significant role played by various instances of unvalued features. As we saw above, FI demands that the phase-interiors be convergent, free from unvalued features (41). Then, the reverse prediction of (41) is that as long as the interior of an SO contains [uF], it cannot define a phase. Thus, if an LI X with an unvalued feature [uF] is introduced to the active workspace, prototypically [uCase], SOs that contain X cannot constitute phases for Transfer, unless Agree assigns a value to [uF]. The natural expectation then is that the distribution of unvalued features determines phase cycles. Indeed, it was shown above that CP, vP and KP (while not TP or VP) are correctly singled out in accordance with (41) as phases in the derivation sketched above. Given that cyclicity is a desired feature of such computational systems as FL, then, we can even regard unvalued features, though uninterpretable *per se*, as indirectly contributing to computational efficiency by providing local determinations of phase cycles. Then, it is the overarching principle of FI that not only sets the correct demarcation of phases, but also provides an important rationale for the *prima facie* imperfection of FL, unvalued features, a desirable conclusion in the pursuit of the SMT.

It is worth recalling that when the notion of uninterpretable features is introduced into the theory of syntax, Chomsky (2000:§3.5) enumerates (a) uninterpretable features of LIs and (b) the ubiquitous dislocation phenomenon as "two striking examples" of imper-

fections in human language, i.e., unexplained departures of the SMT. He then speculates that FL adopts (a) as the mechanical implementation of (b), with the hope of reducing two imperfections to one. In this approach, he resorts to the familiar stipulation that Move is a costly, 'last resort' operation applying only when necessary to ensure checking of uninterpretable features. However, it has been recognized since Chomsky (2004) that bare phrase structure (Chomsky 1994 *et seq.*) provides arguably the simplest possible conception of Move: the recursively applicable operation, Merge, is so free and unconstrained that it applies freely and indistinctly to any two terms α and β, irrespective of whether they are independent SOs (EM) or one is part of the other (IM). Therefore, IM should be "as free as EM," (Chomsky 2008:140) and there should not be any stipulation that differentiates the cost of application for EM and IM (see already Fukui and Speas 1986, Fukui 1988 and Saito and Fukui 1998 for the costless nature of free/optional IM; see also Kuno 2003). Stipulations of viral "EPP" properties are thus eliminated in favor of the generalized *edge-feature* of LIs, an undeletable feature that allows its bearer to be subjected to unbounded Merge, be it IM or EM.[23] This points to the conclusion that displacement is not an "imperfection" of human language, but rather its *absence* would be an imperfection (Chomsky 2004:110). Indeed, it should be a departure from the SMT and hence from the goal of biological adequacy to stipulate a separate cost for the applicability of IM. From this perspective, it should be concluded that we cannot blame the existence of uninterpretable features on dislocation anymore. Correspondingly, Chomsky's (2000) earlier reasoning that uninterpretable features are justified as actual triggers of another 'imperfection', displacement, should be eliminated as a misguided stipulation. We are thus left again with uninterpretable features qua imperfections. To the extent that the existence of unvalued features is a real property of FL, as seems thoroughly demonstrated by various morphological agreement phenomena in natural language, then the minimalist question is why UG has to assume such apparent

---

[23]Chomsky (2008). For the notion of edge-feature, see in particular Chapter 3 for relevant discussion; see also Narita (2009a) and Fukui (2008, forthcoming).

'imperfections'. We have to either (i) give up the minimalist goal of biological adequacy in face of such imperfections, or (ii) aim at providing an explicit hypothesis for how they enter into UG in fact as a part of a "best way" to satisfy design specifications of FL.

In this regard, the above discussion points to another important function of unvalued features: their distribution locally demarcates phase cycles. Under our hypothesis that phasehood is defined in terms of convergence (41), the distribution of unvalued features, such as [$u\varphi$] and [uCase], rather constitutes a primary factor of detecting phases. It was shown that (41) indeed provides the correct demarcation of CP-, v*P- and KP-phases. To the extent that phase cyclicity is a highly desirable feature of syntactic computation, contributing to periodic reduction of computational load, then, we may provide an important alternative rationale for the existence of distributed unvalued/uninterpretable features in the architecture of FL, namely as a device to efficiently detect phase cycles. In this approach, we may even regard them as indirectly contributing to the desideratum of computational efficiency, keeping to the desideratum of the SMT.[24]

2.6 Eliminating the Merge-over-Move Stipulation along with Numeration and Lexical (Sub-)arrays

In the previous discussion, we reached the conclusion that IM and Agree must apply simultaneously with Transfer (see (25) and (31)). It was also noted that application of EM is not constrained that way, due to the lack of the copy-identification requirement that is associated with IM and Agree. This way, EM is truly unconstrained, freely applicable without being confined to cyclicity by Transfer. However, it should be noted here that we are *not* stipulating that IM is an operation that is inherently more costly than EM. No such stipulation is necessary, and is to be excluded by the SMT. Derivational simultaneity of IM and Transfer (25) is rather derived from the interplay of the IC, the NTC, FI and the

---

[24]I thank Noam Chomsky (p.c.) for bringing this point to my attention.

interpretive demand of copy-identification, and hence there is no principled reason to add a further stipulation that IM is differentiated in its computational cost from EM.

In the early exploration on bare phrase structure, Chomsky (1994, 1995) put forward the hypothesis that has come to be referred to as the *Merge-over-Move principle* (MOM). MOM holds that Move (= IM) is an operation that is computationally more costly than EM, so that syntax always applies EM whenever there arises a choice between EM and IM. It is further speculated that this cost-difference arises probably because Move is in fact a complex operation that comprises Merge, feature-checking, and pied-piping (Chomsky 1995, 2000; see also Chomsky 2001) or because Move is necessarily accompanied by subsequent phonological operations that eliminate all but one copy at the phonological component (Kitahara 1997). However, as noted, it came to be recognized since Chomsky (2004) that once syntax is given Merge as set-formation (9), both EM and IM come for free. Barring further stipulations, both modes of Merge apply freely, hence IM should be as free as EM (Chomsky 2008:140). In this line of reasoning, then, we are compelled to conclude that there should be no stipulation to the effect of MOM. In this section, I will argue that the earlier stipulation of MOM can be naturally eliminated in favor of cyclic derivation by phase.

Chomsky (1994, 1995) argues that MOM explains the contrast in examples like (47).

(47)   a.   *There$_i$* seems [t_i to be *a man* in the room].

       b.   **There* seems [*a man$_i$* to be t_i in the room].

According to the analysis proposed by Chomsky (1995), syntax starts its derivation by constructing *Numeration*, a set of pairs (LI, *i*), where LI is an item of the lexicon and *i* is its index that indicates the number of times that LI is selected from the Lexicon in a given derivation. The indices are reduced by 1 when the relevant LI is chosen and introduced to the active workspace (this operation is called *Select* in Chomsky 1995), and they are reduced

to zero at the end of derivation. For example, at the point where the infinitival T′ in (48) is constructed, the Numeration N for the derivation that yields the sentence in (47a) is of the form in (49).

(48)   [to$_T$ [be [a man] in the room]]

(49)   N = {(C, 1), (T$_{present}$, 1), (*there*, 1), (*v*, 1), (*seem*$_V$, 1), (*to*$_T$, 0), (*be*, 0), (*a*, 0), etc.}

In the model of Chomsky (1995), it was stipulated that the infinitival T *to* has a so-called 'EPP-feature' (or [–D]) that requires its specifier position to be filled by some determiner phrase. At the point in (48), this EPP-requirement can be met either (i) by externally merging the expletive *there* drawn from *N* in (49), or (ii) by moving the associate DP [*a man*]. (i) and (ii) would yield the derivations in (47a) and (47b), respectively. The severe ungrammaticality of (47b) suggests that the movement option (ii) is unavailable at (48), which Chomsky argues is naturally predicted by MOM: the presence of (*there*, 1) in (49) indicates that both options of EM and Move are available, and syntax always chooses EM in favor of MOM, precluding the derivation in (47b). This way, MOM together with the concept of Numeration was argued to receive empirical support from the data like (47).

However, it is pointed out in Chomsky (2000:106) that MOM in tandem with Numeration makes a wrong prediction with regard to simple examples like the following (his (7b)).

(50)   There is a possibility [that *proofs*$_i$ will be discovered *t*$_i$].

At the point where the embedded clause is constructed, movement of *proofs* to Spec-T is possible, even though the Numeration at that stage would still contain an expletive *there* (or (*there*, 1)), and thus MOM would preclude Move from applying within the embedded clause, contrary to the full acceptability of (50).

One might try to defend MOM by saying that MOM is just a relative preference

but not an absolute requirement, thus its effect can be overridden if only the choice of IM yields a convergent derivation. However, this line of approach turns out to be problematic in face of examples like (51)-(52) of the sort discovered by Alec Marantz and Juan Romero (see, e.g., Uriagereka 2008:Ch.1 for relevant discussion).

(51)   a.   (...and) the fact is [that there is a monk in the cloister]

        b.   (...and) there is the fact [that *a monk$_i$* is t_i in the cloister]

(52)   a.   A ball-room was [where there was a monk arrested]

        b.   There was a ball-room [where *a monk$_i$* was t_i arrested]

It is apparent from these examples that one and the same Numeration can in fact yield more than one convergent derivation that seem to differ only in the choice of expletive-insertion and subject raising at a particular point of derivation. Therefore, these examples show that MOM cannot be sustained with the original concept of Numeration, which in turn render the original MOM-based account of (47) unsustainable.

Still in defense of MOM, Chomsky (2000) argues that this problem can be technically circumvented if we speculate that the initial lexical array is further partitioned into smaller chunks corresponding to derivational phases. These partitions are called *lexical subarrays*. For example, *LA$_i$* and *LA$_j$* below represents the lexical subarrays corresponding to the embedded CPs in (51a) and (51b), respectively. (Mainly for expository purposes, here I keep to the assumption that the initial lexical array takes the form of Numeration in the sense of Chomsky 1995, and hence that lexical subarrays are subsets of Numeration. However, note that Chomsky 2000 refrained from adopting his earlier hypothesis that lexical arrays are structured in the form of Numeration. See also Chomsky 2000: fn. 41.)

(53)   *LA$_i$* = {(*that*, 1), (T, 1), (*be*, 1), (*a*, 1), (*monk*, 1), (*in*, 1), (*the*, 1), (*cloister*, 1), (*there*, 1)}

(54)   *LA$_j$* = {(*that*, 1), (T, 1), (*be*, 1), (*a*, 1), (*monk*, 1), (*in*, 1), (*the*, 1), (*cloister*, 1)}

LA_i and LA_j differ in that only the former contains a token of *there*. Now, these lexical subarrays yield one and only one convergent CP-phase as shown with square brackets in (51a) and (51b). Note that there arises no choice between expletive-insertion and subject-raising in these CP-phases: *there*-insertion applies if and only if the lexical subarray contains a token of *there*, period. Here, presyntactic selections of lexical subarrays render MOM by and large irrelevant for the derivations of these CP-phases. On the other hand, Chomsky (2000) argues, MOM is still applicable within each phase of a lexical subarray, and the MOM-based account of (47) can be sustained in its original form: given that TP cannot be a phase on its own, the only phase involved in the derivation of (47a) is the matrix CP itself, and hence the initial lexical array with a token of *there* equals the lexical subarray associated with the CP. Within this phase cycle, then, MOM derives the priority of EM of *there* over movement of *a man* at the point of (48), hence the latter derivation that would yield (47b) is precluded. In this line of reasoning, Chomsky (2000) defends the idea that the contrast between (47a) and (47b) still supports the stipulation of MOM—a conclusion that is now recognized as undesirable, given the natural hypothesis that EM and IM are just instances of one and the same operation and hence there should not be any difference in computational cost between them.

Now, I would like to maintain that we can eliminate recourse to MOM in the account of (47), in favor of our conclusion in (25), namely that IM but not EM is constrained to apply simultaneously with Transfer. This hypothesis gives rise to ordering of operations in syntactic computation, in which instances of EM in a given phase apply first, and then all instances of IM (and Value) apply in tandem with Transfer at that phase.

(55)   Instances of EM apply prior to instances of IM and Agree in a given phase.

Applied to the matrix C-phase in (47), then, (55) is sufficient to derive the priority of external merger of *there* over internal merger of *a man*: *there* must be introduced by EM prior to the movement of *a man*, thus there arises no choice between EM and IM at the point of (48),

simply as a consequence of (55). MOM need not be stipulated in this account, and the ungrammaticality of (47b) is derived from the simple fact that internal merger of *a man* is not possible at (48), given that infinitival TP is not a phase. EM keeps applying until the matrix C is introduced, and at the phase level determined by it, IM dislocates the closest goal *there* to the EPP-position.

(56) a.

b.
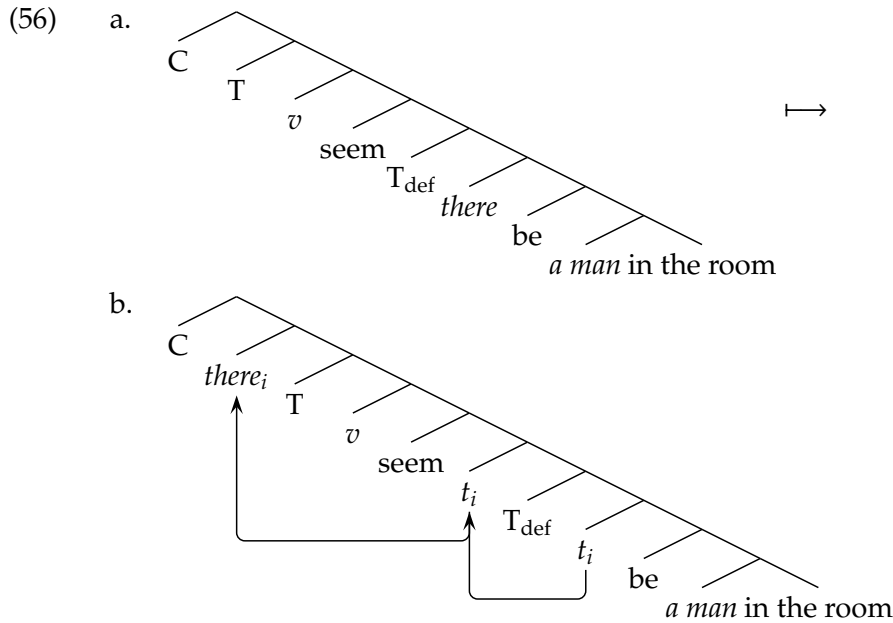
Moreover, it should be noted that this approach can eliminate recourse to lexical subarrays, too. As discussed above, the concept of lexical subarray is introduced by Chomsky (2000) mainly as a technology to prolong the life of MOM, which apparently fails to account for examples like those in (50)–(52). However, once we free ourselves from the stipulation of MOM in favor of (55), these examples also cease to be problematic. The alternation between CP with an expletive subject and one without it (say *there was a man in the room* v.s. *a man was in the room*) arises simply from the unboundedness of EM: a derivation with external merge of an expletive results in the former while one without it results in the latter. The concept of lexical subarray loses its relevance in this account. More generally, pre-syntactic assembly of LIs, of any form, becomes an extraneous, unmotivated stipulation in phase theory. Then, the natural conclusion is that such a technology should

also be dispensed with in favor of phase theory.

2.7   Copy-identification as Part of Transfer

At this point, it is instructive to attend closer to the original concept of Numeration provided by Chomsky (1995). As noted, it was formulated as a set of pairs (LI, *i*) that serves not only as input to narrow syntactic computation (referred to as human language computation (C_{HL}) therein), but also as a device to keep track of the number of distinguished occurrences of LIs in a given derivation. Specifically, Chomsky (1995) put forward the hypothesis that occurrences of one and the same LI introduced to syntax by distinct applications of *Select*, a procedure that draws a lexical item LI from the Numeration and reduces its index by 1, are marked as distinct from each other. Thus, he notes,

> Suppose the lexical item LI in the numeration *N* has index *i*. If a derivation is to be generated, Select must access LI *i* times, introducing it into the derivation. But the syntactic objects formed by distinct applications of Select to LI must be distinguished; two occurrences of the pronoun *he*, for example, may have entirely different properties at LF. *l* and *l'* are thus marked as distinct for C_{HL}, if they are formed by distinct applications of Select accessing the same lexical item of *N*. *Note that this is a departure from the inclusiveness condition*, but one that seems indispensable: it is rooted in the nature of language, and perhaps reducible to bare output conditions. (*ibid*, p.227, emphasis mine)

Importantly, it is explicitly acknowledged here that the mechanism of distinctness marking on occurrences of LIs, applying as part of Numeration and/or Select, is a departure from the IC. The idea that some mechanism of index-assignments is responsible for distinctness marking has earlier roots, e.g., the treatment of coreference and binding by means of coindexation as proposed by Chomsky (1980a, 1981, 1986b), Fiengo and May (1994) among others, for which the same problem of inclusiveness arises.

More generally, copy-identification, subsuming formation of chains of movement, is one of the fundamental descriptive problems that in some way or another led linguists to add the stipulations of referential indices and distinctness marking. Given that the indices and distinctness markings undergo manipulation in the course of derivation, they cannot

be part of the initial input provided by the Lexicon and thus violate the IC, irrespective of whether they are assigned to syntactic constituents as part of Numeration-construction or by some other mechanisms. There is no doubt that these are devices of much descriptive convenience,[25] but the worrisome aspect of them is that they are nevertheless departures from the IC. Further, they may also violate the NTC, if they inspect and modify the indices of constituents already formed.

However, it is not clear whether such extraneous mechanisms of distinctness marking are really 'indispensable' in the theory of phases: see, e.g., Chomsky (2007a:10). This is especially true given that one of the core empirical motivations for the hypothesis of Numeration-formation, namely the formulation of MOM, is undermined in favor of phase theory. Recall further the conclusion we reached in §2.4.1 that the derivational simultaneity of IM and Transfer paves the way for strict adherence to the IC. As noted above, the distinction between IM and EM is derivational in nature: applying to two SOs α and β, the former mode of Merge-application searches inside Y and selects α from terms of β, while α and β are two then independent SOs in the case of EM, but these two modes of Merge possibly result in SOs that are representationally indistinguishable. We concluded that the efficient utilization of the bifurcation between EM and IM by CI and SM thus requires minimization of time-difference between applications of IM and Transfer, resulting in derivational simultaneity (see (25)). In this model of derivation by phase, a simple procedure is available for detecting copies, which holds the following, adapted from Chomsky (2007a:10).[26]

---

[25]Thus, Chomsky notes elsewhere that "chain properties can be reduced in significant part to identity if lexical arrays are enriched to numerations." (2000:note 41) Only 'in significant part', because distinctness marking by Numeration can only take care of identity of LIs, leaving the problem of copy-identification for phrasal constituents unsettled. However, it may be of some interest to note that the problem of copy-identification for phrasal constituents can be totally dispensed with if syntax permits no movement (internal merger) of phrasal constituents, as I will argue in Chapter 3.

[26]Therefore, there is no need to stipulate a distinct rule of copy-formation, contrary to Hornstein's (2009) theory of syntax.

(57)   All and only elements (feature-values or SOs) introduced in simultaneity with
         Transfer are copies of elements within the interior of the phase.

This algorithm naturally achieves identification of copies formed by IM without any ex-
traneous mechanisms of index-assignments, and hence it is still in conformity with inclu-
siveness, satisfying the SMT. I contend that this simple algorithm of copy-identification is
also available for determining the identity of feature-values copied by Agree, which again
applies simultaneously with Transfer (31).

It is simply a matter of fact in the framework of bare phrase structure that copy-
formation by IM and Agree is exploited by CI and SM. At PHON, copy-formation by IM
yields the ubiquitous displacement phenomenon, and copies of feature-values formed by
Agree represent morphological covariation phenomena in various languages. At SEM,
IM yields discourse-related properties such as scope and operator-variable interpretations,
constituting an integral part of the duality of semantics, correlating closely with the EM
v.s. IM bifurcation. Therefore, FL should be designed in such a way that the derivational
information about copy-formation is appropriately delivered to CI and SM. This much can
really be regarded as "rooted in the nature of language, and perhaps reducible to bare
output conditions"—it is just that the inclusiveness-violating mechanism of indexation
and distinctness marking is not warranted by such considerations. In conjunction with
this, recall that the copy-forming operations, IM and Agree, apply simultaneously in the
derivation with Transfer. It was argued in §2.4 that the derivational simultaneity arises
as a consequence of computational efficiency, minimizing the time-difference between
copy-formation and Transfer. Pushing this line of reasoning even further, I contend that
copy-identification applies *as integral part of* Transfer. The derivational simultaneity of IM,
Agree and Transfer simply follows from this hypothesis. Along these lines of reasoning,
the formulation of Transfer is further refined as in (58):

(58)   *Transfer* (final version):

Applied to an SO Σ, Transfer identifies copies within Σ, and subjects the largest possible convergent term Σ′ of Σ (the interior of Σ) to Interpret and Spell-Out, eliminating it from the derivational workspace.

The following chapters will investigate further properties of Transfer. We will see that this hypothesis of Transfer yields a number of predictions that are desirable both in terms of breadth of empirical coverage and theory-internal simplicity, approaching the desideratum of the SMT.

2.8   Concluding Remarks

This chapter attempted to approach the problems posed by FI (3), the IC (4) and the NTC (5) from the perspective of the modern minimalist framework.

(3)   *Full Interpretation (FI)* :

Every constituent of SEM and PHON contributes to interpretation.

(4)   *Inclusiveness Condition (IC)* :

No new features are introduced in the course of linguistic derivation.

(5)   *No-Tampering Condition (NTC)* :

No elements introduced by syntax are deleted or modified in the course of linguistic derivation.

It was argued that the strict adherence to these principles of computational efficiency derives a number of nontrivial consequences on the technical workings of syntactic derivation. Specifically, I argued that the theories of bare phrase structure and phase cyclicity of the Chomskyan sort pave the way for adhering to these derivatives of computational efficiency. The phase-theoretic conclusions we reached in this chapter are summarized

below:

(59)   a.   ***Derivational Simultaneity*** :

Interpret, Spell-Out, and copy-formation operations (IM and Agree) apply

simultaneously in the form of Transfer.

b.   ***Transfer*** (final version):

Applied to an SO Σ, Transfer identifies copies within Σ, and subjects the largest

possible convergent term Σ' of Σ (the interior of Σ) to Interpret and Spell-Out,

eliminating it from the derivational workspace.

c.   ***Strict satisfaction of Inclusiveness and No-tampering*** :

No violation of the IC (4) or the NTC (5) is warranted: referential indices and

distinctness marking, pre-syntactic assembly of LIs in the form of Numeration

or lexical subarray, etc., are thus naturally dispensed with in favor of phase

theory.

All in all, the overarching conclusion of this chapter is that, so long as we try to approach

the theory of bare phrase structure 'from below', i.e. from the minimal set of assumptions,

our theory of syntax must accommodate the notion of derivational simultaneity of Spell-

Out, Interpret, IM and Agree, a sum of which apply cyclically phase-by-phase in a given

derivation. Various technical devices such as indices, Numeration and lexical subarrays

are shown to be dispensable, keeping closely to the conditions imposed by (3)–(5). To the

extent that these results have certain explanatory force, as I will argue in the following

chapters, we can rather regard the combination of FI, the IC and the NTC as providing a

foundational rationale for the system of bare phrase structure and derivation by phase.

It should be stressed that we are putting forward these technical analyses primarily

as a way to achieve an empirically adequate account of the generative capacity of FL. Recall

that the problems of inclusiveness and no-tampering have their roots in the old concerns of

descriptive and explanatory adequacy. It is important to recognize these issues in order to

properly restrict the generative power of syntax, and theories of transformations have been attempting to discover empirically adequate constraints on the applicability of insertion and tampering. Specifically, the following question is posed at the beginning of this chapter.

(60)   Why are applications of insertion and tampering severely restricted in the syntax of natural language?

The present chapter attempted to approach this problem by posing the minimalist desiderata of the IC and the NTC. Earlier theories of transformations were by their very nature systems of insertion and tampering, and therefore, it is a pressing empirical problem for any framework of transformational generative grammar to explicate how the theory can restrict applications of insertion and tampering on principled grounds. In a nutshell, this chapter is an attempt to provide the simple answer in (61) from the perspective of the SMT.

(61)   Insertion and tampering are disallowed because there are no such operations in syntax as a matter of principle.

The concepts of the IC and the NTC are rather minimalist formulations of this overarching empirical hypothesis. It was observed that the framework of bare phrase structure, when accompanied by the hypotheses in (59), can significantly reduce departures from the IC, the NTC, and FI as well. It was claimed that the strict adherence to these principles of computational efficiency virtually exorcises insertion and tampering from the architecture of FL, satisfying the SMT.

Putting forward the claims in (59) is, of course, an empirical hypothesis that should be tested against the facts about human language. The task for the following chapters is to show how the theory of phases and bare phrase structure proposed in this chapter can be proven to be adequate in terms of its empirical predictions and also its contribution to the SMT.

Chapter 3

Endocentric Structuring of

Projection-free Syntax

3.1 Introduction: Towards Projection-free Bare Phrase Structure

The theory of bare phrase structure under exploration holds that the compositional structure of human-language expressions is exhaustively characterized by recursive application of Merge. Merge is a binary set-formation operation that takes two syntactic objects (SOs), say α and β, and creates a set $\{\alpha, \beta\}$.

(1)   Merge(α, β) =   a. $\{\alpha, \beta\}$

                                                 b.   (order irrelevant)

$$\overset{\displaystyle\wedge}{\underset{\alpha \qquad \beta}{}}$$

By assumption, Merge can apply recursively, i.e., the output of Merge can constitute further input to the same operation: thus, Merge can combine $\{\alpha, \beta\}$ with some other SO, say γ, forming another set-theoretic structure $\{\gamma, \{\alpha, \beta\}\}$. Under bare phrase structure, then, all phrase-markers generated by syntax are reduced to set-theoretic objects of various orders. The definition of syntactic objects (SOs) is reproduced here:

73

(2) α is a *syntactic object* (SO) iff

    a. α is an LI, or

    b. α is a set $\{\beta, \gamma\}$, where β and γ are SOs.

The question then is how much the properties of SEM and PHON, the interface representations for CI and SM, can be exhaustively characterized by SOs. Obviously, the minimal theory of syntax will characterize SEM and PHON as keeping closely to SOs as defined in (2), insertion of new features and tampering of internal structures being kept to the bare minimum. It was specifically argued in the previous discussion that so long as we keep to the conception of SOs as defined in (2), the theory of bare phrase structure maximally satisfies the Inclusiveness Condition (IC) and the No-tampering Condition (NTC).

(3)   ***Inclusiveness Condition (IC)*** :

    No new features are introduced in the course of linguistic derivation.

(4)   ***No-Tampering Condition (NTC)*** :

    No elements introduced by syntax are deleted or modified in the course of linguistic derivation.

These constraints are put forward not only as deductive consequences of the SMT but also as necessary parts of the hypothesis that syntax assumes virtually no operations of insertion and tampering, thus meeting the tripartite criteria of descriptive, explanatory and biological adequacy.

Note that no known condition of set theory requires that SOs, being sets, are associated with any 'labels' or distinct nonterminal symbols. As the 'label-less' set-theoretic representations in (1) indicates, Merge simply serves for compositional structuring of constituents, to the exclusion of labeling as well as precedence. Then, only further stipulations can guarantee assignment of any such extraneous properties to set-theoretic objects gener-

ated by Merge. Note further that such addition of new features to SOs is a straightforward violation of the IC (3), which constitutes an integral part of our principled eradication of insertion (and tampering) rules from the theory of syntax. Correspondingly, the insertion of labels in bare phrase structure is a worrisome departure from the SMT, subsidiary enrichment of UG (Chomsky 2007a:23; see also Collins 2002, Seely 2006). Then, we must ask if there is any strong empirical motivation that demands the theory of syntax to provide a distinct mechanism of labeling separate from the operation Merge.

The sixty years of research in the generative tradition have cemented linguists' convention of drawing labeled tree-diagrams. The idea of labeled phrase structure owes its origin to the earlier transformational grammar advanced by Chomsky (1955/1975, 1957, 1965), in which it was assumed that base structures of sentences (the kernel of the language, or 'deep structures') are generated by the system of *phrase structure rules* (PSRs), each of which takes a nonterminal symbol (such as NP, VP, and S) as its input and expands it into a string of terminal and/or nonterminal symbols. The PSR-based conception of phrase structure thus holds that all phrasal SOs are labeled by such nonterminal symbols as a necessary consequence of compositional structuring. The *X-bar theory* developed by Chomsky (1970 *et seq.*) further advances the idea that assignment of labels to phrasal nodes is achieved by *projection* of the features of X^o up to the categories of X' and X''. According to this conception of labeled phrase structure, nonterminal labels like X' and X'' are 'projections' of X: N' is an intermediate projection of N, V'' is the maximal projection of V, and so on. The class of possible labels is thus radically reduced to the combination of locally projected features of lexical items and bar-level indices under this hypothesis, but the concept of labeled phrase structure was essentially carried over from PSRs by and large intact.

It should be acknowledged that the hypothesized feature-percolation via bar-level projections is proposed in the X-bar theory as a simple device to capture the *endocentricity*

(*headedness*) of phrase structure, the essential idea of which can be summarized as (5) (cf. Lyons 1968):

(5)   ***Endocentricity (headedness) of phrase structure*** :

The interpretive properties of a phrase Σ are determined largely by the features of a prominent lexical item within Σ (referred to as the *head* of Σ).

In general, there is a certain central element, called the *head*, for each phrase that determines the essential properties of the phrase, thus a verb phrase is predicative due to the property of the head verb, referential properties of a noun phrase is determined by the head noun, and so on. One may argue that labeling by projection is still necessary even in bare phrase structure to capture the facts about endocentricity. However, such an argument would go through only if it is further shown that labeling by projection is *the* necessary device of commendable theory-internal simplicity to encode endocentricity. This was arguably the case in the earlier PSR-based syntax, where labels were anyway generated as a necessary component of compositional structuring (consider, e.g., S in S → NP VP), but not any more in bare phrase structure, where labeling is bound to be a departure from the SMT, as we saw above.

Note that Chomsky (1994, 1995), in advancing the earlier theory of bare phrase structure, attempted to incorporate the concept of labeled phrase structure into the definition of Merge while avoiding the stipulation of feature-percolation. His hypothesis was essentially that labeling of each SO is achieved by copying an LI within it. It was stipulated in the definition of Merge that when two LIs X and Y are combined, forming {X, Y}, one of the LIs should be necessarily further copied to the edge of this SO, forming either {X, {X, Y}} or {Y, {X, Y}}:

(6)   Merge(X, Y)   =   a.   {X, {X, Y}}, or

b.   {Y, {X, Y}}

If we straightforwardly represent these output representations in terms of innocent, structurally equivalent but nevertheless visually convenient tree-diagrams, we would get the following:

(7)   Merge(X, Y)   =   a′.   (order irrelevant)

X
    X   Y

b′.

Y
    X   Y

However, it was misleadingly stipulated in Chomsky (1994, 1995) that the output set-theoretic objects of Merge in (6a,b) are somehow mapped to labeled phrase structure of the following sort:

(8)   a″.   X   (order irrelevant)

X   Y

b″.   Y

X   Y

The stipulated mechanism of mapping from the real output of Merge in (6)/(7) to set-theoretically inexplicable labeled structures in (8) has never been clarified in the literature, presumably because of the historically intensified familiarity of labeled phrase structure. But if we attend to just the real set-theoretic object in (6)/(7), {X, {X, Y}} and {Y, {X, Y}}, the following should be rather clear: it was hypothesized in Chomsky (1994, 1995) that labeling of phrase structure is achieved by copying one of the Merge-mates by means of *internal Merge* (IM). If we were to adopt this hypothesis, we would be forced to conclude that virtually every application of Merge, EM or IM, should be accompanied by such subsidiary IM. This virtually nullifies the effect of the EM v.s. IM bifurcation, and correspondingly we lose the simplest means of capturing the duality of semantics at CI (§2.4.1). Furthermore, such proliferation of copy-formation would significantly increase the computational load

of syntax, so long as the IC (3) makes no room for any representational means of signaling the identity of copies created by IM, be it referential indices or distinctness marking upon Select/Numeration-formation (§2.7).[1] There is then little question as to the fact that the definition of Merge in (6) is simply a departure from the SMT, and we are forced to stick to the simpler formulation of Merge in (1), which is further forced by the principle of minimal computation (§2.2).

In this context, it should be noted that there are some proposals in the framework of bare phrase structure that make little to no recourse to labeling by projection nor copying by IM for the account of endocentricity. For example, I contend that the series of work by Chomsky (2000, 2008) can and should be understood as paving the way for the projection-free account of endocentricity: Chomsky (2000) hypothesizes that determination of the head of a set-theoretic object $\{\alpha, \beta\}$ correlates with selectional or probe-goal dependency between α and β, an idea followed by a number of researchers. Under this proposal, $\{\alpha, \beta\}$ is headed by (the head of) α if α selects or probes β. Collins (2002) is right in pointing out that this sort of selection/probe-based mechanism can be adapted to encode the centrality of 'head' LIs without making recourse to labels/projections, and his locus principle is a particular attempt to pursue this line of approach. Still after the selector/probe-based mechanism of head-detection became unavailable due to the elimination of selection from the theory of narrow syntax (Chomsky 2004:112-113; see also Pesetsky 1982), Chomsky (2008) puts forward another algorithm of head-detection (still called the 'labeling algorithm') whereby $\{\alpha, \beta\}$ is headed by (the head of) α if (i) α is an LI, or (ii) β is internally merged to form $\{\alpha, \beta\}$. Chomsky (p.c., lectures at MIT in fall 2010) further suggests that (ii) should be eliminated from this algorithm, reducing head-detection to minimal search of an LI for each phrase (i), a proposal to which we will return in §3.4 (see also Piattelli-Palmarini et al. 2009:52ff).

---

[1]Recall further our conclusion that in order to let copy-identities reach SEM and PHON, then, there will be no way but to subjecting SOs to Interpret and Spell-Out simultaneously with the relevant copy-formation operations, a requirement that constitutes an important rationale for the theory of phases articulated in Chapter 2 and explored below.

See also Narita (2009a) for a different approach. These are all attempts to account for endocentricity without reference to any extraneous mechanism of projection, maximally respecting the SMT.[2]

Building on the prospect of the projection-free approaches to endocentricity, this chapter aims to make a step toward constructing a theory of bare phrase structure without labeling/projection. It will be argued in this chapter that, when coupled with the theory of phases developed in the previous chapter, projection-free syntax is indeed superior to projection-bound syntax both in terms of empirical coverage and theory-internal simplicity.

3.2 The H-α schema in Projection-free Syntax

In order to approach projection-free syntax, I will start our discussion by examining how the notion of labeling/projection is employed in the framework of bare phrase structure. It will be specifically pointed out that, despite his remark that labeling/projection is possibly a dispensable notion, Chomsky's theory of syntax nonetheless fails to eradicate recourse to projection, most notably the one involving the 'edge-feature' of lexical items.

3.2.1 Lexical Items and the Edge-feature

Any computational system of discrete infinity such as human language has to assume a recursively applicable generative mechanism, whose simplest formulation is represented by Merge. However, in order for syntax to even start applying Merge at all, there must exist some elements which can constitute input to Merge and are themselves not constructed by Merge. Such SOs constitute computational 'atoms' of the Merge-based recursive generation of SOs. Following the standard terminology, we called these SOs *lexical*

---

[2]Note also that there are some proposals on somewhat less local instantiations of feature-projection, such as WH-feature-percolation (Nishigauchi 1990b and others) and focus-projection (Selkirk 1995 among others). These mechanisms are again residues of earlier labeled phrase structure, and various proposals have been made to overcome the relevant descriptive demands without making recourse to projection: see Cable (2007, 2010) and the following discussion for the Q-based account of WH-dependency; see also Uriagereka (2008) and Irurtzun (2006, 2007) for some suggestive approaches to focus.

items (LIs). How the human mind can acquire the correct set of LIs from impoverished primary linguistic data at all is an important empirical question (see §3.5 for quite relevant discussion), but at the very least, each LI must be associated with a property that permits it to be subjected to Merge. Chomsky calls this property the *edge-feature* of the LI. The hypothesis that LIs can be subjected to Merge is an ineliminable assumption for any theory of human language, and this hypothesis can be justifiably restated as the notion that LIs are associated with the EF. Then, we can postulate as an axiom that there exist LIs with EFs. Adopting the standard terminology, we define the *Lexicon* as a component of human language that stores LIs and hands them to syntactic computation.

Minimally, the EF is a feature that enables LIs to be subjected to Merge. What is a natural hypothesis that derives this effect? Regarding the characterization of EFs, Chomsky (2008) specifically provides the following remark.

> "For an LI to be able to enter into a computation, merging with some SO, it must have some property permitting this operation. A property of an LI is called a *feature*, so an LI has a feature that permits it to be merged. Call this the *edge-feature* (EF) of the LI. ... When merged with a syntactic object SO, LI forms {LI, SO}; SO is its complement. The fact that Merge iterates without limit is a property at least of LIs—*and optimally, only of LIs*, as I will assume." (Chomsky 2008:139, emphasis mine)

Here, Chomsky states that (9) is the effect that any conception of EFs should capture.

(9)   The EF is the feature that permits its bearer to be merged with some SO.

One way to make sense of (9) is to assume that the EF is the feature that constitutes the locus, or trigger, of Merge-application, a formulation reminiscent of the earlier 'EPP-feature' of Chomsky (1995) that states, "I need a specifier." Generalizing to the merger of 'complement' along the line suggested by Chomsky, we may say that the EF is the property of LIs that states, "I may have a sister."

More importantly, Chomsky also proposes a rather strong hypothesis regarding the distribution of EFs, which is (10):

(10)   The EF is a property only of LIs.

I agree with Chomsky in regarding (9) and (10) as a minimal set of assumptions. As stated above, the assumption that LIs are associated with EFs is ineliminable, since otherwise Merge cannot even start assembling LIs. However, once we grant the association of EFs with LIs, no further stipulation is necessary to ensure compositional structuring. Thus, without any such stipulation, it is LIs and only LIs that are associated with EFs with the property (9), as I will assume in this chapter.

It should be noted that once we adopt (10), (11) arises as a logical consequence, inevitable unless some further UG-enriching stipulation is provided:

(11)   No phrases (non-LIs) have EFs.

As an illustration, imagine a simple derivation where syntax operates Merge on two LIs, say X and Y. Both of these LIs are associated with EFs, so both of them can be merged with the other. The result is a set comprising them, {X, Y}. Importantly, the resultant SO is not an LI, thus by assumption (10), this phrasal SO is precluded from having an EF. This situation can be schematized as in (12), where the subscript e stands for the EF:

(12) $X_e + Y_e \quad \longmapsto \quad \{X_e, Y_e\}$
$\qquad\qquad\quad$ Merge $(* \{X_e, Y_e\}_e)$

The absence of EFs should be true for any phrase (non-LI) in general.

Now, let me point out that, by the combination of (9) and (10)/(11), we are forced to conclude (13):

(13)   Only LIs are permitted to be merged with some SO.

That is, any instance of Merge must take an LI as at least one of its inputs, utilizing its EF as a locus of the Merge operation combining it with the other SO. That is, all instances of Merge obey the form {H, α}, where H is an LI and α is an SO (be it another LI or a phrase).

I will refer to this constraint, which is in itself just a direct consequence of (10), as *the H-α schema*.

(14)   ***The H-α schema***: Merge(H, α) → {H, α}.

Merge must take at least one LI as its input.

The H-α schema (14) requires that at least one of the Merge-mates be an LI for any instance of Merge. Note that the H-α schema makes a very strong prediction (15):

(15)   *{*XP, XP*}: No two phrases (non-LIs) can be merged.

At face value, this prediction seems to be contrary to the observation that instances of 'XP-XP structures' seem to be abundant in natural languages, and to be falsified by simple sentences like [[the man] [kissed the girl]]. Then, at first glance, the simplest hypotheses regarding EFs cannot be sustained.

Interestingly, Chomsky (2007a) in effect avoids this 'XP-XP' problem by making recourse to labeling/projection. He notes,

> "If an element Z (lexical or constructed) enters into further computations, then some information about it is relevant to this option: at the very least, a property that states that Z can be merged [i.e., an EF, HN], but presumably more, it is commonly assumed. The optimal assumption is that this information is provided by a designated minimal element of Z, a lexical item W (Z itself, if it is an LI), which is detectable by a simple algorithm; the *label* of Z, the head projected in X-bar theories." (*ibid*, 9)

Chomsky claims here that there is a 'simple algorithm', called the labeling algorithm in Chomsky (2008), that can identify for an SO Z a designated minimal LI within Z that provides for Z, among other things, the property that states that Z can be merged with some SO, namely the EF with the properties (9) and (10). Simply put, the hypothesis put forward by Chomsky is that the EF is projected from designated LIs to phrasal SOs.

(16)   The EF for a phrasal/non-LI SO is projected from a designated LI (i.e., the label) of that SO.

(17) $\{W_e, SO\} \rightarrow \{W_e, SO\}_e$

Chomsky regards (16) as necessary if one would like to make room for XP-XP merger applying in narrow syntax, while sticking to the hypothesis that the EF is a property only of LIs. As soon as one adopts this hypothesis, however, one must posit some algorithm of projection/feature-percolation, hence departing from the SMT. In effect, Chomsky's adoption of (16) can be regarded as a residue of earlier theories of labeled phrase structure, constituting a major obstacle on the road to achieving a truly projection-free syntax.

3.2.2   Phase Cyclicity and *{XP, XP}

Instead of admitting (16) into the theory of syntax, however, I will claim in the rest of the present chapter that the H-α schema (14) (or *{XP, XP} (15)) can nevertheless be sustained in projection-free syntax, if we adopt the conclusion reached in Chapter 2 that syntactic derivation proceeds cyclically phase by phase (Chomsky 2001, 2004, 2008; see also Uriagereka 1999). Stated in the now familiar terminology introduced in the previous chapter, a *phase* is an SO to which the operation called *Transfer* applies upon its completion. To recapitulate, Transfer is basically the composite of Interpret and Spell-Out, applying in simultaneity with copy-formation operations (IM and Agree), and it cyclically strips off the interior of the phase, sending it to the mappings to SEM and PHON.

(18)   *Transfer*:

Applied to an SO Σ, Transfer identifies copies within Σ, and subjects the largest possible convergent term Σ' of Σ (the interior of Σ) to Interpret and Spell-Out, eliminating it from the derivational workspace.

As we discussed, Transfer is arguably a necessary component of minimal syntax, arising as a result of strict adherence to the conditions imposed by the IC, the NTC, and FI. Importantly, I provided various arguments for the hypothesis that the phase-interior subjected to Transfer

will be literally 'forgotten' (as remarked already by Chomsky 2008:143), i.e., eliminated from the derivational workspace. By doing so, Transfer has the effect of periodically reducing the domain in the active workspace by subjecting the interior of a phase to Interpret and Spell-Out. For example, if a phase head X takes a phrase YP as its sister, Transfer eliminates YP from the derivational workspace at the completion of the phase headed by X, and subjects it to Interpret and Spell-Out.

(19)   a. $\longmapsto$ b.
                X_e YP   Transfer X_e
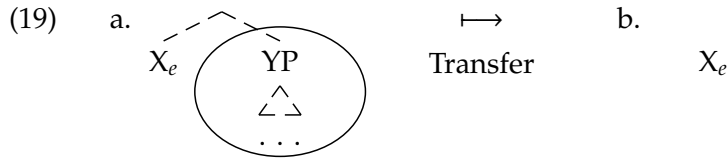
As briefly remarked at the end of §2.5.1, I put forward the hypothesis that Transfer subjects to Interpret and Spell-Out not only YP (and elements within it) but also all the syntactic relations established with regard to it, maximally respecting the effect of periodic reduction of computational loads. I specifically assume that the information subjected to Transfer includes the 'constitute' relation that YP has established with X by forming {X, YP} (the definition of 'constitute', (15) of Chapter 2, is reproduced here).

(20) α and β *constitute* γ (or conversely, γ *consists of* α and β) iff γ is the output of Merge(α, β).
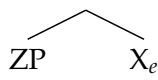
Thus, after Transfer applying to a phase {X, YP}, only the phase head X itself will be left for further computation. The situation will be different in cases where the X-phase contains one or more specifiers to its edge, a possibility to which I will return later.

A question arises as to how different this resultant 'bare LI' X is before and after the application of Transfer. This is an open question, but it is clear at least that such an LI is still syntactically accessible as a matter of fact. In particular, it should be able to constitute further input to Merge even after Transfer. This state of affairs can be naturally captured by adopting Chomsky's (2007a, 2008) hypothesis that the EF of LIs is uniformly undeletable

throughout the derivation in syntax (Chomsky 2007a:11).

(21)   The EF is undeletable throughout the derivation in syntax for all LIs.

Thus, after Transfer, nothing precludes X from utilizing its EF again to be merged to some

SO, say ZP, in a way that still conforms to the H-α schema.

(19)   c. X_e $\longmapsto$ d.

                       Merge with ZP   ZP X_e

In the bare phrase structure framework where *complement* and *spec(ifier)* mean nothing

more than *first-merged* and *later-merged* (Chomsky 1994 *et seq.*; see Chomsky 2007a:11), this

'second-merged' ZP counts as a specifier of X, informally speaking. It appears that notions

like 'complement' and 'specifier' play little role in syntax, but I will sometimes use these

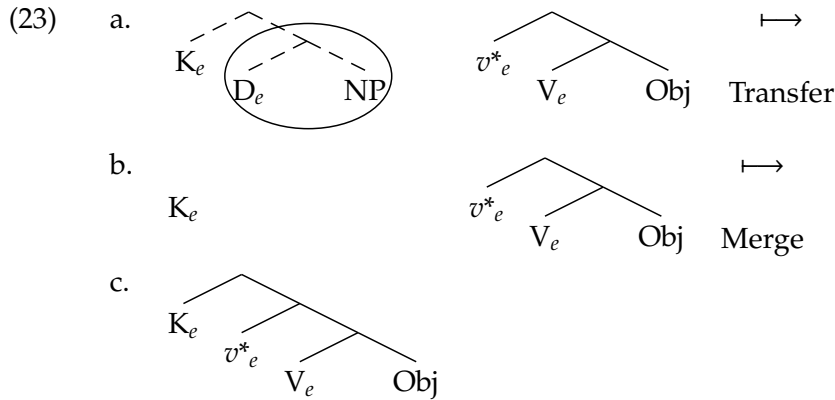terms in this chapter mainly for expository convenience.

    Given these considerations, I will assume that a phase head X, upon completion of

X's phasal computation, can still be merged with some other SO, due to the undeletability

of EFs, and that this in effect provides room for apparent XP-XP merger.[3]

    Consider, e.g., the merger of the external argument to the edge of v^*, a typical

instance of apparent 'XP-XP' external merger. For reasons to be discussed, I will adopt

the hypothesis that nominals are phases headed by the functional category K(ase) with an

unvalued Case-feature (Bittner and Hale 1996a,b, Neeleman and Weerman 1999, Asbury

2008, Caha 2009 and references cited therein).

---

[3]The idea that Transferred phases count as atomic elements has been proposed in various places. For example, Uriagereka (1999) and Nunes and Uriagereka (2000) propose that Transfer/Spell-Out has the effect of eliminating all but the topmost label of an XP from the derivational workspace. This proposal been adopted by, e.g., Sheehan (in press) among others. However, it should be clear from the present discussion that I am here explicitly departing from their label-based conception of the 'atomization' effect of Transfer. Instead of making recourse to labeling/projection, I am claiming here that the relevant atomizing effect arises simply as a result of eliminating not only the phase-interior XP but also the 'constitute' relation it has established with the phase head H by forming {H, XP}.

(22)

$$K_e \quad\quad +\quad\quad v^*_e$$
$$D_e \quad NP \quad\quad\quad V_e \quad Obj$$
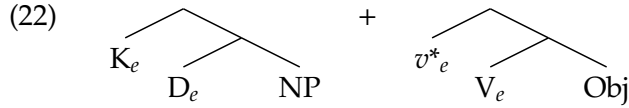
Such an instance of EM as such would be precluded by the H-α schema. However, if syntax

can make use of Transfer and reduce at least one of the 'XPs' to a simplex LI, the external

merger can be achieved without violating the H-α schema. For example, if we adopt the

assumption that K can constitute its own phase (see (43) of Chapter 2 for a more detailed

description of the K-phase), Transfer applying at the completion of the K-phase strips off

all the structural information related to {D, NP} to the interfaces, including the 'constitute'

relation that it holds to {K, {D, NP}} with K. After the Transfer of {D, NP}, the K-phase in

effect becomes a bare LI K and can be merged to {v^*, {V, Obj}} while still conforming to the

H-α schema. This derivation is sketched in (23).

(23)   a.

$$K_e \quad\quad\quad\quad v^*_e \quad\quad\quad\quad \longmapsto$$
$$D_e \quad\quad NP \quad\quad V_e \quad Obj \quad \text{Transfer}$$

   b.

$$K_e \quad\quad\quad\quad v^*_e \quad\quad\quad\quad \longmapsto$$
$$V_e \quad\quad Obj \quad \text{Merge}$$

   c.

$$K_e$$
$$v^*_e$$
$$V_e \quad\quad Obj$$

It should be noted that Transfer of K's complement does not always ensure the

reduction of the K-phase to a simplex LI. For example, if some operator Op internal to the

nominal undergoes successive cyclic A'-movement to the edge of K, resulting in a syntactic

object Σ = {Op, {K, [$_{YP}$ ... t_{Op}...]}} (with t_{Op} a copy of Op), Transfer of K's complement YP

only reduces Σ to {Op, K}, which is still phrasal.

(24)

$$\text{Op} \quad\quad\quad\quad\quad\quad \longmapsto \quad\quad \text{Op} \quad K_e$$
$$K_e \quad\quad\quad\quad\quad \text{Transfer}$$
$$... t_{Op} ...$$
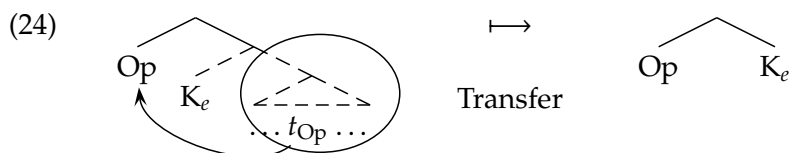
However, if we adopt the now standard assumption that v^*_e with a transitive verb is a phase head, and has an ability to Transfer its complement (Chomsky 2000, 2001, 2008), this Transfer can strip off all the structural information related to its complement {V, Obj} to the interfaces, including the one that it is merged with v^*. Thereby, it reduces the v^*-phase to a bare LI v^* that can utilize its EF for further application of Merge, say one with the phrase in (24).

(25)   a.

       Op K_e v^*_e V_e Obj   Transfer

   b.

       Op K_e v^*_e Merge

   c.

              v^*_e

       Op K_e
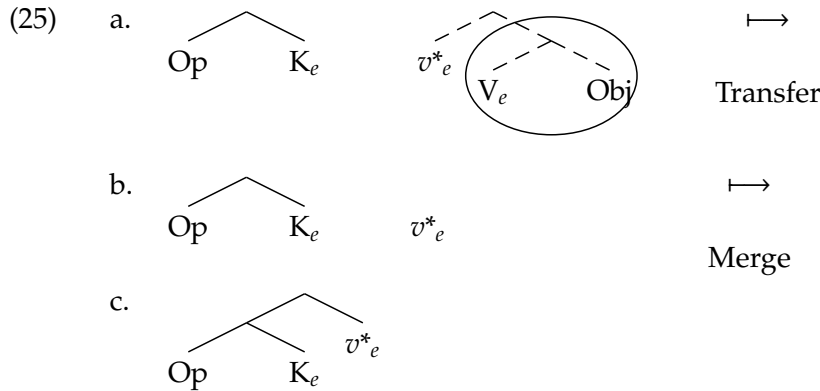
In (25), Transfer of {V, Obj} by v^* (25a) in effect enables v^* to be merged with the subject KP.
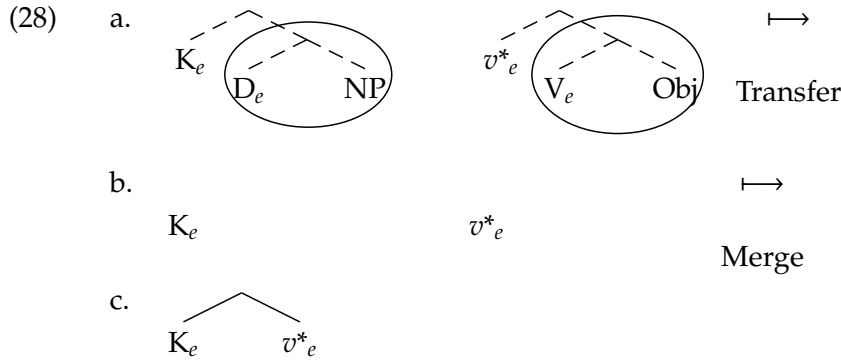
      Incidentally, I refrained from the characterization of v^*-phase as the domain of fully realized argument structure, or of 'propositionality'. As we discussed in §2.5.2, the only significant condition for phasehood is the convergence of its interior, which simply follows from the overarching principle of Full Interpretation (FI):

(26)   An SO Σ can be a phase only if the interior of Σ is convergent (i.e., containing no unvalued features).

(27)   ***Full Interpretation (FI)*** :

      Every constituent of SEM and PHON contributes to interpretation.

I will specifically assume that the external argument can be merged with $v*$ and receive the agent/causer θ-role from $v*$ before or after Transfer has applied to the complement of $v*$.[4] Thus, it should also be possible that both of the phases in (22) undergo reduction by cyclic Transfer prior to the application of EM.

(28)   a.

   K_e D_e NP $v*_e$ V_e Obj   Transfer

   b.

   K_e $v*_e$ $\longmapsto$

                                    Merge

   c.

   K_e $v*_e$

In a nutshell, apparent cases of XP-XP (external) merger are still susceptible to derivations complying with the H-α schema, as long as at least one of the two XPs is a phase without an edge and thus can be reduced by Transfer to a simplex LI. This points to a more general conclusion that cyclic Transfer is crucially in service of recursive structure-embedding by Merge.

### 3.2.3   Internal Merge and the H-α Schema

In bare phrase structure, the movement transformation is reduced to IM, which creates a new occurrence of the moving element at the edge of the target phrase. IM leaves a copy of the former in its original position, thus it yields the 'copy theory of movement' (Chomsky 1993, 1995). IM comes as free as EM, since only stipulations can preclude Merge from taking as its input either two independently constructed SOs (EM) or two SOs one of which is contained within the other (IM) (Chomsky 2004, 2007a, 2008). Our discussion so far was restricted to cases of EM, but the H-α schema should apply to cases of IM as well.

Consider a case of internal merger of β to the edge of α, as schematized in (29).

---

[4]See Marantz (1984) for the idea that external θ-roles are assigned by VP.

(29)
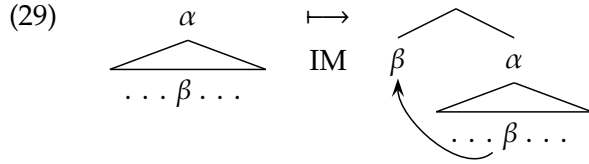
Here, α contains an original occurrence of β, and the application of IM creates another occurrence of β at the edge of α, leaving the copy of β in its original position. How can such an application of IM comply with the H-α schema? Crucially, note that α here is by definition a phrasal/non-LI SO, given the very fact that it contains an occurrence of another SO, namely β. Then, it follows from the H-α schema that the moving element, namely β, must always be an LI, since instances of XP-XP merger are precluded by the H-α schema as we saw above. Thus, any instance of IM must actually take the following form, where β is restricted to an LI (H).

(30)
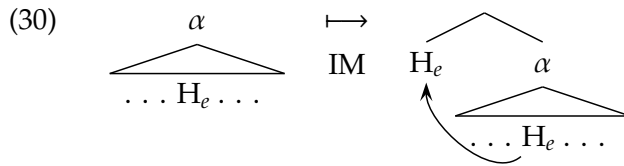

We are thus forced to conclude (31):

(31)   Only LIs can undergo IM.

Prima facie, any instance of 'XP'-movement would seem to falsify (31). However, recall the conclusion from the previous section that apparent cases of XP-XP merger are still compliant with the H-α schema, as long as either one of the two XPs can constitute its own phase without an edge, utilizing cyclic Transfer to render the phase head an LI with an EF. The same should hold for cases of IM, too.

Thus, for example, consider a case of A-movement in (32a), where apparent 'XP-XP' internal merger in (32b) applies. Here and henceforth, I will sometimes represent the relevant copy-formations by IM in a familiar traditional notation with traces and indices, but it should be understood throughout that we resort to these technical notations just

for expository convenience. Recall the discussion from Chapter 2 that no such devices, deviations from inclusiveness and no-tampering, are necessary in the theory of phases under development.

(32)   a.   [*The boy*]$_i$ t_i hit Mary.

b.
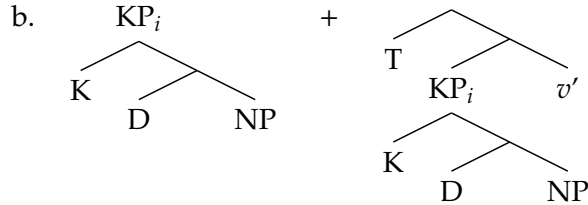
Although XP-XP merger is excluded by the H-α schema, note that the moving constituent is by assumption headed by a phase head K. Suppose that as early as the completion of the K-phase, the complement of K gets Transferred, rendering the subject KP a bare LI K. Then, what T attracts in the later derivational step is just a bare LI K, and thus it can move to the 'EPP' position while still conforming to the H-α schema. The relevant derivation is sketched below.

(33)   a.   Transfer of {D, NP} at the completion of the K-phase.

b.   Recursive Merge builds T'.   c.   IM

I propose that cases of A-movement are actually analyzed as involving IM of a phase-head LI K.

Note that the interior Σ of the K-phase surfaces at the moved position. I contend that Σ gets interpreted as a sister of K by CI and SM (33a), prior to the application of IM (33c),

which accounts for the fact that Σ gets pronounced in proximity to an occurrence of K (I will return to the problem of linearization in §4.4). What remains to be explained is the fact that the highest occurrence is typically chosen to be pronounced in association with Σ (see Nunes 2004 for some related discussion). We might speculate that the choice of the highest occurrence is in part motivated by reduction of the burden of processing: all but one copy of a chain is to be erased at PHON, presumably due to minimization of computation and production (cf. Chomsky 2008:146), and processing would be further eased if the highest (which is typically the leftmost) is chosen to be pronounced, provided that the problem of finding the gap(s)/trace(s) of moved elements can rely on the directionality of left-to-right processing.

How about A'-movement? Let's consider the following example, where a PP containing a WH-element undergoes WH-movement into C-Spec.

(34)   [*At which station*]$_i$ was John reading a newspaper t_i?

Such examples have been customarily analyzed in terms of feature-percolation, say of the WH-feature. According to this traditional analysis, entertained by, e.g., Nishigauchi (1990a, 1991, 1990b) among many others, the WH-feature of *which* gets percolated up to some higher phrasal node, here as high as the PP node, and as a result, C's attraction of the [WH]-feature can 'pied-pipe' the whole PP. The analysis is schematized as follows:

(35)   Traditional WH-feature percolation analysis:
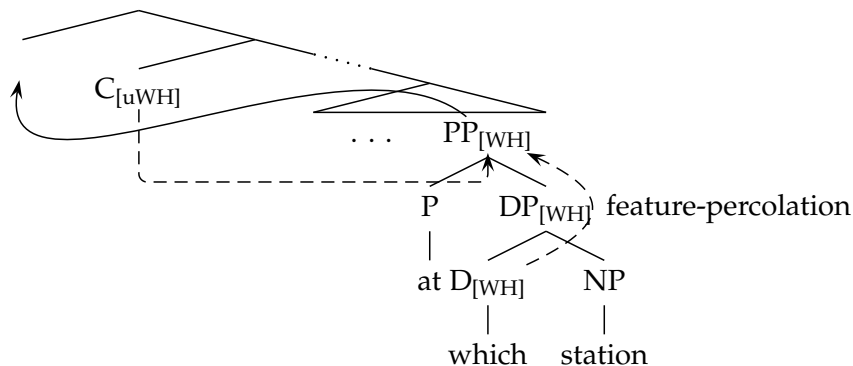
However, once we follow Chomsky (2004, 2008:140) in assuming that IM applies (as) freely (as EM) unless barred by some other constraints, the WH-feature Agree-relation to the moving WH-phrase is no longer a necessary condition for WH-pied-piping, which in turn undermines the necessity of assuming WH-feature percolation. Note further that Cable (2007, 2010) argues, providing ample data from Tlingit, that the apparent WH-pied-piping phenomenon should be reanalyzed as a movement of a category headed by a separate functional category that he calls Q (see already Watanabe 1992a,b, Hagstrom 1998, Kishimoto 2005). According to his analysis, which I will adopt in the present study, the WH-element is first licensed as a goal of probing by Q (say of the WH-feature), and QP in turn constitutes a goal for C's probing (say that of the Foc(us)-feature). The relevant derivation is schematized as follows:

(36) Cable's (2007, 2010) Q-based analysis:
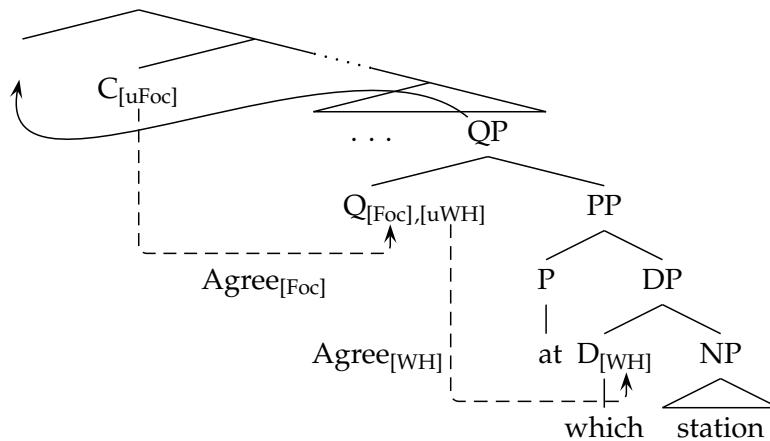
If we further assume that Q can also constitute a phase, then WH-movement is also susceptible to the account sketched above, complying with the H-α schema. Just like other phases, the complement of Q gets Transferred as soon as the Q-phase is constructed, rendering the phasal SO to a simplex LI Q. Later, C probes Q and attracts it to its specifier position. The relevant derivation is summarized as follows:

(37) a. Transfer at the completion of the Q-phase:
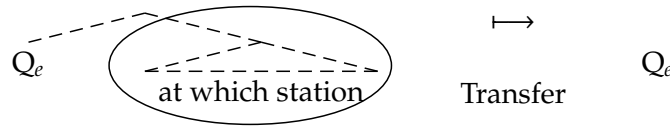
b. Recursive Merge builds the C-phase, applying IM:
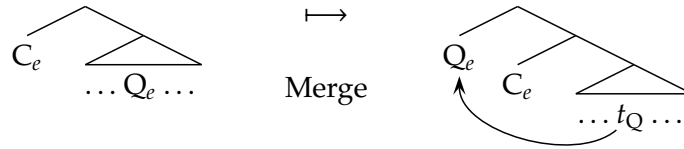
Thus, the Cablean Q-phase provides us with a suitable technical apparatus for the analysis of WH-movement in conformity with the H-α schema, as well as for the elimination of the notion of WH-feature-percolation in the theory of syntax.

However, this chapter draws a stronger conclusion than Cable's: while Cable provides a significant step toward elimination of WH-pied-piping via his Q-analysis, he did refrain from drawing a stronger conclusion that his Q-analysis can offer a step toward the elimination of the notion of XP-movement altogether. Thus, Cable notes,

> "Note that, following the definition in (7) [which says: "Pied-piping occurs when an operation that targets the features of a lexical item L applies to a phrase properly containing L^{MAX}," p.6; HN], I do not include under the rubric of "pied-piping" all instances of phrasal movement. That is, I accept as uncontroversial the existence of a mechanism of feature projection, which places the features of a head onto the projections of that head. What is at issue is any mechanism that places the features of a head onto nodes outside the projections of that head. This is a significant distinction, because feature projection is arguably indispensable, while the latter sort of devices are of little utility outside of deriving pied-piping structures." (Cable 2010:chapter 1, note 21, p.211)

Indeed, Cable's system still makes heavy use of labeling/projection to achieve XP-movement. Then, taking advantage of Cablean Q, I claim that the H-α schema and phase cyclicity naturally eliminate the notion of XP-movement altogether, in line with, but taking a more radical step than, Cable (2010).

What the discussion tells us is that as long as syntax can utilize cyclic Transfer to keep the moving element a bare LI (either a phase head or a non-phase head), syntax can still apply IM in conformity with the H-α schema (14), without departing from the

hypothesis that bare phrase structure is projection-free. Importantly, no syntactic operation of 'pied-piping' or 'XP-movement' needs to be involved in the analysis: what happens in apparent phrasal movement is actually IM of a phase-head LI, with which previously interpreted/spelled-out interior materials are tightly associated. Under the assumption that only LIs can undergo IM, we are rather forced to assume this line of approach to apparent phrasal movement in general.

In this context, it should be noted that there are cases of pied-piping of apparent 'specifiers'. The possessor of nominal phrases is a typical instance:

(38)   a.   [The student's $\varnothing_D$ father]$_i$ seems [t_i to be angry at Mary].

        b.   [Whose $\varnothing_D$ mother]$_i$ did you see t_i?

The H-α schema forces us to assume that the moving nominal phrase is actually a phase-head LI accompanying the previously transferred interior. The surface form suggests that the phase-interior that it is associated with actually contains the possessor. This fact is problematic for the traditional idea that nominals are headed by D(eterminer) and that D takes the possessor as its specifier (Brame 1981, 1982, Fukui and Speas 1986, Abney 1987): if D is the highest phase-head for a nominal, Transfer applying to the sister of D cannot affect the possessor at the edge of D.[5] This is one of the reasons why I am forced to assume, in pursuing the H-α schema, that there is a phase-head LI that occupies the highest position of all nominal arguments, which I assume is K(ase) (Bittner and Hale 1996a,b, Neeleman and Weerman 1999, Asbury 2008 and references cited therein; cf. Chomsky's 2007a n^*), and that its phase-interior domain contains the possessor as well as D. I further assume that this category constitutes the locus of Case-features as well as valued φ-features, a necessary assumption which ensures that its Case-feature can be checked via Agree applying in the

---

[5]Note that the analysis of WH-movement as in, e.g., (38b), would involve IM of a covert phase head Q as we suggested above. Cable indeed argues that the Q-based analysis eliminates the need to assume the pied-piping operation for WH-movement: Q merges on top of the nominal phrase, Transfers its complement including the possessor, and undergoes A'-movement without any problem. However, such a Q-based analysis is unavailable in examples like (38a), where even A-movement of nominal phrases 'pied-pipes' the possessor.

next phase.

The overall conclusion is that internal merger of phase-head LIs yields apparent XP-movement at PHON. Importantly, the option of undergoing ph(r)asal movement should be a privilege only of phase-head LIs, as long as we assume that cyclic Transfer is the only device to 'anchor' an XP (the interior of a phase) to its sister LI (the phase-head). This hypothesis can be adopted to derive a rather straightforward typology of possible 'XP-movement'. For instance, movement of nominals and WH-movement are proposed to be reanalyzed as internal merger of K and Q that 'anchor' their complement XPs, as we saw above. Moreover, the long noticed observation that TP as opposed to CP and *v*P strongly resists movement can be regarded as a straightforward consequence of this approach. For example, the following data from English suggest the clear distinction in syntactic independence between phasal CPs/*v*Ps and non-phasal TPs.

(39)   *English*: Topicalization

      a.   (i)   I believe that John criticized Mary.

            (ii)   [$_{CP}$ that John criticized Mary]$_i$, I believe t_i.

            (iii)   *[$_{TP}$ John criticized Mary]$_i$, I believe that t_i.

            (iv)   (... and) [$_{v\text{P}}$ criticize Mary]$_i$, I believe John (really) did t_i.

      b.   (i)   I prefer for John to take care of his mother.

            (ii)   [$_{CP}$ for John to take care of his mother]$_i$, I prefer t_i.

            (iii)   *[$_{TP}$ John to take care of his mother]$_i$, I prefer for t_i.

(40)   *English*: Cleft (Chomsky 2001: fn. 13, attributed to Luigi Rizzi)

      a.   It is [$_{CP}$ C PRO to go home (every evening)]$_i$ that John prefers e_i.

      b.   *It is [$_{TP}$ t_j to go home (every evening)]$_i$ that John$_j$ seems e_i.

(41)   *English*: Pseudocleft

      a.   What John prefers e_i is [$_{CP}$ C PRO to go home (every evening)]$_i$.

b.   *What John$_j$ seems e_i is [$_{TP}$ t_j to go home (every evening)]$_i$.

Further, the same account can be extended to data like the following, if we assume Chomsky's (2007a, 2008) analysis of ECM adapted from the earlier work by Postal (1974) and Lasnik and Saito (1991), whereby the ECM infinitival is TP and its subject raises to the VP-Spec position. TP and VP are both non-phases, and thus they resist phrasal movement.

(42)   a.   I believe John to have criticized Mary.

b.   *[$_{VP}$ John [$_{TP}$ to have criticized Mary]]$_i$, I believe t_i.

c.   *[$_{TP}$ to have criticized Mary]$_i$, I believe John t_i.

d.   *It is [$_{VP}$ John to have criticized Mary]$_i$ that I believe e_i.

e.   *It is [$_{TP}$ to have criticized Mary]$_i$ that I believe [$_{VP}$ John e_i].

f.   *What I believe e_i is [$_{VP}$ John to have criticized Mary]$_i$.

g.   *What I believe John e_i is [$_{TP}$ to have criticized Mary]$_i$.

These results are readily derived from the H-α schema in combination with the standard assumption that Cs and vs but not Ts are phase heads.[6]

3.3   The H-α Schema in Lieu of Fragmented Conditions on Extraction Domains

This section will argue that a number of empirically motivated island constraints can be derived from the H-α schema.

---

[6]Note that not only "strong" v*Ps but also "weak" vPs can be preposed, as shown in (i).

(i)   a.   [criticize Mary]$_i$ John did t_i (yesterday).
b.   [arrive at the station]$_i$ John did t_i (yesterday).
c.   [hated by Mary]$_i$ John is t_i (currently).

Thus, in current terms, v in general can Transfer its complement and move, irrespective of whether it is within transtive/unergative or unaccusative/passive constructions. This situation can be readily captured if we assume that {v, VP} can constitute a phase after the Theme KP with [uCase] is evacuated out of VP for later Nominative-Case assignment by C-T.

3.3.1 Freezing effects

As we have seen above, the H-α schema predicts that only LIs can undergo IM.

(30)

$$\alpha \quad\quad \longmapsto$$

$$\ldots H_e \ldots \quad\quad \text{IM} \quad H_e \quad\quad \alpha$$

$$\ldots H_e \ldots$$
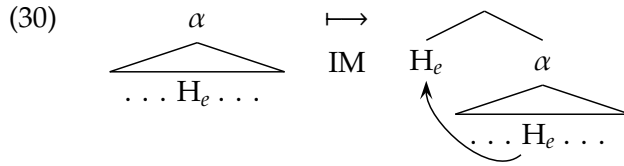
Under the H-α schema, all the cases of apparent 'XP'-movement must be analyzed as instances of internal merger of a phase-head LI that has subjected its phase-interior domain to Transfer. Let us assume with Uriagereka (1999) and Chomsky (2000, 2004, 2008) that Transferred domains are inaccessible to further syntactic computation, a constraint often referred to as the *Phase-Impenetrability Condition* (PIC):

(43) **Phase-Impenetrability Condition (PIC)** :

After Transfer applies to a phase Σ, the interior of Σ becomes inaccessible to further syntactic computation.

Then, it is predicted that all moved ph(r)ases exhibit the *freezing effect* (Culicover and Wexler 1980, Uriagereka 1999 and many others):

(44) A moved phase constitutes an island for extraction.

This is simply because all the moved ph(r)ases must have become 'bare LIs' by means of cyclic Transfer before undergoing IM. Thus, the H-α schema derives the effect of (44) for free.

There are a number of facts that support (44). For example, it has been observed by a number of researchers that an in-situ subject is not an island, whereas a raised subject is (Lasnik and Park 2003, Stepanov 2007, Gallego 2010b, Boeckx 2008a,c). (45) from Lasnik and Park (2003) illustrates this observation.

(45)   (Lasnik and Park 2003)

    a.   *Which candidate$_i$* were there [posters of t_i] all over the town?

    b.   **Which candidate$_i$* were [posters of t_i]$_j$ t_j all over the town?

If we follow Chomsky (1995, 2000), Hornstein (2009) and others in assuming that the merger of an expletive *there* into the subject position can allow the associate *posters of <which candidate>* not to be affected by IM, subextraction from this in-situ subject is predicted to be possible, as shown in (45a), which is in sharp contrast with the case of subextraction from the EPP-raised subject, as in (45b). The following examples point to the same conclusion.

(46)   a.   *Of which candidate$_i$* were there [posters t_i] all over the town?

    b.   **Of which candidate$_i$* were [posters t_i]$_j$ t_j all over the town?

(47)   a.   *Whom$_i$* were there [posters of t_i] all over the town?

    b.   **Whom$_i$* were [posters of t_i]$_j$ t_j all over the town?

(48)   a.   *Of whom$_i$* were there [posters t_i] all over the town?

    b.   **Of whom$_i$* were [posters t_i] t_j all over the town?

    A similar observation can be made for ECM constructions.

(49)   a.   *Which candidate$_i$* did you believe there to be [posters of t_i] all over the town?

    b.   **Which candidate$_i$* did you believe [posters of t_i]$_j$ to be t_j all over the town?

Essentially the same explanation is available here, too, if we follow Chomsky (2007a, 2008) in assuming that the ECMed K-phase undergoes A-movement to somewhere above the infinitival T *to* at the *v**-phase level (see already Postal 1974, Lasnik and Saito 1991, 1992 for the raising-to-object analysis of ECM). Related to this, it has been observed that an in-situ object allows extraction from within, but an ECM-raised object does not (Takahashi 1994, Bošković 1997, 2002a). Thus the following contrast:

(50) a. *Who$_i$* did you see [pictures of t_i]?

  b. **Who$_i$* did you believe [pictures of t_i]$_j$ to be t_j the cause of the riot?

This contrast suggests that the accusative KP can stay in situ in a simple transitive clause, whereas the ECMed accusative KP must move, as suggested by Chomsky (2000, 2001), Boeckx (2008b) and others.

The contrast between sentential subjects and sentential objects is another clear case in point. Thus, although the two examples in (51) are nearly synonymous, the CP moved to the Spec-T position resists A′-subextraction, as shown by the severe unacceptability of (52b).

(51) a. It is obvious [that John likes you].

  b. [that John likes you]$_i$ is t_i obvious.

(52) a. *Who$_i$* is it obvious [that John likes t_i]?

  b. **Who$_i$* is [that John likes you]$_i$ t_i obvious?

(52b) is once attributed to the so-called sentential subject constraint by Ross (1967), but readily accommodated by the freezing-based account.

In general, then, the islandhood of EPP-raised subjects emerges as a result of IM, as readily predicted by the freezing effect (44). (NB: To provide a full-fledged account of freezing effects, we will need to supply another constraint that I will refer to as **{t, t}*, to which we will return in §5.6.2).

Note that (44) makes a reverse prediction that subject nominals do not exhibit freezing effects if the language lacks obligatory EPP raising. I argue that this is a desirable prediction that can be corroborated by, e.g., data from Japanese. It has been observed that Japanese does not exhibit subject condition effects in various constructions (see Kuno 1973, Lasnik and Saito 1992, Ross 1967, Saito and Fukui 1998, Ishii 1997, and references cited

therein). The following examples are from clefts and (long distance) scrambling.[7]

(53) *Japanese*: Lack of subject condition effects in Japanese

    a.   The cleft construction

        [Op_i [John-ga [[Mary-ga t_i katta koto]-ga mondai-da to] omotteiru
             John-NOM Mary-NOM bought NML -NOM problem-CPL that think
        no]-wa sono hon$_i$-o da.
        NML-TOP that book-ACC CPL
        Lit. "It is that book$_i$ that John thinks that [that Mary bought e_i] is a problem."

    b.   scrambling

        ?nani-o$_i$ [John-ga [[Mary-ga t_i katta koto]-ga mondai-da to]
        what-ACC John-NOM Mary-NOM bought NML -NOM problem-is that
        omotteru] no?
        think   Q
        "What$_i$, John thinks that [the fact that Mary bought t_i] is a problem?"

Interestingly, Fukui (1986/1995), Kuroda (1988) and Kato (2006) among others independently provide evidence for the view that Japanese subjects can (at least optionally) stay in-situ. Ishii (1997) among others proposes that these two observations are interrelated, and specifically that the lack of subject condition effects can be attributed to the lack of obligatory EPP in this language. The phase-based account of freezing effects (44) can straightforwardly incorporate Ishii's analysis.

      Data from Romance post-verbal subject constructions provide additional evidence in favor of freezing effects (44). The following data are observed by Uriagereka (1988, 2009) for Spanish.

(54) *Spanish*: transparent postverbal subject (Uriagereka 1988:116; see also Gallego 2007, 2010)

    a.   * [de qué conferenciantes]$_i$ te   parece [que [las propuestas t_i]
        of what speakers   CL-to-you seem.3SG that the proposals

---

[7]See in particular Hoji (1990) for argument that Japanese cleft constructions with a case-marked nominal focus or PP focus involve A′-movement of an empty operator. See also Ishii (1997), Sakai (2000), Hiraiwa and Ishihara (2002, to appear) among others on different analyses of A′-movement in Japanese clefts.

me   van   a   impresionar].
CL-to-me go-3PL to impress
"Of which speakers does it seem to you that [the proposals t] will impress me."

b. (?)[de qué conferenciantes]*ᵢ* te   parece [que me   van   a
      of what speakers   CL-to-you seem.3SG that CL-to-me go-3PL to
    impresionar [las propuestas *tᵢ*]].
    impress   the proposals
    "Of which speakers does it seem to you that will impress me [the proposals

    t]."

(55)   *Spanish*: transparent postverbal subject (Uriagereka 2009)

Qué partidoᵢ te   hizo gritar [(el) que hayas   perdido *tᵢ*]
What game   you made scream   the that have.you lost
'What game has it made you scream that you lost?'

If we assume with Gallego (2010b) (who builds on Uriagereka 1988, Belletti 2001 and others) that the postverbal subject in these examples stays in the base-generated Spec-*v*P position whereas the preverbal subject moves to Spec-TP, the relevant contrast in (55) can be subsumed under the same account in terms of the freezing effect.

Note that the H-α schema predicts that the freezing effect arises only derivationally, and it does not preclude the possibility of remnant movement, an instance of IM of some ph(r)asal XP that contains a trace/copy of some smaller constituent α (den Besten and Webelhuth 1990, Müller 1996). Thus, remnant movement is readily ruled in by our analysis.

(56)   a.   [How likely t_i to win]$_j$ is John$_i$ t_j?

        b.   [Fired t_i by the company]$_j$, John$_i$ indeed was t_j.

See Müller (1996), Takano (2000), Abels (2007, 2009) among many others for further constraints on remnant movement.

3.3.2 Subject Condition Effects and Beyond

It has been widely observed that a subject XP (whether it is nominal or clausal) resists subextraction from within. Some relevant examples are given in (57).

(57)   a.   *Which article$_i$ was [that John wrote t_i] known to everyone?

          b.   *Which person$_i$ were [pictures of t_i] on sale?

A familiar description of the relevant facts, essentially since Huang's (1982) CED, holds that any subject XP constitutes an island (see also Uriagereka 1999, Nunes and Uriagereka 2000, to which we will return in Chapter 4). However, it is known that this statement is too strong. For example, Stepanov (2007) reports a number of examples from various languages that exhibit subextraction from subjects. I here provide some relevant examples from Japanese that show subextraction from the (sentential) external argument.

(58)   *Japanese*:

        a.   Cleft

           [*Op$_j$* [John-ga t_j okane-o   karita   koto]-ga Bill-o   kizutuketa
                  John-NOM   money-ACC borrowed NML -NOM Bill-ACC hurt
           no]-wa   *Mary-kara$_j$* datta.
           NML -TOP Mary-from CPL.PAST
           "It was *from Mary$_j$* [*Op$_j$* that [that John borrowed money t_j] hurt Bill]."

        b.   Scrambling

           sono hon-o$_i$ John-ga [[Mary-ga t_i katta   koto]-ga Bill-o   kizutuketa
           that book-ACC John-NOM   Mary-NOM   bought NML -NOM Bill-ACC hurt
           to]   omotteru.
           that think
           "*That book$_i$*, John thinks [that [that Mary bought t_i] hurt Bill]."

Adopting the standard assumption that the external argument (sentential or not) is base-generated in Spec-v*, the sentential subject in these examples should constitute a *bona fide* example of a noncomplement that assumes an 'escape-hatch' for successive-cyclic subextraction. Such a 'transparent' noncomplement should have no place in Uriagereka's

LCA-based account.

However, the H-α schema provides an account of CED effects that incorporates the basic tenet of Uriagereka's analysis but still accommodates the existence of transparent noncomplements. According to the H-α schema, external merger of two phrases, say XP and YP, is impossible unless at least one of the two phrases is reducible to an LI by means of cyclic Transfer. Thus, prior to the application of EM, either XP or YP must constitute a phase without an edge, so that it can be reduced to a simplex LI after Transfer. Unlike Uriagereka's analysis in terms of the LCA, the H-α schema does not specify which phrase to be reduced to an LI. Specifically, reduction of the v^*-phase to v^* as in (59a) will keep the edge of the subject accessible for further computation, allowing successive-cyclic subextraction. The derivation is sketched in (59) (here, I adopt the now widely accepted assumption, also discussed above, that nominative subjects can stay in-situ in Japanese; see Kuroda 1988, Fukui 1986/1995, 1988, Fukui and Speas 1986, Lasnik and Saito 1992 among many others):

(59)   a.

b.

c.   d.
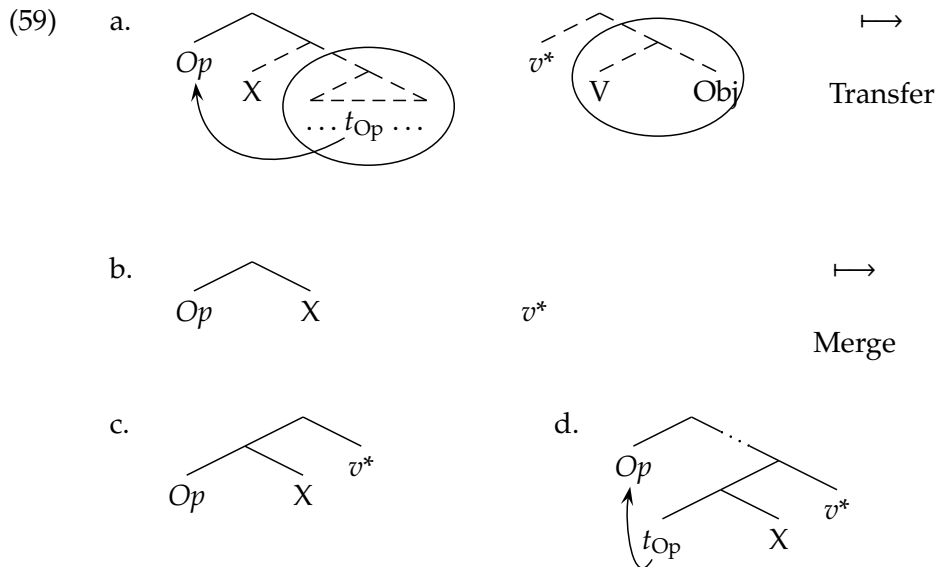
I claim that this is essentially what happens in the examples of subextraction from an external argument in (58).

This analysis holds that the reduction of v^*-phase to a simplex LI (59a) is a prerequisite for later subextraction from the subject external argument. It is specifically predicted

that the *v**-phase cannot have any edge in a derivation where subextraction from the external argument applies. Then, it is predicted that Transfer at (59a) should render the *v**-phase an island for extraction, due to the inability to assume an 'escape-hatch' for successive-cyclic movement. This prediction is schematized in (60).

(60)   * ... Op ... X ... {[$_{Subj}$ ... t_{Op} ...], [v* ... t_X ...]}

Japanese provides crucial evidence in favor of this novel prediction. (61) gives a relevant minimal pair from Japanese cleft constructions ((61a) = (58a)).

(61)   *Japanese*: Cleft and scrambling ((61a) = (58a))

    a.   [Op_j [John-ga t_j okane-o   karita   koto]-ga <u>Bill-o</u>   kizutuketa
              John-NOM   money-ACC borrowed NML -NOM Bill-ACC hurt
        no]-wa   *Mary-kara$_j$* datta.
        NML -TOP Mary-from CPL.PAST
        "It was *from Mary$_j$* [Op_j that [that John borrowed money t_j] hurt <u>Bill</u>]."

    b. *?[Op_j <u>Bill-o$_i$</u> [John-ga t_j okane-o   karita   koto]-ga t_i kizutuketa
               Bill-ACC John-NOM   money-ACC borrowed NML -NOM   hurt
        no]-wa   *Mary-kara$_j$* datta.
        NML -TOP Mary-from CPL.PAST
        "It was *from Mary$_j$* [Op_j that <u>Bill$_i$</u>, [that John borrowed money t_j] hurt t_i]."

Recall first that Japanese does allow subextraction from external arguments, as shown in (61a). Specifically, we can construct a cleft sentence (61a) from the underlying sentence comparable to (62a) by A'-movement of *Op* out of the sentential external argument. In addition, it is widely known that Japanese allows optional scrambling of an object to a sentence initial position, thus, everything else being equal, alternation between (62a) and (62b) is freely available in this language.

(62)   *Japanese*:

    a.   [John-ga Mary-kara okane-o   karita   koto]-ga <u>Bill-o</u>   kizutuketa.
         John-NOM Mary-from money-ACC borrowed NML -NOM Bill-ACC hurt
        "[That John borrowed money from Mary] hurt <u>Bill</u>."

b. <u>Bill-o$_i$</u> [John-ga Mary-kara okane-o karita koto]-ga t_i kizutuketa.
   Bill-ACC John-NOM Mary-from money-ACC borrowed NML -NOM hurt

"<u>Bill$_i$</u>, [That John borrowed money from Mary] hurt t_i."

I follow the standard assumption that object scrambling can target either an outer Spec-*v* or some higher position (say Spec-T or Spec-C) (Saito 1985, 1992, 2003, Kuroda 1988, Fukui 1986/1995, 1988, Fukui and Speas 1986, Saito and Fukui 1998, Miyagawa 1997, 2003 among many others; but see also Ueyama 1998, Saito 2005, Fukui and Kasai 2004). Application of scrambling is purely optional in most cases (Saito 1989, Saito and Fukui 1998), and it usually does not interfere with any other syntactic operation. However, the curious fact remains that scrambling of the object crossing the sentential subject is disallowed in (61b), where a null operator moves out of the sentential subject. This state of affairs is indeed predicted by the H-α schema, as in (60): subextraction from the subject clause entails that Transfer has 'atomized' the *v**-phase, allowing no edge, prior to the external merger, thus even scrambling cannot apply to elements within the *v**-phase. No previous account of the CED effect made predictions on this sort of 'complement island', which lends further support for an H-α schema-based approach.[8]

---

[8]Naoki Fukui (p.c.) provides the following examples, noting that the contrast is less obvious than the one in (61) (the reported judgments are his; I personally found the contrast still significant to an extent comparable to (61)).

(i)   *Japanese*:

a.   [*Op$_j$* [seefu-ga t_j wairo-o uketotta koto]-ga <u>ookuno kokumin-o</u> fungekisaseta
         government-NOM bribe-ACC received NML -NOM many citizens-ACC infuriated
     no]-wa *sono uyokudantai-kara$_j$* datta.
     NML -TOP that right.wing.group-from CPL.PAST
     "It was *from that right-wing group$_j$* [*Op$_j$* that [that the government received bribe t_j] infuriated <u>many citizens</u>]."

b.   ??[*Op$_j$* <u>ookuno kokumin-o$_i$</u> [seefu-ga t_j wairo-o uketotta koto]-ga t_i fungekisaseta
         many citizens-ACC government-NOM bribe-ACC received NML -NOM infuriated
     no]-wa *sono uyokudantai-kara$_j$* datta.
     NML -TOP that right.wing.group-from CPL.PAST
     "It was *from that right-wing group$_j$* [*Op$_j$* that <u>many citizens$_i$</u>, [that the government received bribe t_j] infuriated t_i]."

It is not obvious how lexical adjustment can weaken the contrast predicted by the H-α schema in some cases for some speakers. What I think is relevant to this problem is the fact, reported by Ueyama (1998), Hoji (2003),

3.3.3 Adjunct Condition Effects and Beyond

So far, I have put aside considerations regarding adjuncts, but the null hypothesis is that the H-α schema also holds for adjunction. That is, there should be no XP-XP merger, regardless of whether the merger is an instance of 'substitution' (argument-merger) or adjunction.[9] Consider, for example, adjunction of an adverbial *because*-clause in (63):

(63)   The man criticized Mary [because she failed the exam].

Adverbial clauses like the one in (63) are obviously phrasal, so external merger of an adverbial clause to the main clausal spine would count as an instance of XP-XP merger, and the H-α schema would necessitate reduction of one of the XPs to an LI. If the adverbial clause

Fukui and Kasai (2004), Agbayani et al. (to appear) and others, that there are cases where some speakers of Japanese analyze KP-/PP-scrambling as purely phonological and involving no syntactic movement (see also Saito 1985, 1989 and others that scrambling is 'semantically vacuous'). At this point we do not have a full-fledged account of when and in what environments such instances of phonological scrambling are available (see the above-cited references for relevant discussion). The difficulty in testing predictions made by the H-α schema is then to make sure that we are really looking at syntactic scrambling in relevant examples. One way to control the ambiguity is to test the relevant contrast in structures of the form in (iii), where the object clearly binds a pronominal variable within the subject from the scrambled position.

(ii)   . . . [$_{Subj}$ . . . pronoun$_{*i}$ . . .] . . . Obj$_i$ . . .

(iii)   . . . Obj$_i$ . . . [$_{Subj}$ . . . pronoun$_i$ . . .] . . . t_{Obj} . . .

Indeed, under this sort of control, the predicted contrast becomes really strong.

(iv)   *Japanese*:

a.   [Op_j [soko-no$_{*i/k}$ kogaisya-ga t_j TOB-o   uketa koto]-ga   <u>60%-izyoo-no syoosya-o</u>
          its   subsidiary-NOM   TOB-ACC made NML -NOM 60%-or.more companies-ACC
        fungekisaseta no]-wa   *IBM-kara$_j$* datta.
        infuriated   NML -TOP IBM-from CPL.PAST
        Lit. "It was *from IBM$_j$* [Op_j that [that its$_{*i/k}$ subsidiaries received takeover bids t_j] infuriated <u>more than 60% companies</u>]."

b.   *[Op_j <u>60%-izyoo-no syoosya-o$_i$</u>   [soko-no$_i$ kogaisya-ga t_j TOB-o   uketa koto]-ga t_i
          60%-or.more companies-ACC   its   subsidiary-NOM   TOB-ACC made NML -NOM
        fungekisaseta no]-wa   *IBM-kara$_j$* datta.
        infuriated   NML -TOP IBM-from CPL.PAST
        Lit. "It was *from IBM$_j$* [Op_j that <u>more than 60% companies$_i$</u>, [that its$_i$ subsidiaries received takeover bids t_j] infuriated t_i]."

[9]Chomsky (2004) proposes that these two types of Merge correspond to *set-Merge* and *pair-Merge*, respectively.

is chosen, the adverbial *because*-clause is reduced to an LI. For expository convenience, I simply assume that *because* is the relevant phase head.[10]
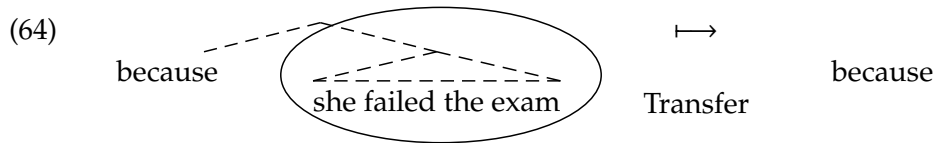
(64)

As long as the Transfer domain within the adjunct (here the complement of *because*) does not contain any uninterpretable feature, then it is a convergent domain, and it should thus be able to constitute a phase. In conformity with the H-α schema, application of cyclic Transfer (64) in effect enables the external merger of the *because*-clause. As suggested by Uriagereka (1999) and Nunes and Uriagereka (2000), the required Transfer is presumably responsible for the strong opacity of these adjuncts (see Cattell 1976, Huang 1982, Chomsky 1986a, Uriagereka 1999).

(65) a. *This is *the girl$_i$* that John failed the test [because he was thinking about t_i].

b. *I know *what$_i$* the man criticized Mary [after she said t_i].

c. *It was *this flaw$_i$* that the man criticized Mary [due to t_i].

Uriagereka and Nunes claim that the CED effect in these examples can be readily attributed to the obligatoriness of Transfer-based reduction of these adjuncts to simplex nodes. I will essentially follow their reasoning and attribute the unacceptability of the examples in (65) to the PIC.

Contra the widely accepted view that adjuncts are always strong islands (see Cattell 1976, Huang 1982, Chomsky 1986a, Uriagereka 1999), however, it has been observed that not all adjuncts exhibit CED effects (see Chomsky 1982, Boeckx 2003, Truswell 2007a,b among many others). Chomsky (1982:72) already observed some apparent counterexam-

---

[10]But see Emonds (2009) for the hypothesis that most adjuncts are indeed headed by (often covert) P (adposition).

ples to the adjunct condition, like those in (66).

(66)   (Chomsky 1982:72)

    a.   Here is the influential professor that John went to college [in order to impress

          e].

    b.   The article that I went to England [without reading *e*]

    c.   The man that I went to England [without speaking to *e*]

Truswell (2007a,b) provides a thorough survey of apparent CED-violating subextraction from adjuncts. Some examples from Truswell are reproduced in (67). A similar set of examples can be constructed in noninterrogative contexts as well, as shown in (68).

(67)   (Truswell 2007a,b)

    a.   *Which book$_i$* did John design his garden [after reading t_i]?

    b.   *What$_i$* did John arrive [whistling t_i]?

    c.   *What$_i$* did John drive Mary crazy [whistling t_i]?

    d.   *What$_i$* did John drive Mary crazy [trying to fix t_i]?

    e.   *What$_i$* are you working so hard [in order to achieve t_i]?

    f.   *Who$_i$* did John travel to England [to make a sculpture of t_i]?

    g.   *Whose attention$_i$* is John jumping up and down [in order to attract t_i]?

    h.   *What$_i$* did Christ die [in order to save us from t_i]?

    i.   *What$_i$* did you tap your nose [in order to signal t_i to Mary]?

    j.   *What$_i$* did you come in [to talk to us about t_i] today?

(68)   a.   This is *the book$_i$* that John designed the garden [after reading t_i].

    b.   It was *Hey Jude$_i$* that John arrived [whistling t_i].

   c.   I bet I know *what$_i$* John drove Mary crazy [trying to fix t_i].

    d.   It is *this goal$_i$* that students are working so hard [in order to achieve t_i].

e. Mary told me that Sam is *who$_i$* John traveled to England [to make a sculpture of t_i].

f. There's *the taxi$_i$* that John was jumping up and down [in order to attract t_i].

g. This is *the matter$_i$* that I came in [to talk to you about t_i] today.

These examples transparently show that UG should not characterize adjuncts as exceptionless islands, contra Huang (1982), Uriagereka (1999), Nunes and Uriagereka (2000), Chomsky (2004), and Stepanov (2007). In terms of the H-α schema-based account, these examples show that certain adjuncts, including bare and PP gerundives (e.g., *(after) reading the textbook*) and purpose-clauses (e.g., *in order to save us from the guilt*), are allowed not to undergo reduction to a simplex node at the point of adjunction. How can we ensure this result, while still keeping the H-α schema-based account of CED effects for the type of adjuncts like those in (65)?

I would like to propose that, unlike the adverbial clauses such as the ones headed by *because*, which I assume adjoin to positions higher than v/v^*, these transparent adjuncts are allowed to adjoin low in the clausal spine, below v/v^*. I claim that this difference in adjunction sites allows us to make the cut between the former class of 'high' adjuncts (exhibiting CED effects) and the latter class of 'low' adjuncts in question (exempt from CED effects).

First, consider cases with transparent adjuncts, e.g., the adjunction of the PP-gerundive *after reading ⟨which book⟩* in (29a).

(69) a.   b.
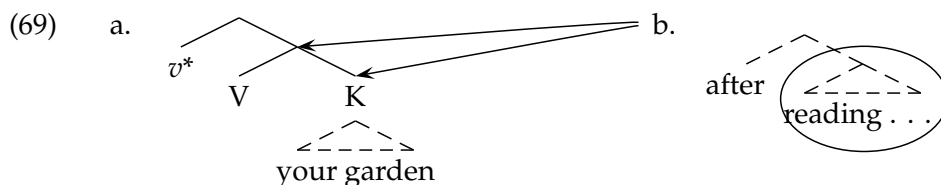
By assumption, the adverbial is a type of 'low' adjunct that can be adjoined within the domain c-commanded by v^*. For the case in question, I argue that there are two potential

adjunction sites, either to {V, K} or to K, with K being a phase-head residue of its own phase. I specifically assume that such adjuncts can adjoin *as low as in the sister of K*. Given that it is outside of the phase of K but still in the scope of *v/v**, there is no principled ground that excludes this position from being a potential adjunction site.

(69) is a case of 'XP-XP' merger, thus the H-α schema requires cyclic Transfer to reduce at least one of the XPs to a simplex LI. The Wh-phrase within the adjunct (69b) is by assumption a phase headed by covert Q (Cable 2007, 2010), thus reduced to Q after the Q-phase-level Transfer. In order for Q to undergo successive cyclic movement to the edge of interrogative C, Q should first evacuate to the edge of the ph(r)ase in (69b). As a result, a phase with Q at its edge can never be reduced by Transfer to an LI. The H-α schema therefore predicts that it should be the target of adjunction in (69a) that should be reduced to an LI. While the node {V, [K...]} is not a suitable candidate for a phase, given the unvalued Case-feature on K, the reduction of the K-phase to K via cyclic Transfer is feasible, assuming the Theme KP constitutes its own phase without an edge. The derivation is schematically shown in (70).

(70)   a.

b.

c.
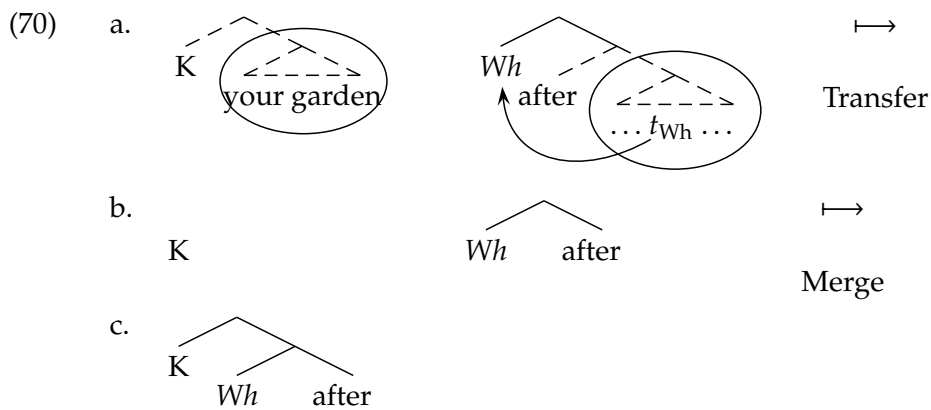
This reduction allows the edge of the adverbial to remain accessible for further operations, a desirable result.

On the other hand, I assume that high adverbials that show CED-effects, like finite clauses headed by *because* or *after*, are adjoined relatively high in the clausal spine,

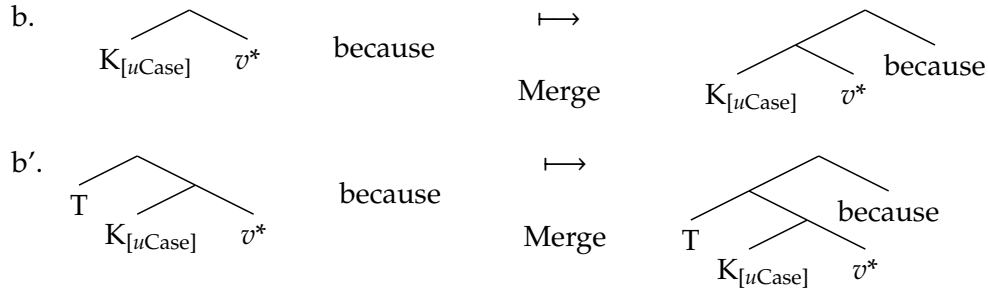specifically positions that can take scope over v/v^*.[11] This adjunction would take the form of (71):[12]

(71) a.

C
T
$K_{[u\text{Case}]}$ v^*

the man   criticize Mary

b.

because
. . .
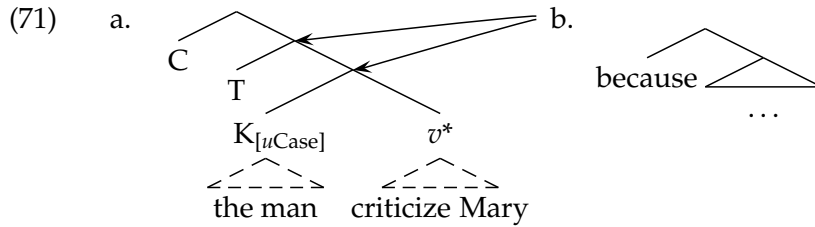
Here, I claim that reduction of the adverbial clause (71b) to an LI is necessitated by the H-α schema. The reason is the presence of the uninterpretable Case-feature on the subject K. There is evidence that Nominative Case assignment by T is contingent on the presence of C (Iatridou 1993, Watanabe 1996; Richards 2007b and Chomsky 2007a, 2008 specifically argue that the unvalued φ-features responsible for Nominative Case valuation are introduced by C and inherited onto T). The uninterpretable Case-feature of the subject K remains unchecked until the introduction of the phase head C, and so neither of the potential adjunction sites indicated by arrows in (71) can define a convergent phase. Therefore, it is impossible for these phrases to be reduced to simplex LIs by means of cyclic Transfer, and the H-α schema forces syntax to execute reduction of the adverbial (71b) to achieve external merger in (71).

(72) a.

because

she failed the exam

$\longmapsto$
Transfer

because

---
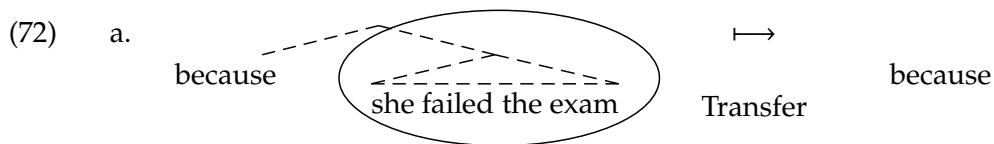
[11]The Condition C effect in examples like *He$_{*i}$ got sick {because/since/after} John$_i$ ate that fish* shows that these adjuncts are necessarily located below the subject (which is at so-called 'Spec-T') at SEM, thus it seems that adjunction of these adverbial clauses to a position that is not c-commanded by the raised subject is precluded for some independent reason. If we assume with Chomsky (2007a, 2008) that every operation except EM takes place at the phase level, it naturally follows that EM of the adverbial to T′ or v^*P necessarily precedes the introduction of C and the EPP-driven subject raising.

[12]Adjunction to the subject K in this context is precluded, because it would result in a structure like {{K, adjunct}, v^*}, locating the relevant high adjunct under the scope of v/v^*. This by definition violates the above-mentioned condition on the scope of high adjuncts.

b.

$K_{[u\text{Case}]}$ v^* because

$\longmapsto$
Merge

$K_{[u\text{Case}]}$ v^* because

b'.

T
$K_{[u\text{Case}]}$ v^* because

$\longmapsto$
Merge

T
$K_{[u\text{Case}]}$ v^* because

Consequently, the adverbial phase cannot assume any edge for successive-cyclic movement, precluding subextraction.

Moreover, I would like to point out that the H-α schema-based account makes a novel prediction that adjunction of transparent low adjuncts is unavailable for ECM-clauses. If we assume with Chomsky (1995, 2000, 2007a, 2008) that the subject of the infinitival clause receives Accusative Case from v^* in the ECM construction (see already Chomsky 1981; see also Postal 1974, Lasnik and Saito 1991 for the raising-to-object analysis of ECM), the configuration to which the low adjunct adjoins will have the following form.

(73)   a. v^*

V   'TP'

... $K_{[u\text{Case}]}$ ...

b.   adjunct

. . .
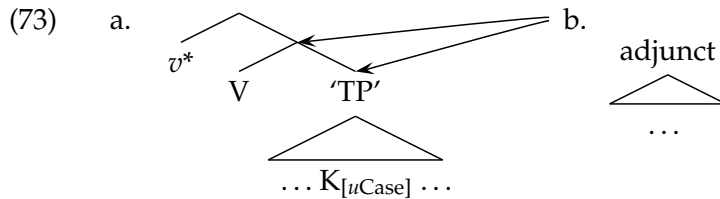
The unvalued Case-feature of K within the ECM infinitival effectively makes any of the potential adjunction sites in (73a) nonconvergent, thus the H-α schema again predicts that the low adjunct (73b) should constitute a phase without an edge (reducible to an LI by Transfer), prohibiting subextraction out of it. This prediction is indeed borne out by the unacceptability of (74a) (Bridget Samuels, p.c.):

(74)   a.   ?**Which trial**$_i$ did the DA prove [the suspect to have been at the scene of the crime] [in order to conclude t_i]?

    b.   *Which trial*$_i$ did the DA prove [that the suspect was at the scene of the crime] [in order to conclude t_i]?

    c.   *Which trial$_i$* did the DA decide [to call the suspect for psychiatric examination]

           [in order to conclude t_i]?

Crucially, note the contrast between (74a) and (74b)–(74c). On the one hand, the finite *that*-clause in (74b) is a CP that can constitute its own phase without an edge (*that* being the complementizer). Hence, its reduction to an LI can support a convergent derivation involving subextraction from the low adjunct. The same applies to the control infinitival CP [*PRO to call the suspect for psychiatric examination*] in (74c). On the other hand, an ECM-infinitival TP is different from finite and control CPs in that it cannot define its own phase due to the unvalued Case-feature on the raising object *the suspect*, thus its necessarily phrasal status precludes WH-subextraction from applying through the edge of the low adjunct. The data in (74) hence constitute another piece of strong support for the H-α schema based account of the adjunct condition effect.

      Importantly, (74a) and (74b) are virtually identical as far as their semantic interpretations are concerned, and so the contrast in this minimal pair constitutes a strong piece of evidence that the nature of the contrast is *syntactic* rather than semantic or extragrammatical. Thus, it speaks against one of the most detailed studies of the (un)availability of subextraction from adjuncts, namely Truswell (2007a,b), which puts forward a semantico-centric account of the relevant facts. Specifically, Truswell proposes the following semantic condition and argues that subextraction from adjuncts is allowed as long as it satisfies this condition.

(75)   *The Single Event Condition* (Truswell 2007a,b):

        An instance of WH-movement is acceptable only if the minimal constituent containing the head and the foot of the chain describes a single event.

He elaborates the theory of event composition to ensure the result that the minimal constituent containing the WH-chain constitutes a description of a single 'macro-event'. How-

ever, the Single Event Condition has presumably nothing to say about the contrast be-
tween (74a) and (74b), given that (74a) describes as 'single' a 'macro-event' as (74b) does.
Truswell's argument against the relevance of syntax is correspondingly undermined. At
least, then, phase cyclicity and the Single Event Condition may be principles that support
each other in accounting for the facts about subextraction from adjuncts. One may also
argue that a more interesting interpretation of these two components of explanation is to
speculate a connection between them: rather than seeing the semantic condition as an
unexplained axiom that *a posteriori* allows subextraction from adjuncts, we may rather see
the phase-based successive cyclicity, independently motivated by the principle of mini-
mal computation, as the source of semantic coherency ('single-eventhood') of the relevant
syntactic domains.[13]

3.3.4   Further Generalization on Phrase Intervention

We saw in the previous sections how H-α schema-based phase cyclicity can pro-
vide a uniform account of the freezing principle, the subject condition, and the adjunct
condition. In addition, it also provides a novel account of the 'complement-island' effect
discussed in (60), as well as the cases of certain permissible subextraction from adjuncts
studied by Truswell (2007a,b) among others. Let me further point out that the prediction
made by the H-α schema for these data can be generalized to the following derivational
constraint that we may call the *Phrasal Sister Condition (PSC)*:

(76)   *Phrasal Sister Condition (PSC)*:

No syntactic operation (IM or Agree) can relate X and Y in the structure

$\ldots$ X $\ldots$ {[$_\alpha$ $\ldots$ Y $\ldots$], [$_\beta$ $\ldots$]} $\ldots$ (order irrelevant)

---

[13]See also Narita and Samuels (2009) for data that suggest the relevance of phase cyclicity to prosodic
phrasing.

where β is phrasal (not an LI or a phase which has been reduced to an LI by Transfer).

(76) can be straightforwardly deduced from the H-α schema, simply because it predicts that Merge cannot combine two phrases α and β unless Transfer reduces at least one of them to a simplex LI: if β is irreducibly phrasal, then α must be a simplex LI. This single generalization can cover all the conditions on extraction domains discussed so far. First, the freezing effect corresponds to cases where α is a moved ph(r)ase and β is the target phrase of IM. By the definition of IM, β is irreducibly phrasal in (77) by virtue of containing the original occurrence of α.

(77)
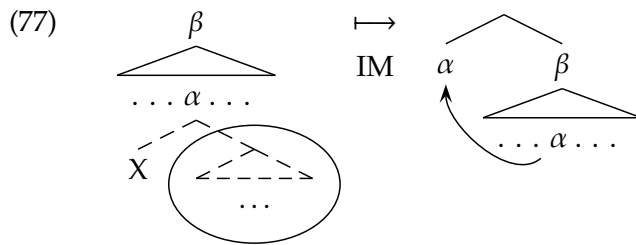
Next, in a derivation where subextraction from the in-situ external argument XP applies, the operator moves to the edge of the XP for successive cyclic movement, thus making the XP irreducibly phrasal. If β stands for this XP, then it is predicted that extraction from α is unapplicable, which derives the 'complement-island' effect discussed in §3.3.2 (see the derivation sketched in (59)).

(78)
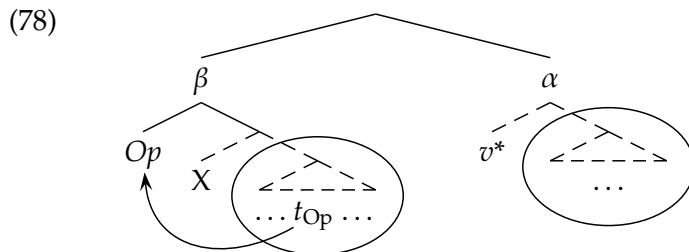
Further, the case of impermissible subextraction from adjuncts can be fit into the generalization in (76), where β stands for the main clausal spine containing an un-Case-marked K(P), and α the relevant opaque adjunct.

(79)
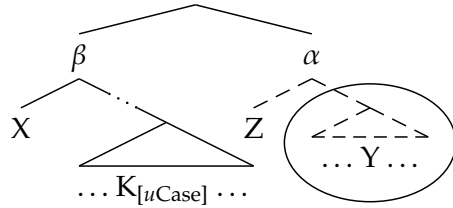
I would like to argue that the following data from raising constructions lend further support to the prediction by the PSC (76).

(80) a. *Many books$_i$* seem [to John] (still) [to be t_i in the room].

b. **Who(m)$_i$* do *many books$_j$* seem [to t_i] (still) [to be t_j in the room]?

c. [*To whom*]$_i$ do *many books$_j$* seem t_i (still) [to be t_j in the room]?

The raising predicate *seem* can optionally take an experiencer argument PP, as exemplified by *to John* in (80a). What is interesting is the fact pointed out by Groat (1999) that this *to*-PP is opaque to syntactic operations: e.g., subextraction of its complement yields ungrammaticality as shown in (80b). Notice that PP is generally not an island for extraction in English, as shown in (81), and hence previous theories that characterize English PPs as transparent (van Riemsdijk 1978, Abels 2001) would have little to say about the contrast in (80). By contrast, in order for my account to make sense of the transparency of these PPs, it is sufficient to hypothesize that they are like low adjuncts discussed in §3.3.3, in that they can be base-generated at the bottom of the *v*-phase.

(81) a. *Which city$_i$* are you heading [to t_i]?

b. *Who(m)$_i$* did you give a present [to t_i]?

c. *John$_i$* was spoken [to t_i] (by Mary).

I contend that (80) can be rather straightforwardly accounted for by the PSC (76). To begin, the experiencer PP is presumably located above the raising infinitival TP and below the raised subject KP, as shown by the binding facts in (82).

(82) a. *John$_i$* seems [to {*himself$_i$/him$_{*i}$*}] [$_{TP}$ to be t smart]

  b. *They$_i$* seems [to {*each other$_i$/them$_{*i}$*}] [$_{TP}$ to be t smart]

Thus, the raising construction behind examples like those in (80) should involve something like the structure in (83).

(83)
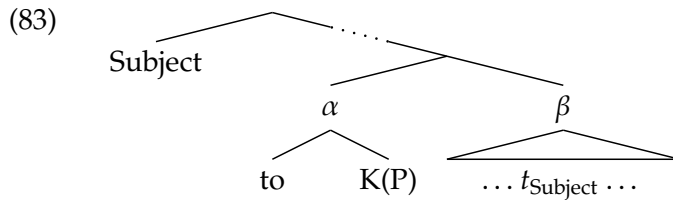
This is another instance of apparent 'XP-XP' structure, for which the H-α schema makes a now familiar prediction: either α or β should constitute an edgeless phase reducible to a simplex LI by means of Transfer. Raising of the subject out of β indicates that β does not constitute a phase and hence is irreducibly phrasal, so it should be α, the experiencer PP, that constitutes such an edgeless phase. The P-phase (or the Q-phase taking PP as its complement) can move by itself, as shown in (80c), but extraction from such a PP should be disallowed due to the PIC, yielding the unacceptability of (80b). This way, (83) constitutes another case supporting the PSC (76), where an irreducibly phrasal SO β bars movement out of its sister α.

Note that a similar observation can be made for cases where a raising verb takes a finite clause complement. The data are from Groat (1999:30-31) and Hornstein (2009:144).

(84) a. **Who(m)$_i$* does it seem [to t_i] [that it is raining]

  b. [*To whom*]$_i$ does it seem t_i [that it is raining]

These data are also susceptible to the same account based on the PSC (76), if we adopt Rosenbaum's (1970) hypothesis that the expletive *it* in such a configuration is base-generated at the edge of the embedded *that*-clause and moves to the subject position.[14]

(85)
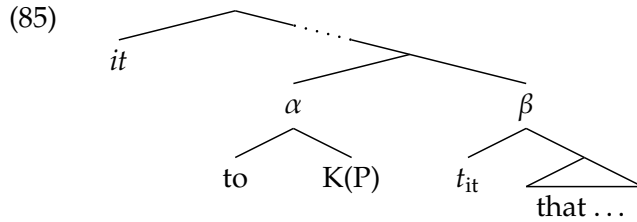
3.3.5 Locality beyond Lexicalism and Barrierhood

Boeckx and Grohmann (2007) contend that Chomsky's (2000, 2001, 2004) original notion of phases can basically be regarded as a replacement of earlier notions of 'bounding nodes' (Chomsky 1973, 1986a), as far as the account of syntactic locality is concerned. In Chomsky's phase theory with the PIC, each phase would act as an island for subextraction, unless the moving element is evacuated to its edge before the completion of that phase. Specifically, in order to allow interphasal movement to proceed in a successive cyclic fashion, Chomsky proposes that phase-head LIs are associated with some *P(eriphery)-features* (EPP-features, OCC features, etc.) that somehow allow movement from inside the phase to its edge. In Chomsky (2007a, 2008), the notion of P-feature is dispensed with, and its effect is attributed to the generalized undeletable EF. However, it is often acknowledged that Chomsky's conception of phases alone is insufficient to account for *any* island effect, once the theory loosens the barrierhood of phases by incorporating P-features or undeletable EFs. This point is made clear by, e.g., Ceplová's (2001:2-3) following remark, cited in Boeckx and Grohmann (2007:213):

> "In the current theory, 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

---

[14]In pursuing this account, a question may be raised as to how we can make sense of the contrasting acceptability of examples of the following sort:

(i)   ?*Whom$_i$* was it {told/indicated} [to t_i] [that John is a genius]?

I contend that unlike the *that*-clause in the raising construction in (84), which I claim is unambiguously analyzed as forming the abstract Rosenbaumian composite with *it* of the sort in (85), the sentence-final *that*-clause in (i) can also be marginally analyzed as an appositive clause which functions as specifying the content of the proposition denoted by *it*.

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)." (emphasis added)

In response to this worrisome aspect of phase theory, Chomsky (2001, 2004), Ceplová (2001) and many others attempt to delineate proper restrictions on the distribution of active P-features. However, their proposals have lost force in the theory of generalized EFs (Chomsky 2007a, 2008), where the notion of P-feature is eliminated and IM applies freely due to the undeletability of EFs. In this approach, the fundamental insufficiency of the phase-based account of locality observed by Ceplová thus remains unremedied, which even leads Boeckx and Grohmann (2007) to conclude that "locality doesn't offer any argument for phases."

However, what we saw above speaks directly against Boeckx and Grohmann's(2007) conclusion. It was shown that, once the theory of phases is coupled with the H-α schema, the PIC does provide a simple and unified account of CED effects that achieves even broader empirical coverage. The fundamental conclusion that we reached in the preceding discussion is that phase cycles are crucially in service of recursive structure-embedding that is compliant with the H-α schema, or more generally with the absence of projection in UG. Therefore, as long as we agree that the SMT leads us to emancipate bare phrase structure from the residual stipulation of labeling/projection borrowed from PSRs, locality of the CED sort does lend strong support to the theory of phases, *pace* Boeckx and Grohmann.

Before proceeding to the next section, I would like to briefly remark that the H-α schema-based account of locality distinguishes itself in an important respect from the earlier proposals by Ross (1967, 1986), Chomsky (1973, 1986a), Kayne (1981, 1983, 1984), Huang (1982) and many others. These proposals essentially hold that barrierhood is *lexical* in nature, and prototypically characterized by referring to features and projections of certain LIs. Since the groundbreaking work by Ross (1967, 1986), it has been dominantly

presumed that locality constraints have to do with inherent barrierhood of certain SOs: projections of certain LIs constitute islands in certain environments while others do not. Ross first provided an impressive list of fine-grained islands, each associated with lexical and categorial specifications along with positional restrictions, such as the complex NP constraint, the sentential subject condition, and so on. Chomsky's (1973) theory of subjacency was essentially an attempt to reduce the catalog of fragmented islands, where NP and S are singled out as the set of bounding nodes that restrict transformations from relating elements crossing them. Rizzi (1978) and van Riemsdijk (1978) further argue that such lexical specification of bounding nodes is parametrized across languages. Later development of bounding theory culminating with Chomsky (1986a) essentially shares the same insight, namely that certain lexical projections but not others are bounding nodes and/or barriers, with an interesting twist that the barrierhood is somewhat relaxed in certain (again lexically specified) contexts, such as proper government (Kayne 1981, 1983, 1984, Huang 1982) or L-marking (Chomsky 1986a) by certain LIs. Chomsky's (2000, 2001, 2004) P-feature-based conception of successive cyclicity discussed above can be seen as just another instance of such lexicalist approaches to barrierhood. An important addition of a different sort to bounding theories is Rizzi's (1990) *relativized minimality*, whose effect of closest attraction is relativized to lexical features ([WH]-features, φ-features, focus-features, etc.) (see also Chomsky 1995, Fukui 1999; see also Chomsky's 2000, 2001, 2008 defective intervention condition).

In contrast to the lexicalist orientation of earlier bounding theories, however, the H-α schema-based account sketched in this chapter holds that characterization of the relevant locality effect makes virtually no recourse to lexical constitution of SOs. It holds that no phrases or LIs are specified as barriers *per se*, and no relevant computation is relativized to features of LIs. Under this account, the effect of the PIC arises for an SO Σ under the condition that the sister of Σ takes a certain structural shape (namely a phrase), regardless of lexical contents. In this sense, the account is purely structural and non-lexical,

thus complying with Boeckx's (2010a) criticism of the overuse of lexical features in syntactic theory. The relevant locality constraints are now tied to the efficient cyclicity of derivation by phase, a desired result that was correctly envisaged by Chomsky (2000, 2001) and others (see in particular Uriagereka 1999, 2008, 2009, Nunes and Uriagereka 2000, and Toyoshima 1997) but not achieved until the H-α schema is supplemented to the theory of phases, as we discussed above.

3.4   Edge-features, Endocentricity and Full Interpretation

Chomsky (2007a:23) suggests that label(ing)/projection is possibly a dispensable notion, and reference to labels is a departure from the SMT. Chomsky's theory of syntax nonetheless fails to eradicate recourse to labeling/projection. Most notably, Chomsky hypothesizes that at least the EF of an LI, whose properties are summarized in (86), can percolate up to phrasal SOs by the medium of labeling/projection, as we observed in §3.2.1.

(86)   a.   The EF is the feature that permits its bearer to be merged with some SO.

        b.   The EF is a property only of LIs.

However, I argued that the theory of phase cycles adopted from Chomsky (2004, 2007a, 2008) paves the way for keeping to the minimal hypothesis that syntax assumes no mechanism of EF-projection. It was further shown that syntax without EF-projection straightforwardly derives what I have called the H-α schema, which is shown to be associated with richer empirical coverage. Thus, phase cyclicity undermines one of the core motivations behind labeling/projection, namely the percolation of EFs. Notice further that phase theory in tandem with the H-α schema eliminates the notion of XP-movement and pied-piping altogether. Thereby, it undermines another major function of labeling/projection, namely delineation of maximal projections (or pied-piping domains) susceptible to syntactic operations such as internal Merge. The feasibility of eliminating labeling/projection is

correspondingly enhanced.

As noted at the beginning of this chapter, the approach to projection-free syntax is in part motivated by the fact that various accounts are proposed to capture the effect of endocentricity of phrase structure without recourse to labeling/projection. Among others, Chomsky (p.c., lectures at MIT in fall 2010) proposes that the mechanism of head-detection can be radically reduced to minimal search of an LI for each phrase. The hypothesis can be stated as in (87):

(87)   ***Minimal head detection (MHD)*** :

The head of an SO Σ is the most prominent LI within Σ.

The effect of (87) is essentially that for any syntactic object {H, α}, where H is an LI and α an SO, H is the head of {H, α} (cf. the first clause of Chomsky's 2008 labeling algorithm; see also Piattelli-Palmarini et al. 2009:52ff). As noted, no notion of projection is implied in this minimal theory of endocentricity, keeping to the SMT. Minimal search of LIs involved here may be an effect of the third factor of language design (Chomsky 2005) like principles of computational efficiency, as suggested by Chomsky.

Phrases headed by *v*, C, *n*, P and so on are all interpreted differently at CI, which strongly suggests that the effect of endocentricity figures in semantic interpretation. Correspondingly, if (87) is the correct mechanism of head-detection, it should be operative at least at CI (and less clearly at SM). Indeed, Chomsky goes on to suggest that only SOs whose heads can be identified by (87) can receive legitimate interpretation at CI. Under this hypothesis, any XPs whose heads cannot be detected by (87) would violate the principle of FI:

(88)   ***Full Interpretation (FI)*** :

Every constituent of SEM and PHON contributes to interpretation.

It should be noted in this context that Chomsky still assumes that Merge can apply to two XPs (by the medium of labeling/projecting EFs onto XPs). Any SO of the form {XP, XP} would be headless according to (87), and thus it would fail to receive interpretation.[15] In this line of approach, then, the strength of FI must be correspondingly undermined.[16]

The strength of FI, as well as that of minimal head detection (87), is a topic we will further investigate in Chapter 5. However, let me note at this point that, if the H-α schema put forward in this chapter is on the right track, it follows that syntax *cannot* generate any interface-illegitimate objects of the form {XP, XP} for which (87) fails to determine heads. In regard to head detection, then, syntactic computation governed by the H-α schema is 'failure-proof' so to speak, and SOs generated thereby always satisfy FI. Correspondingly, the H-α schema's contribution to the SMT is corroborated, to the extent that FI constitutes a desirable component of optimal language design.

The desideratum of guaranteed FI-satisfaction can be regarded as a straightforward consequence of the H-α schema. This line of reasoning is in part motivated by the above-mentioned hypothesis that the H-α schema itself can be independently deduced from the properties of EFs. That said, however, it must also be admitted that the question of *why* the EF should have the properties in (86) is left open in this approach for the time being. This problem is especially burdensome, for the EF as formulated in (86) exhibits quite peculiar properties, distinct from all other features of LIs, as pointed out by Fukui (forthcoming) and Boeckx (2010b).

As a way to address this issue, let me also note that, alternatively, we can instead entertain a reverse causal analysis of the EF and FI. Specifically, we may hypothesize that the H-α schema and the related properties of EFs are deduced from the cooperation of

---

[15]Chomsky actually regards this as a desirable result, noting typical XP-XP structures like those involving intermediate copies of successive cyclic WH-movement and EPP-raised subjects as instances of SOs that appear to receive no interpretation at CI.

[16]Chomsky proposes that traces of movement are invisible for head-detection, thus an XP-XP structure can be salvaged for head-detection if one of the XPs moves.

FI and the interface requirement that the head of each phrase be detected by minimal search (87) at CI. That is, we may also postulate that it is rather FI in its full force that compels syntax to obey the H-α schema, and further strongly configures the properties of EFs associated with LIs. Thus, the interplay of the three factors is open to bidirectional interpretation as in (89):

(89)   Constant satisfaction of FI $\rightleftharpoons$ the H-α schema $\rightleftharpoons$ properties of EFs

I will leave the choice between these two theories of the H-α schema open for future research. However, note that both of these approaches share the same conclusion that UG should not assume any mechanism of projection: first, the theory of EFs requires that the EF must not project, since otherwise we cannot rule out XP-XP merger. If phrases are labeled by projected features of head LIs, the absence of EF-projection would be at best mysterious. Moreover, the minimal search-based account of head detection makes sense only when labeling by projection is *not* the mechanism used to account for endocentricity. Both approaches thus point to the fundamental hypothesis that bare phrase structure is projection-free, which is what we have been arguing for throughout the present chapter.

Further consequences of the approach suggested in this section, based on the minimal search-based definition of endocentricity (87) and FI (27), will be explored in Chapter 5.

3.5   Further Consequences on the Problem of Learnability

We saw that the postulation of the H-α schema derives a number of predictions that receive strong empirical support. Before concluding the present chapter, I would like to discuss some further consequences of the H-α schema in the acquisition of the Lexicon.

Throughout the present chapter, I have used the term 'lexical item' (LI) to refer to the 'atomic' elements of computation that are stored in the mental lexicon and bear

some property that allows them to be subjected to Merge, a feature that is called an edge-feature (EF). Each LI may further contain, in addition to the EF, some formal features that differentiate one LI from the other in their functions and distributions in syntactic derivation, as well as some other intrinsic features that can contribute to interpretation either at CI or SM (semantic and phonetic features). That LIs have EFs follows rather straightforwardly in this line of reasoning, but that said, we really don't know what the internal composition of LIs can be. How 'complex' are LIs? Does UG specify any constraint on the possible compositions of LIs? How can the composition of each LI be learned/acquired by the child through experience?

Opinions vary as to what the smallest unit of linguistic computation is, or where the computation 'bottoms out'. Linguists have proposed all sorts of different conceptions of the lexicon, but there is no sign that the controversy will ever find a point of agreement. However, despite linguists' continuous failure to pin down a reasonable and agreeable set of universal linguistic primitives, human infants still acquire the lexicon of one or another I-language with remarkable speed and uniformity, a familiar poverty-of-the-stimulus fact. How could this ever be possible? If there is indeed some set of primitives that are readily accessible to infants, why are trained adult linguists bound to fail to find the slightest trace of them?

The H-α schema might hint at a clue to this learnability problem. Consider, for example, recent experimental results discussed in Yang and Gambell's work (see Yang 2002, 2004, Gambell and Yang 2003 among others), according to which the general mechanism of statistical data analysis (presumably a constituent of the third factor of FL design) provides a reasonable first-cut segmentation of words in primary linguistic data, when it works in tandem with the principle (presumably determined by UG) that each phonological word bears a single primary accent. Given the relative ease of detecting phonological words in primary data (say, the three phonological words in /ðəˈbɔɪzˈkɪstəˈgɚl/, *the-bóys kíssed a-gírl*), it seems reasonable to suppose that children acquire these readily detectable

units as the first provisional candidates for LIs. From there, the H-α schema provides a preliminary analysis of sentential structures comprised of these words, synthesizing them in accordance with the H-α schema (e.g., {*the-bóys*, {*kíssed*, *a-gírl*}} or the like). In addition to such a 'bottom-to-top' application of the H-α schema-based structural analysis, the H-α schema might also give a clue to a 'top-down' decomposition of words to smaller units. For example, the phonological word *the-bóys* may be eventually analyzed as {the, bóys}, further to {the, {boy, s}}, and even to {the, {-s, {n, $\sqrt{boy}$}}} under the guidance of the H-α schema. Each of such H-α schema-based reanalyses of phonological word structures may lead to a corresponding revision of the list of provisionally analyzed LIs. For example, a learner of English may start with the provisional list of unanalyzed phonological words as his first lexicon, say {*the-bóys*, *kíssed*, *a-gírl*, ... }, but continuous revisions will be made to this list as the acquisition proceeds.

Various cues from experience (distributional or semantic) may be taken as evidence for such decomposition. Among other things, the H-α schema predicts that any instance of a movable element should involve a topmost phase-head LI, sometimes covert: for example, provided with the H-α schema, the distribution of nominals, some of which involves A- or A'-movement, might be sufficient to indicate that there is some category that appears at the topmost edge of the relevant nominal structure, namely K. Whereas learners of languages like Japanese will identify it with overt case-particles (like *o* 'ACC' and *ga* 'NOM'), learners of languages like English may assign a morpheme with zero phonetic content to this category. Arguably, such functional categories are relatively easy to detect via overt evidence, such as morphological manifestation and movement. So to speak, they satisfy what Fukui and Sakai (2003:327; see also Thráinsson 1996) calls the *visibility guideline for functional categories*, which holds that a functional category has to be visible (i.e., detectable) in the primary linguistic data, thus simplifying the problem of learnability.[17]

---

[17] Incidentally, it is not unreasonable to speculate that this sort of learnability consideration independently supports the connection between obligatoriness of A-movement in languages without case-particles like English, and optionality of A-movement in languages like Japanese with overt case-particles. See also Fukui and

In this respect, the H-α schema is both a very 'soft' and a very 'hard' constraint on the possible form of linguistic structure. It is quite 'soft' in the sense that it allows both leftward and rightward branching of the H-α structure, as in, e.g., {the, {{*n*, $\sqrt{lecture}$}, yesterday}} in English, or {{that-GEN, {yesterday-GEN, {*n*, $\sqrt{lecture}$}}}, ACC} (*so-no kinoo-no koogi-o*) in Japanese. More importantly, the analysis parses the relevant H-α branching structure without involving any extra notions like 'head-of', 'adjunct-of', or 'specifier-of'. As we have seen throughout the present chapter, these notions are just unnecessary and stipulative residue of the earlier phrase structure grammar that are simply unavailable in projection-free bare phrase structure. On the other hand, the H-α schema is so 'hard' a constraint that it disallows any vacant structural slots like unfilled 'specifiers' and 'complements', which cannot be excluded in X-bar-theoretic phrase structures. It precludes the child from speculating about any extraneous notions like labeling/projection, head-of, complement-of, specifier-of, maximal-projection-of and so on. Moreover, it strongly restricts the possible instances of 'displacement' in natural languages, and it instructs the child that any apparent instances of 'XP'-movement should involve cyclic application of Transfer, specifically triggered by (probably covert) phase-heads that encapsulate Transferred domains for phonological purposes.

This way, the H-α schema provides a strong restriction on the space of possible syntactic structures, and hence a strong bias for the initial linguistic analysis entertained by the child. The problem of explanatory adequacy is correspondingly simplified, approaching the desideratum of principled explanation of FL.

3.6 Concluding remarks

In this chapter, it was argued that reference to labels is a prominent departure from the SMT, in that it necessitates some violation of the Inclusiveness Condition (and

Takano (1998, 2000), Fukui and Sakai (2003).

presumably also the No-tampering Condition). I proposed that the theory of phase cycles adopted from Chomsky (2004, 2007a, 2008) paves the way for keeping to the minimal hypothesis that syntax assumes no mechanism of labeling/projection. It was argued that cyclic Transfer can reduce a phase to a phase-head LI, and that such a reduced phase head LI can be subjected to another application of Merge, without percolation of EFs. It was argued that phase cyclicity is crucially in service of recursive compositional structuring in projection-free syntax. It was further shown that syntax without EF-projection straightfor-wardly derives what I have called the H-α schema, which is shown to be associated with richer empirical coverage. Thus, phase cyclicity undermines one of the core motivations behind labeling/projection, namely the percolation of EFs. Notice further that phase theory in tandem with the H-α schema eliminates the notion of XP-movement and pied-piping altogether. Thereby, it undermines another major function of labeling/projection, namely delineation of maximal projections (or pied-piping domains) susceptible to syntactic op-erations such as IM. The feasibility of eliminating labeling/projection is correspondingly enhanced.

Pursuing this line of approach, this chapter attempted at eliminating the notions of labeling and projection altogether from the theory of bare phrase structure, keeping to the SMT. This chapter specifically entertained the following reasoning: if the premise in (90) holds, then the consequences in (91) are straightforwardly derived.

(90)   UG does not assume any mechanism of labeling/projection.

(91)   a.   The H-α schema (14) holds for all instances of Merge.

        b.   Only phases can undergo 'XP'-movement (viz. movement of a phase head that has subjected its complement to Transfer).

        c.   All moved ph(r)ases constitute islands (the freezing effect).

        d.   CED effects arise for moved subjects but not for in-situ subjects.

        e.   CED effects arise for 'high' adjuncts but not for 'low' adjuncts.

      f.   The PSC (76) is derived, unifying the account of CED effects and more.

      g.   Syntax generates SOs that constantly satisfy FI with regard to MHD.

      h.   The H-α schema provides a strong analytic means for the acquisition of LIs.

To the extent that the empirical payoffs summarized in (91) meet descriptive and explanatory adequacy, the explanatory force of (90) is correspondingly corroborated. The overall conclusion, in a nutshell, is that the hypothesis of projection-free syntax not only keeps more closely to the SMT but also achieves even richer empirical coverage than theories of projection-bound syntax.
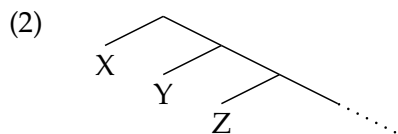
Chapter 4

Linearization of Unlabeled Structures

4.1 Introduction

I proposed in Chapter 3 that syntax does not include any projection or labeling. Specifically, it was claimed that the derivational constraint referred to as the H-α schema paves the way for eliminating the notion of projection from the theory of syntax, giving rise to unidirectional H-α branching structures like the one in (2).

(1)   ***The H-α schema***: Merge(H, α) $\rightarrow$ {H, α}.

       Merge must take at least one LI as its input.

(2)

There is no doubt that this proposal has a number of nontrivial ramifications on the proper analysis of what has been attributed to the effect of labeling. The following chapters will explore some of such ramifications of the present proposal. Specifically, the present chapter will discuss one of the fundamental operations that take place in the mapping to PHON, *linearization*, which has been predominantly presumed to make recourse to labels.

Presumably due to the modality restriction imposed by SM, the hierarchical, '2-dimensional' structure generated by syntax is 'unpronounceable'.[1] Such an unpronounceable input must be transformed to a corresponding pronounceable output of some form, satisfying SM-interface conditions.[2] Linearization refers to the necessary phonological mapping of an input hierarchical SO to a corresponding sequence of LIs readable by SM. It is interesting to observe that although past proposals on the mechanism of linearization are diverse, they seem to have reached a consensus that can be summarized in (3):

(3)   Linearization works properly for each SO only if the input is properly labeled.

To take a representative example, we will see in §4.2 that the theory of linearization by the *Linear Correspondence Axiom* (LCA) (Kayne 1994, ch.3) makes crucial recourse to labels. It was Fukui and Takano (1998) who showed that the recourse to c-command is actually eliminable from the Kaynean antisymmetry program, further clarifying the crucial relevance of labeling to linearization (their proposal is that linearization uniformly maps the head-nonhead distinction on two Merge-mates to postcedence (if α projects over β, then $\{\alpha, \beta\}$ is mapped to a string where β precedes α), yielding the universal Spec-Complement-Head word order, with apparent 'head-initial' Spec-Head-Complement order being derived by head-to-Spec movement; see already Takano 1996). Both Kayne's and Fukui and Takano's label-based theories of linearization share the goal of deriving the effect of the so-called *directionality-parameter* from invariant UG axioms. Chomsky's (1981)

---

[1]Such use of words like 'pronounce' is only metaphorical, and readers should bear in mind that PHON is usable by SM not only for articulation/externalization but also for perception. As Uriagereka (2008:8) correctly points out, by positing a single interface-level PHON for both the perception and articulation of phonetic sound, we are essentially subscribing to the theory of so-called 'analysis by synthesis'. See Halle and Stevens (1959, 1963) for earlier proposals, and see Poeppel et al. (2008) and Bever and Poeppel (2010) for contemporary discussion.

[2]Such transformations would necessarily involve violations of inclusiveness and no-tampering, but this much seems inevitable. This imperfect nature of mapping to PHON (i.e., Spell-Out) is presumably rooted in the fact that efficiency for the purpose of SM has only secondary importance in the structring of FL. This in turn seems to follow from one current line of speculation on the evolution of language, which says that FL first evolved primarily for individual thoughts (CI), only later being put to use in externalization by SM (see Chomsky 2007a, 2008, 2007c).

directionality-parameter, some version of which is adopted by a number of researchers even currently (e.g., Epstein et al. 1998, Richards 2004, Fox and Pesetsky 2005), was the first proposal that clearly expressed the crucial relevance of labeling/projections to linearization. All in all, these past proposals on linearization processes rely on the idea that the role of labeling at linearization is indispensable.[3]

That said, the absence of successful label-free proposals in the literature so far by no means proves that theorists of projection-free syntax can never approach the problem of linearization. I will try to lay out in §4.4 my own proposal regarding the issue of linearization that makes no recourse to labeling.

This chapter is organized as follows: First, I will argue in §4.2 that Kayne's (1994) Linear Correspondence Axiom (LCA) has no place in projection-free syntax, despite various researchers' attempts at revamping it (Chomsky 1994, 1995, Epstein et al. 1998, Uriagereka 1999, 2009, Moro 2000, Guimarães 2000, Nunes 2004, Jayaseelan 2008, Sheehan 2009, 2010, in press and Kayne 2009). §4.2.1 and §4.2.3 will discuss the empirical inadequacy of two different accounts of the CED effect based on the LCA, Uriagereka's (1999) and Sheehan's (2009, 2010, in press), that nevertheless exhibit certain similarity to the H-α schema proposed in the previous chapter. §4.2.4 will discuss a possible explanation of why the LCA cannot arise in UG as a mechanism of linearization. I will then turn to my own proposal in §4.4. §4.5 will conclude the chapter.

---

[3] Marc Richards (p.c.) argues that his parametrized desymmetrization mechanism (Richards 2004, 2007a), akin to directionality-parameter in spirit, achieves linearization without any recourse to (3). While I agree in all respects with his criticism against the LCA (and also Fox and Pesetsky's 2005 mechanism of cyclic linearization), his alternative linearization mechanism still makes recourse to the distinction between the predicate (his 'V') and the object (his 'O'), the precise definition of which, I suspect, will involve some mechanism that encodes asymmetry between sister SOs, which would be more or less equivalent to labeling/projection. He argues that {V, O} is a prominent instance of symmetric structure that cannot be properly dealt with by the LCA, but a similar question would be raised as to, e.g., {'S', 'v'}, where S is the external argument and 'v' the phrase headed by v/v*, and correspondingly as to why {'S', 'v'}, formally comparable to {V, O}, does not yield parametrized desymmetrization (as it seems).

4.2 Eradicating the LCA from the H-α Schema

4.2.1 Uriagereka's LCA-based H-α Schema

In Chapter 3, we reached the conclusion that the conception of EFs and the lack of feature-percolation thereof deduce the constraint that each application of Merge must take at least one LI as its input. This amounts to what we called the H-α schema, which severely restricts the possible form of SOs to {H, α}, where H is an LI and α an SO, leading to a recursive 'fractal-like' pattern of the form in (2), modulo phase-by-phase Transfer. Interestingly, this recursive H-α structure has certain resemblance to what Uriagereka (1999) calls a *Command-Unit*, which he claims is to be derived from a version of Kayne's (1994) *Linear Correspondence Axiom* (LCA).

Uriagereka (1999) is arguably one of the first attempts in the literature that seeks to resurrect the notion of 'cycles' in the framework of bare phrase structure. In a nutshell, Uriagereka's proposal is that the *Linear Correspondence Axiom* (LCA) of Kayne (1994), or some simplified version of it, imposes a strong constraint upon syntactic derivation, and that its satisfaction necessitates a cyclic Spell-Out procedure. Uriagereka argues that the requisite process of linearization via the LCA (either applying throughout the derivation (Kayne 1994) or only at PF (Chomsky 1995)) necessitates asymmetric c-command relations among LIs to be properly established during the syntactic derivation, and that syntactic derivation compliant with this LCA-based linearization requirement necessitates phase cyclicity.

To review, consider Uriagereka's exposition of the LCA, which consists of two axioms, summarized in (4).

(4)   The Linear Correspondence Axiom (LCA) (Kayne 1994; rephrased by Uriagereka 1999)

    a.   Base Step: If α c-commands β, then α precedes β.

b.   Induction Step: If γ precedes β and γ dominates α, then α precedes β.

The base step (4a) maps the asymmetric c-command relation between two LIs, say X and Y, to precedence (henceforth 'X precedes Y' is denoted by 'X→Y'). For example, (4a) maps the input representation in (5a) to a sequence of LIs, X→Y→Z→..., since X asymmetrically c-commands Y, Y asymmetrically c-commands Z ..., etc. In structures like (5a) where each instance of merger takes one LI and one non-LI (phrase) as its input, what Uriagereka calls Command-Units, then the same LI can asymmetrically c-command the rest of the LIs within the phrase. Solely the base step of the LCA (4a) suffices to determine the total linear ordering of the LIs. On the other hand, for the LCA to linearize structures like (5b), where two phrases are merged, recourse to the induction step (4b) is necessary, given that there is no direct asymmetric c-command relation established between the LIs of one phrase and those of the other. Thus, for example, X and Y in (5b) can be assigned a precedence relation by (4b), thanks to the fact that the phrase {X, WP} asymmetrically c-commands Y and dominates X.

(5)   a.   b.
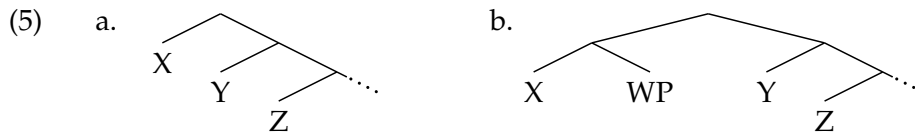
Uriagereka (1999) points out a fundamental problem with the induction step of the LCA (4b), which is that it requires some stipulation to ensure an asymmetric c-command relation between 'sister' phrases that apparently c-command each other. Kayne's (1994) original solution is to incorporate the category-segment distinction *à la* May (1985; see also Chomsky 1986a) into structural representations and stipulate that any specifier/adjunct merger splits the target category into segments, rendering the 'X'-node' (the lower segment of the category) invisible for c-command. In a similar vein, Chomsky (1995) proposes that nonminimal nonmaximal projections of LIs are marked as invisible for linearization along with the LCA, among other operations (see Sheehan 2010, in press for similar approaches).

These proposals point to the fact that some formal manipulations of labels/projections are necessary to make the induction step of the LCA (4b) work properly. However, regarding these label-based stipulations as unexplanatory, Uriagereka (1999) claims that, in order to overcome this difficulty, we should eliminate the induction step from the LCA altogether. He proposes that his simplified version of the LCA, one without the induction step (4b), imposes a rigorous constraint on syntactic derivation, namely that any structure that departs from the unidirectional H-α branching such as (5b) is unlinearizable, and as such, excluded. Therefore, the LCA severely restricts the possible mode of application of Merge, in such a way that only structures that comply with (4a), i.e., Command-Units, can be generated.

As a result, Uriagereka's simplified LCA forces each application of Merge to take one LI H and some other SO α as its input, leading to recursive H-α branching structures. Let me refer to this constraint, which Uriagereka claims is derived from his simplified LCA (4a), as the *LCA-based H-α schema*.

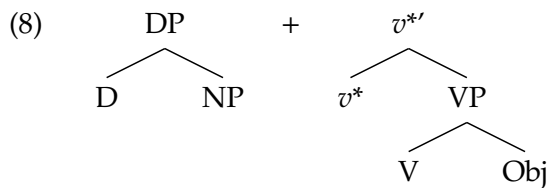(6)   ***Uriagereka's LCA-based H-α schema***:

        Merge(H, α) → {H, α}, where H precedes elements contained in α at PHON.

(6) effectively bans phrasal SOs from merging with a Command-Unit SO, so there can be no complex specifiers or adjuncts. Uriagereka argues that this is a sustainable conclusion if syntax utilizes phase cyclicity (see also Chomsky 2000, 2001, 2008), an idea akin to the proposal I made in the previous chapter. Uriagereka specifically claims that if the specifier or adjunct XP can constitute its own phase and strip off its interior domain by Spell-Out, it leaves nothing but the label of the phrase for further access in the derivational workspace.

(7)   a.   XP   ↦   b.   XP
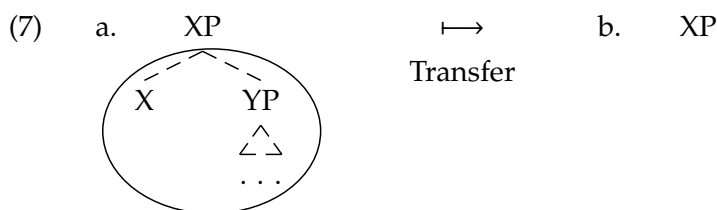
                                        Transfer

Uriagereka's work (1999, 2008, 2009) continues to assume labeled syntax, but the basic effect of Spell-Out/Transfer is supposed to be similar: it makes the phase-internal elements inaccessible to later computation, and thereby effectively reduces a complex phase to a simplex node (or a 'gigantic compound', to adopt his metaphor). It is further stipulated that such an 'atomized' node can directly c-command the LIs within its sister for the purpose of LCA-based linearization.
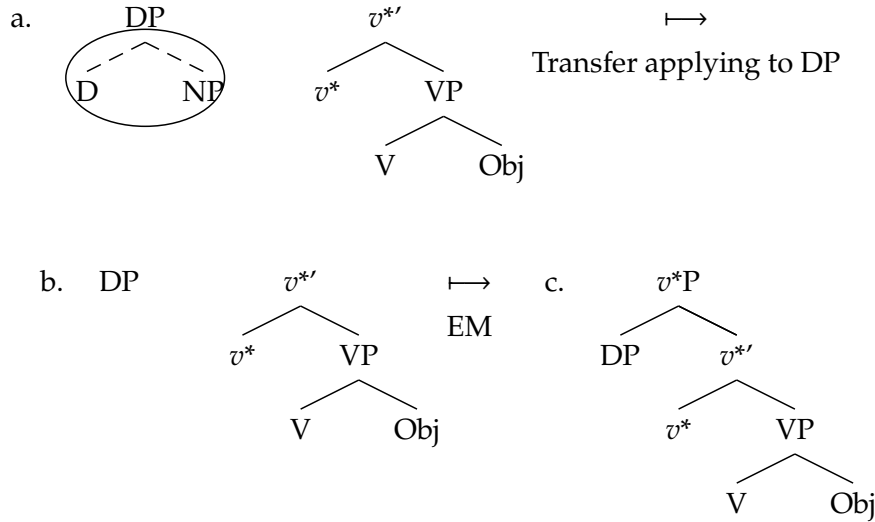
Consider, e.g., the familiar merger of the external argument to the edge of $v^{*\prime}$. To keep to Uriagereka's original hypothesis, I represent the relevant phrase-markers by means of labeled tree diagrams and the DP-analysis of nominal phrases *à la* Abney (1987) that he assumes.[4]

(8)

$$\begin{array}{ccc} \text{DP} & + & v^{*\prime} \\ \diagup\diagdown & & \diagup\diagdown \\ \text{D} \quad \text{NP} & & v^{*} \quad \text{VP} \\ & & \diagup\diagdown \\ & & \text{V} \quad \text{Obj} \end{array}$$

Such an instance of external merger would be a *prima facie* violation of the LCA-based H-α schema, because of the would-be complex 'specifier', the subject DP. However, it is assumed in Uriagereka's system that DPs (or any specifiers for that matter) can constitute their own command-units, and application of Transfer to the subject DP-phase will terminate access to all the elements dominated by DP, reducing the entire phase to a simplex node, basically the maximal label DP. It is further assumed that such a 'stranded' simplex label symbol can then be merged to the edge of v^{*} in conformity with his H-α schema. The relevant derivation is summarized in (9):

(9)   Noncomplement-reduction:

---

[4]Given that Uriagereka assumes labeled phrase structure, the empirical fact that A-movement can pied-pipe Spec-D does not force him to posit a separate functional category K. The cost is of course that Uriagereka has to assume a mechanism of labeling/projection, a departure from the SMT in the framework of bare phrase structure, as I have repeatedly argued.

a.
```
      DP                  v*'                        ⟼
   ⟨          ⟩         /     \            Transfer applying to DP
   (  D    NP  )      v*      VP
                             /   \
                            V    Obj
```

b.　DP
```
                  v*'                  ⟼      c.        v*P
                /     \                     EM        /     \
              v*      VP                            DP      v*'
                     /   \                                 /    \
                    V    Obj                             v*     VP
                                                               /   \
                                                              V    Obj
```

For ease of discussion, let me introduce some informal terminology and refer to such an application of Transfer, which reduces phrasal specifiers and adjuncts to just simplex label symbols, as *noncomplement-reduction*. For Uriagereka, noncomplement-reduction is a necessary derivational step that renders complex noncomplements compliant with his version of the LCA. Cyclic Transfer thus crucially serves the recursive embedding of noncomplements in his theory.

Assuming the equivalent of the PIC, Uriagereka claims that his LCA-based H-α schema forces syntax to execute noncomplement-reduction for all instances of "XP-XP" merger, and that this correctly deduces the effect of Huang's (1982) CED on specifiers and adjuncts (see also Nunes and Uriagereka 2000), as exemplified by (10).

(10)　　a.　*Who$_1$ did [a picture of t_1] cause the problem?

　　　　b.　*Of whom$_1$ did [a picture t_1] cause the problem?

Then, it is predicted to follow from Uriagereka's LCA-based H-α schema that every noncomplement XP (specifier or adjunct) should constitute an island for extraction.

Recall at this point that the previous chapter was devoted to advocating for the H-α schema in (11), which is totally independent of precedence relations among LIs at PHON.

(11)   ***The H-α schema***: Merge(H, α) → {H, α}.

Merge must take at least one LI as its input.

The similarity between Uriagereka's account of the CED effect and our H-α schema based account of subject and adjunct conditions is rather obvious. The two approaches share the idea that a constraint on structure-building derives the effect of Huang's (1982) CED. Specifically for instances of XP-XP merger like the one in (8), I proposed that cyclic Transfer at the K-phase level eliminates the complement of K and all the structural information associated with it, in effect reducing the K-phase to a bare LI K. The derivation is schematically shown in (12):

(12)   a.

b.   c.
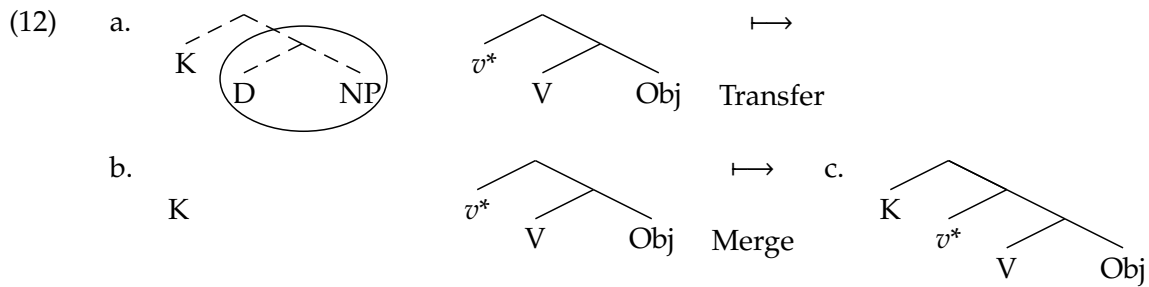
As with Uriagereka, I claimed that Transfer of the interior of the K-phase effectively renders it an island for extraction.

At this point, recall from our discussion in §3.4 that we can deduce the H-α schema without any recourse to the LCA, either from the properties of EFs (13) or, perhaps more promisingly, Minimal Head Detection (14) serving for the consistent satisfaction of Full Interpretation (15), each recapitulated here.

(13)   ***Edge-features (EF)*** (Chomsky's (2008) version):

a.   The EF is the feature that permits its bearer to be merged with some SO.

b.   The EF is a property only of LIs.

(14)   ***Minimal head detection (MHD)*** :

The head of an SO Σ is the most prominent LI within Σ.

(15)   ***Full Interpretation (FI)*** :

Every constituent of SEM and PHON contributes to interpretation.

Then, to the extent that Uriagereka's LCA-based alternative has some empirical appeal that overlaps with the LCA-free H-α schema articulated in Chapter 3, whether or not the H-α schema should be ultimately attributed to the LCA becomes an empirical question.

In the rest of this section, we will see that the H-α schema makes correct predictions regarding the generalizations achieved in the previous discussion *only if* the LCA is eliminated from the set of UG-principles. That is, the LCA is simply *incompatible* with the H-α schema, as we will see. I will then relate this important result to the broader conclusion that the LCA has no place in projection-free syntax, despite Uriagereka's (and in part Sheehan's in press) argument to the contrary.

4.2.2   The Irrelevance of the LCA to the CED Effect

It should be pointed out that the two H-α schemata (our LCA-free one and Uriagereka's LCA-based alternative) make different predictions about how syntax cyclically Transfers derivational cascades to the interfaces. Most importantly, it should be pointed out right away that, if we dissociate the H-α schema from the LCA, then not only noncomplement-reduction but also what we may informally call '*complement-reduction*', i.e. cyclic Transfer of the would-be 'complement' XP, can make an instance of XP-XP merger compliant with the H-α schema. For Uriagereka, noncomplement-reduction is the *only* option whereby 'XP-XP' merger can be in conformity with the linearization requirement imposed by the LCA. In LCA-free syntax, by contrast, complement-reduction is as effective for the satisfaction of the H-α schema as noncomplement-reduction. As for (8), complement-reduction can be applied to eliminate the phase interior domain of the v^*-phase (VP) and reduce it to a simplex phase-head LI, whose EF can then permit it to be merged with the subject KP. The relevant derivation is summarized in (16):

(16)   Complement-reduction:
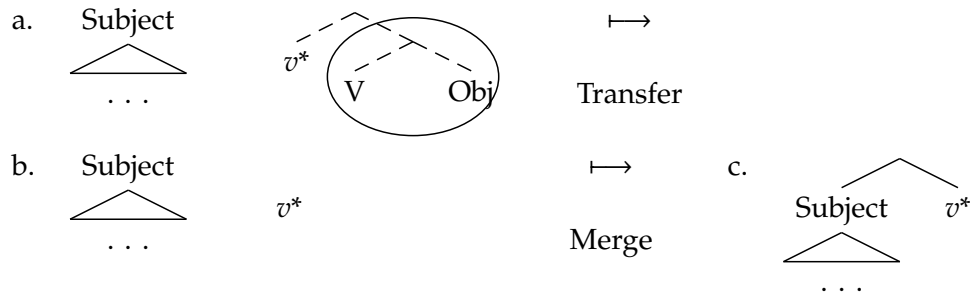
Thus, the two H-α schemata make different predictions with regard to the (un)availability of complement-reduction. Uriagereka's LCA-based H-α schema predicts noncomplement-reduction is the only option, since otherwise the requisite Spec-Head-Complement word order cannot obtain, whereas the H-α schema in LCA-free syntax freely entertains both complement- and noncomplement-reduction. We have already seen a number of cases where application of complement-reduction is crucial, which suggests that the H-α schema and its predictions concerning island effects work properly only if we eliminate the LCA from the theory of UG.

First of all, every noncomplement XP (specifier or adjunct) should constitute an island for extraction under Uriagereka's account of the CED. However, we have seen that this prediction is too strong. Specifically, the in-situ external argument provides *bona fide* counterevidence to that prediction. Here I reproduce the Japanese example that shows subextraction from a (sentential) external argument.

(17)   *Japanese*:

a.   Cleft

[Op$_j$ [John-ga t_j okane-o   karita   koto]-ga Bill-o   kizutuketa
         John-NOM   money-ACC borrowed NML -NOM Bill-ACC hurt
no]-wa   *Mary-kara$_j$* datta.
NML -TOP Mary-from CPL.PAST
"It was *from Mary$_j$* [Op$_j$ that [that John borrowed money t_j] hurt Bill]."

b.   Scrambling

sono hon-o$_i$	John-ga	[[Mary-ga	t_i katta	koto]-ga	Bill-o	kizutuketa
that book-ACC	John-NOM	Mary-NOM	bought NML	-NOM	Bill-ACC	hurt

to] omotteru.
that think
"*That book$_i$*, John thinks [that [that Mary bought t_i] hurt Bill]."

Adopting the standard assumption that the external argument (sentential or not) is base-generated in Spec-v^*, the sentential subject in these examples should constitute a *bona fide* example of a noncomplement that does not undergo Transfer prior to the subextraction of a null operator. Such a 'transparent' noncomplement should have no place in Uriagereka's LCA-based account. In contrast, the LCA-free H-α schema readily allows application of complement-reduction upon the external merger of subject XP and $v^{*\prime}$ as in (22), which will keep the former accessible for further computation, allowing subextraction of *Op*. A sketch of the relevant derivation is reproduced in (18):

(18)   a.

Transfer

b.

Merge

c.   d.
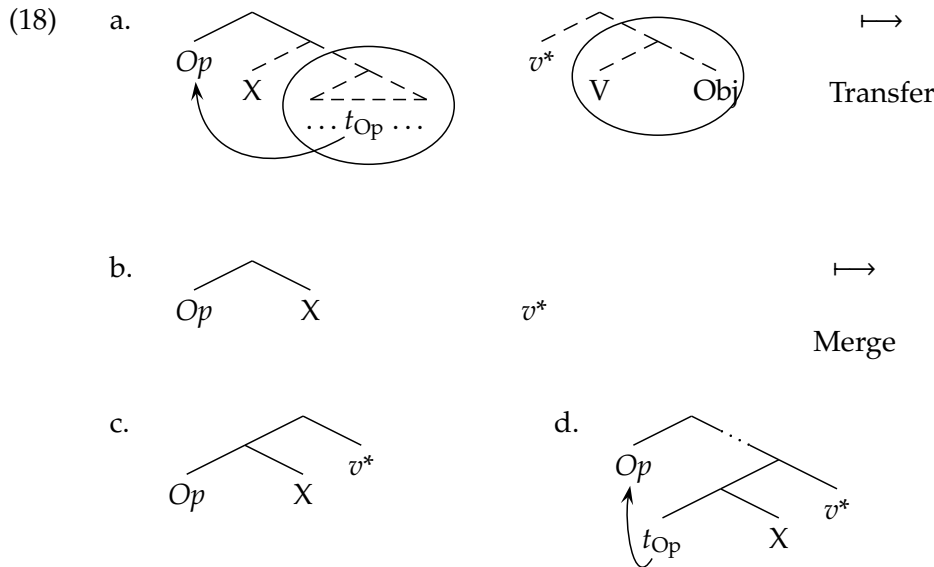
Moreover, as we have seen, it is predicted that the application of complement-reduction to the v^*-phase is a prerequisite for later subextraction from the subject external argument as in (17). Then, the other side of the same coin is the prediction that complement-reduction at (18a) renders the relevant v^*P-complement an island for extraction. This prediction is reproduced in (19).

(19)   * ... *Op* ... X ... {[$_{Subj}$ ... t_{Op} ...], [v^* ... t_X ...]}

(20) reproduces a relevant minimal pair from Japanese cleft constructions ((20a) = (17a)).

(20)   *Japanese*: Cleft and scrambling

    a.   [*Op$_j$* [John-ga t_j okane-o   karita   koto]-ga <u>Bill-o</u>   kizutuketa
            John-NOM   money-ACC borrowed NML -NOM Bill-ACC hurt
        no]-wa   *Mary-kara$_j$* datta.   (=(58a))
        NML -TOP Mary-from CPL.PAST
        "It was *from Mary$_j$* [*Op$_j$* that [that John borrowed money t_j] hurt <u>Bill</u>]."

    b. *?[*Op$_j$* <u>Bill-o$_i$</u>   [John-ga t_j okane-o   karita   koto]-ga t_i kizutuketa
           Bill-ACC John-NOM   money-ACC borrowed NML -NOM   hurt
        no]-wa   *Mary-kara$_j$* datta.
        NML -TOP Mary-from CPL.PAST
        "It was *from Mary$_j$* [*Op$_j$* that <u>Bill$_i$</u>, [that John borrowed money t_j] hurt t_i]."

This state of affairs is readily predicted by our H-α schema in LCA-free syntax: subextraction of *Op* from the subject CP entails that complement-reduction has applied to the v^*-phase at the point of external merger of the CP, thus even scrambling cannot apply to elements within the v^*-phase. By contrast, Uriagereka's LCA-based proposal never predicts any 'complement island' of this sort, which leads to the conclusion that the H-α schema dissociated from the LCA makes an empirically superior analysis of the CED effect on subjects.

Moreover, just like Uriagereka's LCA-based analysis, the LCA-free H-α schema readily predicts the freezing effect, as we have seen in Chapter 3. In bare phrase structure, the movement transformation is reduced to IM. Recall that the H-α schema predicts that only simplex LIs can undergo IM.

(21)

Here, H can be either just an LI, resulting in head-movement, or a phase head that has previously transferred its interior domain to the CI- and SM-interfaces, resulting in apparent 'XP'-movement. Specifically, according to the H-α schema, all instances of apparent phrasal movement must be analyzed as instances of IM of a phase-head LI that Transfers its phase-interior domain. The freezing effect (Ross 1974, Culicover and Wexler 1980, Culicover 1982, Chametzky 2000, Corver 2006 among many others) is thus derived from the H-α schema.

(22)   A moved phase constitutes an island for extraction.

This is simply because all the moved ph(r)ases must have become simplex nodes by means of cyclic Transfer before the application of IM.

       Note that the freezing effect can be derived from either the LCA-based or the LCA-free H-α schema, given that IM always relocates an SO into a noncomplement position. However, these two H-α schemata make different predictions about unmoved noncomplements and moved noncomplements. Suppose we have an externally merged specifier XP (say an external argument) which is to be dislocated to some higher specifier position. In Uriagereka's (1999) LCA-based system, this XP undergoes Transfer and becomes a simplex node before being externally merged to the base specifier position, because noncomplement-reduction always applies at any XP-XP merger in this approach. Thus, any noncomplement should constitute an island for extraction, regardless of whether it moves later or not. However, in the LCA-free approach sketched in Chapter 3, the option of complement-reduction can keep the relevant specifier XP untransferred at the point of EM, as we have seen above. The only restriction imposed by my LCA-free approach is that this XP must be reduced to a simplex node by means of cyclic Transfer before moving to some higher position. Thus, it predicts that the freezing effect arises only for moved noncomplements.

A contrast relevant to these predictions arises in pre- v.s. post-verbal subject KPs in Spanish (see Uriagereka 1988, Gallego 2007, 2010b for discussion). The data here show that subextraction from post-verbal external arguments is strongly preferred to subextraction from preverbal ones.

(23)   *Spanish*: transparent postverbal subject (Uriagereka 1988:116; see also Gallego 2007, 2010)

a.   * [de qué   conferenciantes]$_1$ te   parece   [que [las propuestas t_1]
         of what speakers   CL-to-you seem.3SG that the proposals
      me   van   a impresionar].
      CL-to-me go-3PL to impress
      "Of which speakers does it seem to you that [the proposals t] will impress me."

b.   (?)[de qué   conferenciantes]$_1$ te   parece [que me   van   a
         of what speakers   CL-to-you seem.3SG that CL-to-me go-3PL to
      impresionar [las propuestas t_1]].
      impress   the proposals
      "Of which speakers does it seem to you that will impress me [the proposals

      t]."

Gallego (2007, 2010b) provides arguments that the postverbal subject in these examples stays in the base-generated Spec-v* position whereas the preverbal subject moves to Spec-T (see already Uriagereka 1988, Belletti 2001). Then, the relevant contrast in (23) should be attributed to the freezing effect on the raised noncomplement KP (the external argument in (23a)), which arises only in the H-α schema in an LCA-free world. Uriagereka himself admits that examples like (24), carefully constructed in a way that clearly shows the transparency of in-situ external arguments, are marginally acceptable.

(24)   *Spanish*: transparent postverbal subject (Uriagereka 2009)

      Qué partido$_i$ te   hizo gritar [(el) que hayas   perdido t_i]
      What game   you made scream the that have.you lost
      'What game has it made you scream that you lost?'

Uriagereka notes, "My feeling hasn't changed in two decades: it is not perfect, but it is also not as bad as it would be if the subject were in its canonical subject position." (2009:ch.2) This corroborates the point that we should attribute the contrast in (23) to the derivationally-arising freezing effect, not to the inherent noncomplement status of the in-situ subject. Here too, then, it is shown that the LCA-free H-α schema provides a better prediction of the scope of the freezing effect than Uriagereka's approach.

Furthermore, the empirical inadequacy of the LCA-based H-α schema appears most robustly in the treatment of adjunction. The null hypothesis based on Uriagereka's set of assumptions is that there is no XP-XP merger, regardless of whether the merger is an instance of 'substitution' (argument-merger) or adjunction. Consider, for example, adjunction of an adverbial *because*-clause in (25):

(25)   The man criticized Mary [because she failed the exam].

Adverbial clauses like the one in (25) are obviously phrasal, so external merger of an adverbial clause and the main clausal spine would count as an instance of XP-XP merger, necessitating Transfer of either one of the XPs. If noncomplement-reduction is chosen, the adverbial *because*-clause is reduced to an LI.
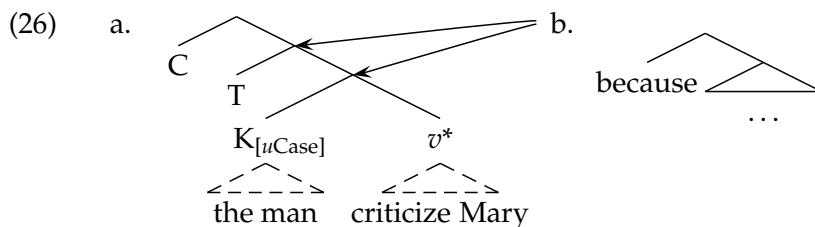
(26)   a.   b.

As long as the Transfer domain within the adjunct (here the complement of *because*) does not contain any uninterpretable features, then it is a convergent domain. It should thus be able to constitute a phase. In conformity with the H-α schema, application of noncomplement-reduction in effect enables the external merger in (26).

Noncomplement-reduction is presumably responsible for the CED effect for adjuncts (see Cattell 1976, Huang 1982, Chomsky 1986a, Uriagereka 1999), as suggested by Uriagereka (1999) and Nunes and Uriagereka (2000).

(27)   a. *This is the girl$_i$ that John failed the test [because he was thinking about t_i].

        b. *I know what$_i$ the man criticized Mary [after she said t_i].

        c. *It was this flaw$_i$ that the man criticized Mary [due to t_i].

Uriagereka and Nunes claim that the CED effect in these examples can be readily attributed to the obligatoriness of noncomplement-reduction for these adjuncts. In §3.3.3, I essentially endorsed the PIC-based approach to the relevant data, and attributed the unacceptability of the examples in (27) to noncomplement-reduction.

As we have seen, however, not all adjuncts exhibit CED effects (Chomsky 1982, Boeckx 2003, Truswell 2007a,b among many others).

(28)   (Truswell 2007a,b)

        a. *Which book$_i$* did John design his garden [after reading t_i]?

        b. *What$_i$* did John arrive [whistling t_i]?

        c. *What$_i$* did John drive Mary crazy [whistling t_i]?

        d. *What$_i$* did John drive Mary crazy [trying to fix t_i]?

        e. *What$_i$* are you working so hard [in order to achieve t_i]?

        f. *Who$_i$* did John travel to England [to make a sculpture of t_i]?

        g. *Whose attention$_i$* is John jumping up and down [in order to attract t_i]?

        h. *What$_i$* did Christ die [in order to save us from t_i]?

        i. *What$_i$* did you tap your nose [in order to signal t_i to Mary]?

        j. *What$_i$* did you come in [to talk to us about t_i] today?

(29)   a. This is *the book$_i$* that John designed the garden [after reading t_i].

b. It was *Hey Jude$_i$* that John arrived [whistling t_i].

c. I bet I know *what$_i$* John drove Mary crazy [trying to fix t_i].

d. It is *this goal$_i$* that students are working so hard [in order to achieve t_i].

e. Mary told me that Sam is *who$_i$* John traveled to England [to make a sculpture of t_i].

f. There's *the taxi$_i$* that John was jumping up and down [in order to attract t_i].

g. This is *the matter$_i$* that I came in [to talk to you about t_i] today.

These examples transparently show that the theory should not characterize adjuncts as exceptionless islands, contra Huang (1982), Uriagereka (1999), Nunes and Uriagereka (2000), Chomsky (2004), and Stepanov (2007).

In terms of the H-α schema-based account discussed in the previous chapter, these examples show that the relevant adjuncts are allowed not to undergo noncomplement-reduction at the point of adjunction. Readers are referred to §3.3.3 for the proposal that the relevant contrast between these two classes of adjuncts arises as a function of the structural heights of the adjunction sites: unlike the finite adverbial clauses like the one headed by *because*, which I propose adjoin to positions higher than *v*, these transparent 'low adjuncts' are allowed to adjoin to positions that are lower than *v*. I argued that this difference would allow us to draw the cut between the class of 'high' adjuncts (exhibiting CED effects) and the one of 'low' adjuncts (exempt from CED effects).

Now, what happens if we employ Uriagereka's LCA-based H-α schema in the discussion above? Recall that under Uriagereka's system, noncomplement-reduction is driven by the LCA-based linearization requirement at PHON. It essentially applies to reduce adjuncts to simplex nodes that can independently asymmetrically c-command the other LIs. Note that the LCA should always map these asymmetric c-command relations to precedence, thus it should be the case that any adjunct always precedes the LIs within the relevant Command-Unit it is adjoining to. That is, adjunction should always be 'to the left'.

This consequence, unavoidable if the H-α schema is tied to the LCA, is highly problematic in its empirical import. First of all, all apparent cases of 'rightward adjunction' that surface with the H(ead)-C(omplement)-A(djunct) order (as in, e.g., *John [$_H$ called] [$_C$ Mary] [$_A$ before Sue came]* in English) must be reanalyzed either as (i) involving the relevant adjunct as a *bona fide* complement, (ii) involving movement of the constituent comprising H-C over (left-adjoined) A, or (iii) involving 'XP'-movement of C over A and 'head'-movement of H over A.

(30)   a.   b.   c.
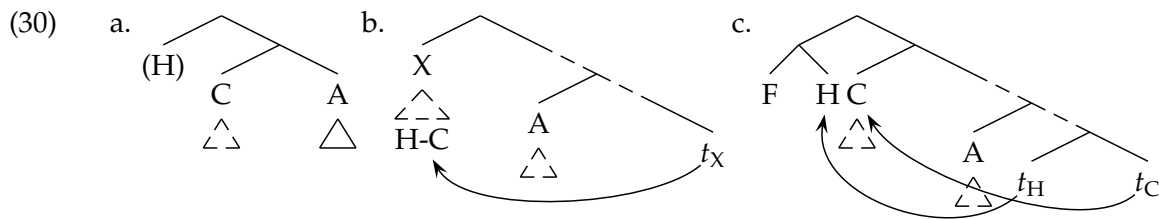
Advocates of the LCA thus bear a serious burden of proof: they must show that *all* apparent cases of rightward adjunction are best analyzed as involving either of these structural representations, and moreover that such structural analyses can be naturally acquired by children through impoverished primary linguistic data. We have never seen in the LCA literature any serious justification of such *ad hoc* reanalyses, which is already indicative of the empirically inadequate nature of the LCA. See Fukui and Takano (1998), Ackema and Neeleman (2002), Richards (2004), Abels and Neeleman (2009) and Narita (2010) for further discussion.

Moreover, advocates of the LCA must be ready to allow massive parametric variation with regard to the availability of movement of the sort in (29). It has been long known since Fukui (1993) that the directionality of adjunction rather strongly correlates with the value of directionality-parameter of the language in question: Rightward adjunction is quite free in head-initial languages like English, whereas leftward adjunction is

freely available in head-final languages like Japanese.[5]

(31) a. John hugged Mary [*passionately*] [*without hesitation*] [*out of love*] ... .

 b. *John ... [*passionately*] [*without hesitation*] [*out of love*] hugged Mary.

 c. John-ga   ... [aizyoo-kara] [tamerai-naku]   [zyoonetuteki-ni] Mary-o
  John-NOM   love-from   hesitation-without passionately   Mary-ACC
  dakisimeta.
  hugged

 d. *John-ga Mary-o   dakisimeta [aizyoo-kara] [tamerai-naku]
  John-NOM Mary-ACC hugged   love-from   hesitation-without
  [zyoonetuteki-ni] ... .
   passionately

Thus, if advocates of the LCA manage to capture the pattern in (31a) by either of the analyses in (30), they must also account for why the same analysis is unavailable in a comparable 'head-final' language.

Furthermore, even if we tentatively grant that these challenges can be met in some way or another, further considerations quickly show that the derivations in (30) make wrong predictions as to the islandhood of C. In Uriagereka's system, it is necessary to apply Transfer to C in any of the derivational options shown in (30), given that either C or H-C occupies a noncomplement position, either by EM or IM. This should mean that any C in the H-C-A word order should constitute an island for extraction, an obviously incorrect result in face of examples like (32).

(32) a. *To whom*₁ was *the letter*₂ [H sent] [C t_2 t_1] [A before Mary received it]?

 b. *At the Tokyo station*₁, *three men*₂ [H arrived] [C t_2 t_1] [A early in the morning].

Countless other examples of the relevant sort can be produced easily. These examples transparently falsify the analysis in (30a), and those in (30b) and (30c) would also be ruled out unless we further stipulate a considerable amount of 'look-ahead' A- and A'-movement

---

[5]Examples of rightward adjunction as in (31d) are marginally acceptable if they are understood as involving so-called 'right dislocation', a paratactic construction which exhibits a number of peculiar properties. See Kuno (1978), Simon (1989), Endo (1996), Tanaka (2001), Kato (2007) for analyses of this construction in Japanese.

that must apply before the Transfer of C. In such a system, moreover, the A- v.s. A-bar distinction of movement cannot be attributed to the distinction between phase-internal and interphasal movement as suggested by Chomsky (2007a, 2008), thus further complication seems unavoidable. I cannot provide a rigorous proof that no such analysis is tenable, but it is advocates of the LCA who should bear the burden of argument that such *ad hoc* analyses are indeed on the right track.

4.2.3   Sheehan's Label-based Resuscitation of Uriagereka's LCA

Our discussion so far indicates that the H-α schema-based account of CED effects, though first envisaged by Uriagereka (1999) and Nunes and Uriagereka (2000) as a corollary of the LCA, can be best achieved in syntax without the LCA. I proposed in Chapter 3 that the H-α schema should be alternatively deduced from the overarching hypothesis that there is no mechanism of projection in bare phrase structure.
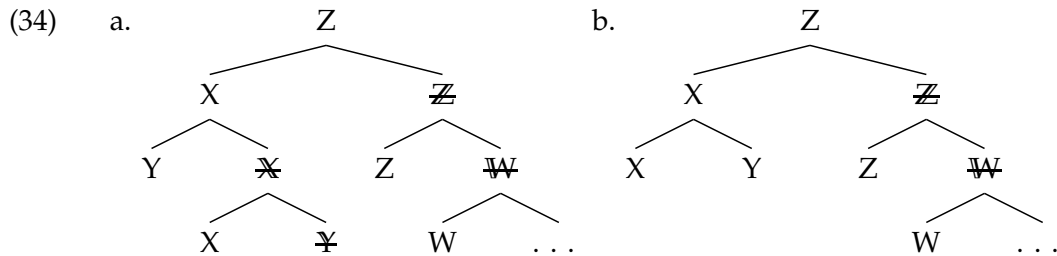
Interestingly, Sheehan (2010, in press) takes a different approach toward resuscitating the LCA-based account of CED effects. Instead of eradicating the LCA, she contends that Uriagereka's simplified LCA should be supplemented by a specific copy theory of labeling.[6] She puts forward the following set of assumptions:

(33)   a.   The label X of an SO Σ is literally a copy of the head X of Σ.

   b.   C-command is defined in such a way that a label X of Σ by itself c-commands the LIs within the sister of Σ.

   c.   Problematic labels/copies are deleted at PF as a last resort in order to enable LCA-based linearization.

   d.   PF-deletion of labels/copies is governed by some economy principle *à la* Nunes (2004) that favors deletion of lower copies over higher copies.

Consider, e.g., the following labeled tree diagrams, which according to Sheehan represent

---

[6]The idea that labeling really is copying of head LIs goes back to as early as Chomsky (1994, 1995).

an externally merged head-final and head-initial specifier XP of ZP. As is a familiar move in the literature of the LCA, Sheehan assumes that the surface head-final word order C-H is derived from the underlying head-initial order H-C by moving C to the specifier of H. To take a representative approach, Kayne (1994) proposes that c-selection forces ubiquitous roll-up complement-to-specifier movement in head-final languages, a proposal adopted by Sheehan.[7]

(34) a.

```
              Z                              b.              Z
        ┌─────┴─────┐                               ┌────────┴────────┐
        X          Z̶                                X                Z̶
     ┌──┴──┐     ┌──┴──┐                         ┌───┴───┐         ┌───┴───┐
     Y     X̶     Z    W̶                          X       Y         Z      W̶
        ┌──┴──┐    ┌──┴──┐                                            ┌──┴──┐
        X    Y̶    W    . . .                                          W    . . .
```

The nodes marked by outline typeface indicate the copies of LIs/labels that undergo deletion at PF. Sheehan specifically proposes that, thanks to (33c), one of the labels that symmetrically c-commands some other LI/label is deleted at PF. This copy-deletion process prefers to apply to 'bar-level' copies rather than to 'XP-level' copies by assumption (33d), thus the intermediate label Z rather than the maximal X label is deleted at PF. By (33b), the label X of XP can directly c-command Z and W in (34a/b), and thus the LCA gives rise to the ordering X→Z→W→... at PF. Moreover, as for the head-final specifier XP in (34a), the complement-to-specifier movement of Y enables Y to asymmetrically c-command X, leading to Y→X at PF. Sheehan argues that the combination of these two instructions are sufficient to get the total linear ordering Y→X→Z→W→..., and therefore that noncomplement-reduction is not forced by the LCA to apply to head-final specifiers. Note that the situation is different for head-initial specifiers as in XP in (34b). Here, the two comparable linearization instructions, X→Y and X→Z→W→..., are still insufficient to derive the total ordering, hence structures like (34b) are not linearizable in terms of the LCA. Sheehan claims that Uriagereka-type

---

[7]Note that Uriagereka's system makes an apparently wrong prediction that this obligatory complement-to-specifier movement should render every complement an island for extraction. Sheehan attempts to avoid this problem in an interesting way but unsuccessfully, as we will see shortly.

noncomplement-reduction is thus forced to apply to head-initial specifiers. This way, Sheehan draws the conclusion that the CED effect due to forced noncomplement-reduction arises only for externally merged head-initial noncomplements. Her system thus can rule in acceptable cases of subextraction from subjects in Japanese (e.g., (17) above) and other head-final languages reported in the literature (see Stepanov 2007), whereas the canonical CED effect for subjects in head-initial languages like English is still correctly derived from noncomplement-reduction.

Sheehan's label-based resuscitation of the LCA makes an interesting contrast with our eradication of the LCA from the H-α schema. In what follows, I would like to mention some problems related to Sheehan's approach.

First of all, because Sheehan advocates for the LCA, her theory inherits all the empirical problems attributed to the LCA. For example, she has to face serious problems like those regarding rightward adjunction discussed above (see Sheehan in press:§6 for some relevant but inconclusive discussion). See Fukui and Takano (1998), Ackema and Neeleman (2002), Richards (2004), Abels and Neeleman (2009) and Narita (2010) for further discussion.

Second, Sheehan's prediction that Uriagereka-type noncomplement-reduction obligatorily applies to head-initial specifiers upon EM is empirically too strong. To take one example, recall the Spanish data in (23)–(24) repeated here, which show that subextraction from post-verbal external arguments is strongly preferred to subextraction from preverbal ones.

(23)   *Spanish*: transparent postverbal subject (Uriagereka 1988:116; see also Gallego 2007, 2010)

a.   * [de qué conferenciantes]$_1$ te   parece   [que [las propuestas t_1]
         of what speakers   CL-to-you seem.3SG that the proposals
         me   van   a impresionar].
         CL-to-me go-3PL to impress
         "Of which speakers does it seem to you that [the proposals t] will impress me."

b. (?)[de qué conferenciantes]$_1$ te parece [que me van a
   of what speakers CL-to-you seem.3SG that CL-to-me go-3PL to
   impresionar [las propuestas t_1]].
   impress the proposals
   "Of which speakers does it seem to you that will impress me [the proposals

   t]."

(24) *Spanish*: transparent postverbal subject (Uriagereka 2009)

Qué partido$_i$ te hizo gritar [(el) que hayas perdido t_i]
What game you made scream the that have.you lost
'What game has it made you scream that you lost?'

Spanish is a *bona fide* head-initial language, thus Sheehan's theory predicts that specifiers in
this language undergo noncomplement-reduction upon EM and become islands for subex-
traction. External argument KPs should constitute islands under the standard assumption
that they are base-generated at Spec-v*, but the acceptability of (34b) clearly shows that the
external argument in the postverbal subject position (in-situ Spec-v*, as argued by Gallego
2007, 2010b) allows subextraction, contra Sheehan's prediction.

Last but not least, the central issue is again that reference to labels/projections is
bound to be a departure from the SMT in bare phrase structure: recall the discussion in
Chapter 3. Indeed, the assumptions which Sheehan adopts in (33) are technical compli-
cations that cannot be reasonably sustained in the pursuit of the minimal theory of bare
phrase structure. Labeled tree-diagrams are just informal proxies for rigorous set-theoretic
representations, however familiar they might look due to the 60 years of generative lin-
guists' conventions. No such structural representations can be generated in bare phrase
structure where the sole structure-building operation is recursively applicable Merge, just a
set-formation operation. Cast in a more accurate set-theoretic representation, the structures
in (34) are nothing more than (35), where no natural places for representational labels are
provided:

(35)   a.   {{Y, {X, Y}}, {Z, {W, ...}}}

        b.   {{X, Y}, {Z, {W, ...}}}

Nonetheless, it is crucial to Sheehan's approach that LIs can extend their c-command domain beyond their sisters by the medium of copies that are assigned (as labels) to certain complex syntactic objects (her (33a) and (33b)). Sheehan (2010, in press) proposes that her 'projection algorithm' formulated as (36) can derive the effect of (33a) and (33b):

(36)   If X selects and merges with Y then copy (the interpretable/unvalued features of) X to dominate {X, Y}.

However, it is not clear where X should be copied to, nor whether this copying process is achieved by IM or a totally different operation. No further exposition (nor justification) is provided in Sheehan (in press) as to the structural relation R (called 'dominate') between the copy of (interpretable/unvalued features of) X and the syntactic object {X, Y}, nor are we told how R can ensure the result that X directly c-commands Z and W in structures like {{X, Y}, {Z, {W, ...}}} (=(35b)). Some stipulative complication of the definitions of labels and c-command is unavoidable. This familiar problem is by no means unique to Sheehan, but rather it is shared by most of the approaches in minimalist syntax that presume some sort of representational labels/projections, either explicitly or implicitly, and I believe that this very difficulty rather strongly vindicates the theoretical move toward eliminating representational labels from the theory of syntax (see Chapter 3; see also Collins 2002, Seely 2006, Chomsky 2004, 2007a, 2008, Boeckx 2009, 2010b, Narita 2009a).[8]

---

[8]Note incidentally that Sheehan's attempt to resuscitate the LCA by means of the copy theory of labeling is partially motivated by her desire to capture (syntactically) a generalization known as the *Final-over-Final Constraint* (FOFC) (see Biberauer et al. 2008, Sheehan 2009, 2010).

(i)   *Final-Over-Final Constraint* (FOFC) (Biberauer et al. 2008:97):
       If α is a head-initial phrase and β is a phrase immediately dominating α, then β must be head-initial.
       If α is a head-final phrase, and β is a phrase immediately dominating α, then β can be head-initial or head-final.

However, given that there exists a simple and plausible parsing-based account of the FOFC effect by Cecchetto

4.2.4 Conclusion: the LCA Has No Place in Projection-free Syntax

It is Uriagereka (1999) who first ingeniously proposes a cyclicity-based account of Huang's (1982) CED. However, it is argued in this section that, despite Uriagereka's claim to the contrary, the model of cyclic Transfer can give a descriptively adequate account of the CED effect only when the H-α schema is dissociated from the LCA. In addition to various problems inherent to the LCA observed in the literature (see Fukui and Takano 1998, Ackema and Neeleman 2002, Richards 2004, Abels and Neeleman 2009 and Narita 2010 to name just a few), I provided various empirical generalizations and arguments that further support an LCA-free approach to the CED effect proposed in Chapter 3. In a nutshell, then, a full-fledged account of the CED in terms of the PIC can survive only in LCA-free syntax.

This conclusion leaves us with the choice between the H-α schema in LCA-free syntax or LCA-bound syntax without the H-α schema. At this point, let us come back to Uriagereka's (1999) criticism of the original LCA (Kayne 1994), whose exposition by Uriagereka (4) is repeated here:

(4)   The Linear Correspondence Axiom (LCA) (Kayne 1994; rephrased by Uriagereka 1999)

    a.   Base Step: If α c-commands β, then α precedes β.

    b.   Induction Step: If γ precedes β and γ dominates α, then α precedes β.

Uriagereka criticizes Kayne's version of the LCA by pointing out that only stipulations can assure asymmetric c-command relations between sister nonterminals that the induction step of the LCA requires. Reference to labels/nonterminals has been the major form of stipulation to this effect, be it the category-segment distinction among nonterminals as in Kayne (1994), or the invisibility of nonminimal-nonmaximal projections as in Chomsky

(2007, to appear), no recourse to the LCA is necessary, contra these authors' claims.

(1995). Then, Kayne's version of the LCA is bound to be a departure from the SMT (as correctly pointed out by Chomsky 2004:110 among others), to the extent that reference to labels/nonterminals is so. In fact, Kayne himself apparently agrees with this criticism by Uriagereka. Thus, he notes in his 2009 paper, "Full integration of the LCA with bare phrase structure will require reformulating the LCA without recourse to non-terminals." (Kayne 2009:note 8) I suspect that this problem eventually convinced him to decide to dispense with the LCA in his account of antisymmetry in Kayne (2010), as we will review shortly in §4.3.

Recall also that Uriagereka's attempt to revamp the LCA is motivated as a way to overcome stipulative references to labels/projections inherent to Kayne's original LCA. In this light, Sheehan's unsuccessful attempt to resuscitate Uriagereka's simplified LCA by adding another set of stipulations on labels is plainly putting things backward. Uriagereka's solution was to eliminate its induction step (4b) and let Transfer 'atomize' phases to simplex LIs, an idea that my proposal in Chapter 3 fundamentally shares with Uriagereka. What Uriagereka got wrong is just that he erroneously attempts to incorporate his theory into LCA-bound syntax, a failure that we successfully overcame by exploring alternative deductions of the H-α schema from the interplay of the property of EFs and the foundational principle of FI (15) (see §3.4; see also Chapter 5 for further exploration of FI). To the extent that Uriagereka is right in pointing out that his modification based on cyclic Transfer/Spell-Out is necessary to sever the LCA from stipulative references to labels, we can draw a much stronger conclusion, which is that the LCA has in fact no place in projection-free syntax. In contrast, the discussion in Chapter 3 is showing that the theory of bare phrase structure, if (and only if) dissociated from the LCA, can also pursue projection-free syntax. Both empirical coverage and theory-internal simplicity achieved by projection-free syntax are urging us to explore it, and therefore, LCA-free syntax.

In a nutshell, the conclusion is that the LCA is a wrong hypothesis as a theory of phonological linearization. This leaves us with the question of *why* the LCA must be wrong,

i.e., why the FL is designed in such a way that it can never accommodate the LCA as a possible principle of linearization. This question may be still pressing, given that the rather remarkable number of advocates in the field suggests that the LCA, despite its descriptive inadequacy discussed in this section as well as in Fukui and Takano (1998), Ackema and Neeleman (2002), Richards (2004), Abels and Neeleman (2009) and Narita (2010), has some intuitive appeal to many researchers. Importantly, Uriagereka's criticism again provides a necessary component of the principled answer to this question: no stipulation involving labels/projections should be available in minimal syntax, and therefore, the LCA (4) is plainly unformulable.

This is a strong argument against Kayne's original LCA, as well as Sheehan's, but it still leaves us with Uriagereka's simplified LCA, which makes little recourse to labeling/projeciton. To make sense of the unavailability of Uriagereka's LCA, it should be noted that reference to asymmetric c-command relations still constitutes a necessary component of Uriagereka's LCA. However, whether asymmetric c-command is available as a relation in syntax is questionable. In fact, Chomsky (2008:141) suggests that there is no notion of c-command available in syntax, and that various considerations that used to build on c-command, such as Agree and also some aspects of binding, can be reduced to minimal search for establishing probe-goal relations, requiring no recourse to c-command. Plausibly, then, the notion of 'c-command' is reduced to just an informal cover term to refer to a certain structural configuration, admitting of no real theoretical significance in syntax. This might provide an answer to the question why Uriagereka's LCA fails: there is no notion of c-command, and therefore Uriagereka's LCA is unavailable.[9]

Moreover, I would also like to point out that there is another ground for the hypothesis that asymmetric c-command cannot figure in the theory of linearization. Recall that the H-α schema generates unidirectional H-α branching of the form {X, {Y, {Z, ... }}}.

---

[9]However, it is arguably the case that Uriagereka's LCA is a theory of linearization that is simple enough to reformulate without recourse to c-command. See §4.4 for relevant discussion. See also Uriagereka (2009).

Importantly, it should be always the case that the most deeply embedded constituent in such a branching structure should assume a structure of the form {X, Y}, where both X and Y are LIs. X and Y c-command each other, so there is no asymmetric c-command available for this structure to be linearized in terms of any version of the LCA. More generally, any structure building in the framework of bare phrase structure should start by merging two LIs. Indeed, symmetric structures of the form {X, Y} are bound to be generated in each phase, and hence, LIs in each phase cannot be exhaustively characterized in terms of asymmetric c-command.[10]

To conclude, I argue that the following three premises are sufficient to exclude the LCA on principled grounds: there is no relation of c-command available in syntax; symmetric SOs of the form {X, Y} are unavoidably generated; and, most importantly, syntax assumes no labels/projections.

4.3 Eradicating Antisymmetry from the H-α Schema

4.3.1 Independence of Antisymmetry and the LCA

We saw in the previous section that the H-α schema can give rise to empirically correct consequences only in LCA-free syntax, and thus the LCA cannot be the correct linearization mechanism in our theory of bare phrase structure. At this point, it should be noted that the eradication of the LCA does not necessarily entail the eradication of antisymmetry in the sense of Kayne (1994).

According to Kayne (2009), "a relatively weaker interpretation of antisymmetry has it that no two human languages can be mirror images of one another, i.e. no pair

---

[10]Guimarães (2000) and Kayne (2009) suggest that one option for Merge is the direct formation of a singleton set {X}, and that the singleton-formation can be reasonably adopted for the purpose of exorcising symmetric {X, Y} structures: thus, {{X}, Y} or {X, {Y}} is sufficiently asymmetrical to ensure asymmetric c-command between X and Y. Although this might technically save the mechanism of the LCA-based linearization, further stipulations are necessary to ensure that all instances of merger of two LIs are mediated by the singleton-set formation. See Kayne (2009) for some relevant discussion. See also Fukui (forthcoming) for a different approach to symmetric structures.

of languages can have the property that one is the exact mirror image of the other (in the sense that each grammatical sentence of one has a grammatical counterpart in the other that is its mirror image, counting by morphemes, say)." The consequence of this weak interpretation of antisymmetry is that for any possible I-language L, there will be no 'mirror-image' language L' of L where each grammatical sentence of L can find its mirror-image counterpart as a grammatical sentence of L'. Let us refer to this conjecture as *weak antisymmetry*.

(37)   Weak antisymmetry:

No two natural languages can be mirror images of one another.

Weak antisymmetry can be counterposed to (38), a stronger interpretation of antisymmetry discussed in Kayne (2009):

(38)   Strong antisymmetry:

If some subtree (with both hierarchical structure and precedence relations specified) is well-formed in some human language, then its mirror image is well-formed in no human language.

Kayne's (1994) LCA is an ingeniously formulated principle that derives both strong and weak antisymmetry, as well as X-bar-theoretic constraints on possible phrase structures. What we have seen in the above discussion is only that none of Kayne's (1994), Uriagereka's or Sheehan's versions of the LCA works for capturing the CED effect, which does not in itself argue against weak and strong antisymmetry as empirical conjectures. Importantly, Kayne (2010) departs from his previous LCA-based conception of antisymmetry and puts forward an alternative hypothesis that deduces strong and weak antisymmetry without recourse to the LCA. What is interesting is that, in his alternative account of antisymmetry, he independently reaches the conclusion that there should not be any instance of XP-XP merger, which is quite similar to our H-α schema. Thus, a closer look at Kayne's (2010)

account of antisymmetry is necessary, a topic that I turn to in this section.

4.3.2   Kayne's (2010) LCA-free but Still Antisymmetric H-α Schema

First of all, it should be noted that, *pace* Sheehan's (2010, in press) argument to the contrary, Kayne himself is well aware of the fact that "Full integration of the LCA with bare phrase structure will require reformulating the LCA without recourse to non-terminals." (Kayne 2009:note 8) As pointed out by Uriagereka (1999) and also discussed in §4.2.1, the problem with non-terminals pertains to the issue concerning the induction step (b) of the LCA (repeated here as (39)), namely that only stipulations can ensure an asymmetric c-command relation between 'sister' phrases that apparently c-command each other, say the one between the external argument and the projection of v^*.

(39)   The Linear Correspondence Axiom (LCA) (Kayne 1994; rephrased by Uriagereka 1999)

   a.   Base Step: If α c-commands β, then α precedes β.

   b.   Induction Step: If γ precedes β and γ dominates α, then α precedes β.

We saw that Uriagereka's simplification of the LCA by means of his precedence-bound H-α schema cannot give rise to empirical predictions regarding the CED effect that Uriagereka (1999) and Nunes and Uriagereka (2000) claim it does. We also saw that Sheehan's label-based alternative cannot achieve a descriptively adequate account of the CED effect, either, even putting aside her complications to notions like c-command and labels. In contrast, Kayne's (2009) approach to the inherent incompatibility between bare phrase structure and the LCA is to discard the LCA-based deduction of antisymmetry. Instead, what he attempts in his (2010) is to "transpose the LCA-based ideas into the more derivational framework of Chomsky (1995) and later work," eliminating recourse to the LCA.

From the outset of his antisymmetry-based research, Kayne (1994, 2004a, 2010) is consistent in assuming that linear order is an integral part of syntax (see also Fukui 1993, 2001, Saito and Fukui 1998, Zwart forthcoming). The LCA holds that precedence relations among syntactic nodes, sets or not, are in one-to-one correspondence with asymmetric c-command relations thereof. Departing from this LCA-based hypothesis, Kayne (2010) proposes that Merge applying to α and β creates not a set of α and β ($\{\alpha, \beta\}$) but rather an ordered pair of α and β, $\langle \alpha, \beta \rangle$, specifying that α immediately precedes β (see already Saito and Fukui 1998). He calls his conception of pair-creating Merge *p-merge*, to differentiate it from Chomsky's set-Merge (and also from his pair-Merge; see Chomsky 2004, 2007a, 2008). For Kayne (2010), p-merge is the only structure-building operation available in syntax. Kayne further claims that precedence established in syntax can and should be understood as *immediate precedence*, a total ordering that has the property in (40).

(40)   Immediate precedence is total:

If X immediately precedes Z and Y immediately precedes Z then X = Y, and also if Z immediately precedes X and Z immediately precedes Y, then X = Y.

In principle, then, p-merge applying to α and β would generate either $\langle \alpha, \beta \rangle$ or $\langle \beta, \alpha \rangle$, which appears to be a *prima facie* instance of the very type of 'directionality-parameter' that antisymmetry is meant to replace. Kayne's task is to show that, for any S(pecifier), H(ead) and C(omplement) p-merge is constrained in such a way that (i) it can generate only $\langle H, C \rangle$ and not $\langle C, H \rangle$, and (ii) S can only be p-merged to the left of H. For the $\langle H, C \rangle$ problem, Kayne proposes the following performance-based principle:

(41)   Probe-goal search shares the directionality of parsing and of production.

Observing that both parsing and production show temporal sequence and therefore a beginning vs. end asymmetry, proceeding from left to right, Kayne claims that (41) yields

(42):

(42)   Probe-goal search proceeds from left to right.

This is claimed to hold "despite the fact that probe-goal search is not literally temporal in the way that parsing and production are." He goes on to claim that (42) is equivalent to:

(43)   Head and complement are invariably merged as ⟨H, C⟩.

given that the probe is the head and that the goal is the complement itself or is contained within the complement, where the notion of probe-goal is generalized over selector-selectee dependency, a rather familiar (though unwarranted) move in the minimalist literature (Svenonius 1994, Holmberg 2000, Pesetsky and Torrego 2001, 2004, 2006, 2007).

        How does Kayne deal with the other target result that S is invariably p-merged to the left of H? Interestingly, one of the building blocks of Kayne's solution to this Specifier problem is the following constraint on p-merge, which is apparently akin to the conclusion reached in Chapter 3.

(44)   The merger of two phrases is unavailable.

The similarity with the H-α schema that I advanced in Chapter 3, repeated here, is rather obvious.

(1)   ***The H-α schema***: Merge(H, α) → {H, α}.
        Merge must take at least one LI as its input.

Let me then formulate Kayne's version of the H-α schema as follows:

(45)   ***Kayne's (2010) H-α schema***:
        p-merge(H, α) → ⟨H, α⟩ or ⟨α, H⟩.

Turning to the merger of a S(pecifier), a typical instance of 'XP-XP merger', Kayne specifically claims that S is not really p-merged with the phrasal complex ⟨H, C⟩ but rather with H itself, resulting in ⟨S, H⟩ where S immediately precedes H.[11] That is, Kayne hypothesizes that 'S H C' does not form a standard constituent, and that p-merge creates two discontinuous phrases, ⟨S, H⟩ and ⟨H, C⟩. According to Kayne, ⟨S, H⟩ follows from the assumption that the immediate precedence relation that p-merge generates is total (40), which naturally leads to (46):

(46)   If H p-merges with X and also p-merges with Y, then X and Y must be on opposite sides of H in terms of immediate precedence.

Recall that Kayne's proposition that C is invariably p-merged to the immediate right of H follows from (42) and ultimately from the performance-based stipulation in (41). If H is further p-merged with an element S that is distinct from C, it follows from the total property of immediate precedence that S should be p-merged to the opposite side of H.[12] Kayne's claim is that this is why S is always merged to the immediate left of H, invariably resulting in ⟨S, H⟩.

        More generally, it follows from (46) that:

(47)   A lexical item H can be p-merged with at most two distinct elements.

Kayne transposes into (47) the earlier LCA-based proposition (48):

---

[11] According to Kayne, "Taking S [...] to merge with H itself would sharpen the sense in which heads are central to syntax, going back to Chomsky (Chomsky 1970). Every instance of Merge must directly involve a head, in the sense that (at least) one of the two syntactic objects merged must be a head. Merge never constructs a set consisting of two syntactic objects each of which is a phrase. From this perspective, (48) [the 'directionality-parametric' proposition that Spell out the specifier S of H to the left/right of the phrase headed by H that S is merging with, HN] is not statable insofar as S(pecifier) is not actually merging with any phrase at all."

[12] This way, the distinctness of C and S is a necessary condition for Kayne's (2010) p-merge-based account of antisymmetry. Incidentally, I suspect this constitutes one of the reasons why Kayne excludes movement of complement to the specifier position of the same head, regarding such an option as no longer viable (*ibid*, note 47). See also Kayne (1994:Ch.6) v.s. Kayne (2004b) on adpositions.

(48)   A single head can host at most one specifier.

(48) partially reconstructs the effect of X-bar theory that was once proposed to be captured by the LCA and the theory of Chomsky-adjunction, incorporating the category-segment distinction on nonterminals.

Note that the S-H-C linear sequence does not form a standard sort of constituent: there are only ⟨H, C⟩ and ⟨S, H⟩, but not ⟨S, ⟨H, C⟩⟩ or anything equivalent to this. Correspoindingly, syntax generates no p-merge-based structure that 'dominates' or 'contains' all of S, H and C. This led Kayne to stipulate the additional definition of 'maximal projection' in (49), so as to capture the apparent constituency of 'S-H-C'.

(49)   The maximal projection of a head H is the maximal set of ordered pairs each of
         which immediately contains H.

However *ad hoc* this may look, it yields the result that attraction of H will pied-pipe both ⟨H, C⟩ and ⟨S, H⟩, or so he claims.

Kayne's p-merge-based account of antisymmetry makes no recourse to asymmetric c-command, or to labels and nonterminals, while it can apparently accommodate most of the empirical predictions attributed to the LCA. The target results for Kayne are, among other things, strong and weak antisymmetry ((38) and (37)), as well as the single-specifier condition (48). Unlike Uriagereka's and Sheehan's LCA, Kayne seems to have no intention of defending his p-merge-based theory against problems raised by the CED and other locality constraints.

It is interesting to observe that he ended up proposing a kind of H-α schema (45) as a way to eliminate recourse to labels and nonterminals. That much is compatible with my conclusion established in Chapter 3, namely that there should be no reference to labeling/projection in the minimal theory of syntax.

However, the cost of his p-merge hypothesis is that it necessarily incorporates precedence into the core of narrow syntax. Recall our initial motivation for bare phrase structure, discussed in §1.2: we have to construct a theory of syntax in which no information about linear order is available. Berwick and Chomsky (2008) argue that the eradication of linear order from narrow syntax constitutes a necessary component of a possible principled explanation for the *structure dependence* of grammatical rules, a very well-founded generalization that all the grammatical rules are formulable only in terms of hierarchical organization (compositional structure) of linguistic constituents. Berwick and Chomsky urge us to demonstrate a principled explanation for the total non-existence of structure-independent (linear-dependent) rules, the simplest of which would just say that there are no linear-dependent rules since linear order really isn't part of syntax. I agree with Berwick and Chomsky in regarding the advent of bare phrase structure (Chomsky 1994, 1995 *et seq.*) as a real step toward this goal, in that it made an important departure from the earlier transformational grammar, where compositional structuring, labeling, and precedence were erroneously encapsulated into PSR-based rule schemata. Then, as far as we agree that we need an explanation of structure dependence and that bare phrase structure paves the way for tackling this problem, we had better refrain from re-importing precedence into the theory of minimal syntax. Correspondingly, Kayne's p-merge based theory of antisymmetry is by and large orthogonal to, if not incompatible with, the present exploration of the minimal theory of compositional syntax.

Counter to the above-mentioned approach, Kayne (2010:note 46) contends that the separation of precedence from syntax has been a mistake in generative lingusitics. He attempts to argue for the relevance of linear order in syntax by pointing to various gaps and asymmetries in crosslinguistic data. The observations he discusses include the ubiquity of leftward (as opposed to rightward) movement, the systematic difference between clitic left- and right-dislocation, the lack of obligatory preposing of lexical objects to the left of their associated pronominal clitics, the overall tendency of prenominal relatives (as op-

posed to postnominal relatives) to be non-finite, the rarity of backward pronominalization as opposed to forward pronominalization, and so on. These typological tendencies are no doubt interesting, but it simply doesn't follow that these typological gaps are necessarily regarded as facts about narrow syntax. Of course, it would be good if we somehow managed to pin down a single factor that derives all these interesting facts about human linguistic behaviors, but I suspect it is too much to expect. The above-mentioned facts are all complicated phenomena involving not only syntax but also numerous other factors pertaining to the use of linguistic capacity, including pragmatics, language-specific morpho-phonology, memory-limitations, principles of parsing and production, and so on. Those peripheral and/or extra-grammatical factors are no doubt operative in linguistic use, and their effects may well be too significant to ignore or abstract away from in the account of surface typological observations. Certainly, then, such observations alone do not reasonably warrant the re-incorporation of precedence into syntax, which is basically tantamount to the withdrawal of a principled explanation for the ubiquitous structure dependence of linguistic computation.[13]

4.3.3 Antisymmetry and Cartography in the Pursuit of Biological Adequacy

So long as we admit that the above-mentioned typological gaps and asymmetries are of limited relevance to the study of syntax, it becomes correspondingly questionable that antisymmetry really is a fact about the core architecture of FL. Recall that Kayne grants two interpretations of the theme of antisymmetry, weak and strong:

(37) Weak antisymmetry:

   No two natural languages can be mirror images of one another.

---

[13]Moreover, there are also typological facts that go in the opposite direction. For example, the prevalence of the head-final word order is quite unexpected on Kayne's grounds.

(38)   Strong antisymmetry:

If some subtree (with both hierarchical structure and precedence relations specified) is well-formed in some human language, then its mirror image is well-formed in no human language.

Weak antisymmetry is a hypothesis about weak generative capacity, concerning the set of word strings generable by the grammar. It holds that there can be no pair of natural languages, L and L′, such that a sequence of words W_1—W_2— ... —W_n is a grammatical sentence of L if and only if its mirror image, W_n—W_{n-1}— ... —W_1, is a grammatical sentence of L′. So far, no such pair of languages was reported in the literature, corroborating weak antisymmetry as an empirical generalization, and we are correspondingly interested in why weak antisymmetry holds. By contrast, strong antisymmetry is a hypothesis about strong generative capacity, concerning abstract phrase structure, and correspondingly, no direct observation on surface strings can speak for or against strong antisymmetry. These two notions of antisymmetry are therefore independent of each other.

Kayne's approach is to maintain that weak antisymmetry arises primarily as a consequence of strong antisymmetry, taking it as an axiomatic component of UG. However, possible accounts of weak antisymmetry may well involve complex interaction of multiple factors. Given the more-or-less shared capacities of parsing, production, perception, articulation, memory-limitations, as well as the genetically determined space of human interests and intentionality, it is quite a reasonable guess that certain temporal aspects of externalization processes show overall tendentious uniformity across human individuals, yielding various degrees of weak antisymmetry. Therefore, the strength of weak antisymmetry by no means guarantees the existence of strong antisymmetry. Correspondingly, it is an open question whether strong antisymmetry is a real axiom of commendable simplicity that can constitute the basis of an account of weak antisymmetry.

In this relation, it is worth noting that presumption of strong antisymmetry has provided one of the main motivations for the recent 'inflation' of functional categories in the syntax of clausal structure (Cinque 1999, 2002, Rizzi 2004). The reason is that strong antisymmetry of the LCA sort entails the following three properties: (A) there is no distinction between Specs and adjuncts (Kayne 1994:Ch. 3), (B) any 'projection' of a single head can host up to only one complement and one specifier/adjunct,[14] disallowing multiple Specs/adjuncts per head (*ibid*, Ch. 2), and (C) there is a single universal Spec-head-complement (S-H-C) word order (*ibid*). If (A)-(C) hold in syntax, then a logical consequence is that any 'deviations' from the S-H-C template, say a surface head-final word order or overt adjuncts, must be each reanalyzed as involving a separate functional head attracting its complement or selecting another specifier. Thus, if we presume strong antisymmetry, the existing variety of word order in an I-language must be seen as evidence for many covert functional projections in that I-language. Indeed, Kayne himself argues that his LCA-based theory should be regarded as providing a partial answer to the question of why there are so many functional heads, which he claims is that "functional heads make landing sites available." (*ibid*, 29-30) In this way, the LCA has been a big patron of the cartography expansion in UG, which would be no less true for Kayne's (2010) p-merge-based alternative conception of strong antisymmetry.

Of course, the LCA was originally proposed as an explanatory principle that deduces a number of apparently unrelated conditions in syntax, effects of the traditional X-bar theory being the crucial one.This has been regarded as a highly desirable result, and many researchers have attempted to maintain the basic insights behind the LCA in various terms while revising it in a number of different ways. Such proposals include Chomsky (1994, 1995), Epstein et al. (1998), Uriagereka (1999, 2009), Moro (2000), Guimarães

---

[14]Under bare phrase structure where recursive Merge can apply freely, forbidding multiple specifiers will require stipulations, departing from minimalist desiderata. In fact, there have been provided a number of empirical arguments for the existence of multiple specifiers, including Fukui (1986/1995), Speas (1986), Fukui and Speas (1986), Chomsky (1995, 2000, 2001), Ura (1996), Saito and Fukui (1998), Fukui and Takano (1998, 2000), to name just a few.

(2000), Nunes (2004), Jayaseelan (2008), Sheehan (2009, 2010, in press) and Kayne (2009), to name just a few. However, now that we established in Chapter 3 that the X-bar theory *is* a departure from the SMT, we can conclude that the empirical ground of the LCA is undermined as well.

Moreover, it should be noted that two decades of investigation have shown that the ramifications of imposing the universal S-H-C template *à la* the LCA are rather daunting. The inflation of functional categories is one of such unfavorable result. To illustrate the situation with Japanese, a strictly head-final language, Kayne's strict universal S-H-C word order forces us to analyze all the instances of surface S-C-H word order in this language as involving obligatory movement of some complement (often remnant) XP to the specifier of a covert functional category. This movement must always apply regardless of the categorial status of the XP. We don't know even how to formulate such a movement requirement in the current theoretical framework, since any movement/attraction operation is considered to select certain syntactic elements, e.g., φ-feature bearers. Even abstracting away from this feature-selection problem, we cannot do anything but further postulate *ad hoc* uninterpretable features or the like on ad hoc functional categories, again departing from the goal of biological adequacy. As Fukui and Takano (1998: 33) conclude, "We cannot think of any independent motivation for movement of complements other than the very reason for getting the surface order right (the C-H order), and it looks as though the postulated functional category attracts complements of any type just to ensure the correct word order under the LCA," or strong antisymmetry more generally.

Head-finality is just one of the several commonly attested deviations from the S-H-C template. It is true that any such deviations can be techically analyzed in conformity with strong antisymmetry as long as we add as many functional categories and uninterpretable features as needed to the cartography of clausal syntax. However, if we were to take that path, namely the marriage of strong antisymmetry and the cartography expansion, the machinery that such analyses employ—covert functional categories and

massive application of phrasal (remnant) movements into their Specs—would become just too powerful to rule out *any* surface word order, thus such a cartographic escape, necessary to save antisymmetry from the difficulties in its crosslinguistic coverage, in effect nullifies Kayne's (2004a) and Cinque's (2005) arguments for the LCA (and strong antisymmetry) from typological gaps in the attested surface word order variations. And the massive overgeneration, inherent to the marriage of strong antisymmetry *à la* the LCA and the cartographic approach, can be blocked only by further stiplations, again departing from the minimalist goal of biological adequacy. See Ackema and Neeleman (2002), Richards (2004), Abels and Neeleman (2009), and Narita (2010) among others for much relevant discussion.

Furthermore, postulating a new functional category X to facilitate strong antisymmetry will by itself complicate UG in various ways: first of all, UG should provide a proper theoretical description of X: To define the working of X, we should postulate:

(50)   A bundle of features of X that collectively determine:

      a.   what category X takes as its complement

      b.   what kind of elements X attracts or selects as its specifier(s)

      c.   how X contributes to interpretation at SEM and PHON

And so on. Moreover, we should come up with some non-stipulative explanation of:

(51)   a.   why X emerges in FL at all

      b.   how the parametric properties of X are acquired through experience for each I-language.

Notice that virtually all previous proposals simply assume the full set of functional categories as predetermined properties of UG.[15] This assumption is a major roadblock on the path to the minimalist goal of biological adequacy, which asks why FL comes to have the

---

[15]The only exception to the best of my knowledge might be Boeckx (2008a), who suggests that the cartographic layers might emerge as a fractal structure obeying the recurring X-bar schematic theme.

properties it has now, not many other imaginable ones.

Quite accurately describing the state of affairs, Fukui (2006:3) notes, "Given their usefulness as a descriptive tool, ... functional elements have sometimes been overused in syntactic analyses, particularly in the late 1980s, a situation that is reminiscent of the overuse of grammatical transformations in the 1970s or parameters in the early 1980s," a statement that is especially true in the Cinquean cartography project (Cinque 1999, 2002, Rizzi 2004). The full set of earlier language-specific or construction-specific transformational rules, postulated with an eye towards meeting observational or descriptive adequacy, once seemed to set the goal of explanatory adequacy hopelessly remote from the early practice of generative grammar. However, the generative enterprise finally overcame the tension between descriptive and explanatory adequacy by giving birth to the Principles-&-Parameters approach to language acquisition. Early transformational rules are now regarded just as "taxonomic artifacts" without any theoretical significance, whose descriptive effects are to be deduced from interactions of UG principles and parameters.[16] The lesson to be learned is that *ad hoc* descriptive specifics are likely to be eventually reduced to simpler and deeper theoretical constructs with a broader range of application. Fukui's passage above expresses a reasonable hunch that essentially the same might be true for the recent overuse of functional categories, too. Indeed, all the past formulations of each and every functional category have been very FL-specific (if not I-language-specific) and 'meaning'-specific ('topic', 'mood', 'aspect', etc., if not construction-specific). The additional assumption that these categories are UG-fixed universals would resolve the significant question of explanatory adequacy in its technical sense, i.e., an encompassing description of the first factor that suffices to determine I-language from available data, but

---

[16]However, simply trading the descriptive burden from transformational rules to corresponding microparameters does not help much in the context of minimalism, since now parameters are just another set of "UG-stipulations" which are hopefully to be eliminated in favor of third-factor principles and properties of external performance systems that FL interfaces with. See Newmeyer (2005, 2006, 2010), Boeckx (2010a,b) for relevant discussion.

it crucially begs the minimalist question of biological adequacy.[17]

The MP is an attempt to show how many of the properties of FL we can attribute to the third factor, minimizing UG-stipulations with the hope for a (more) biologically adequate theory of human language. As it stands now, the cartography of functional categories is definitely FL-specific, hence departing from the SMT. And even if we further stipulate that each of these functional categories is there to enhance, say, expressiveness for the CI performance system (see, e.g., Miyagawa 2010 for an explicit statement of this hypothesis), such a teleological/functional explanation for (51a) is at best partial, and it still begs the rest of the questions in (50)-(51).

Many researchers are well aware of the fundamental tension between the inflation of functional categories (on demand of descriptive pressures) and the minimalist goal of biological adequacy. Thus, for example, Chomsky (1995:240) argues, "Postulation of a functional category has to be justified, either by output conditions (phonetic and semantic interpretation) or by theory-internal arguments. It bears a burden of proof, which is often not so easy to meet." He famously provides a theory-internal argument against postulating a functional category with null CI-interpretation, AGR, based on the principle of FI (15). We should apply such minimalist scrutiny to other functional categories as well. And of course, reluctance to attribute null interpretation to an abstract functional category (e.g., AGR) does not by any means imply willingness to attribute an arbitrarily isolated 'semantic content' to a newly stipulated functional category (topic, focus, subjecthood, force, etc.).

Advocates of the cartographic approach claim that a highly articulated functional hierarchy is necessary to achieve sufficient descriptive and/or explanatory adequacy. We should carefully scrutinize each of these arguments empirically and theoretically (see the discussion below), and seek to find a way to overcome the difficulties the cartography expansion necessarily poses (see (50)-(51)) for biological adequacy. We should regard the

---

[17]Of course, simply replacing functional categories with some post-syntactic covert 'type-shifting' operations does not do any good to this situation; essentially the same question arises for such other language-specific operations.

descriptive technicalities in the cartographic approach not as a final explanation but rather as a first descriptive approximation of the facts to be explained in terms of the three factors in the language design, with the significance of the third factor emphasized (Chomsky 2005).

Fortunately, now, we have seen that minimalist syntax provides no room for the LCA, by far the most dominant statement of strong antisymmetry. Kayne's p-merge-based alternative also suffers from its unexplanatory re-incorporation of precedence into syntax, and correspondingly loses its force in the MP. Therefore, the big patron for the inflation of functional categories, namely strong antisymmetry, has lost its force in syntactic theories, opening the door for more minimal analyses of clausal syntax, keeping closely to the SMT.

4.4   Linearization of Projection-free Syntax

We saw in the preceding discussion that the LCA cannot be the correct theory of linearization. The LCA is but only one of the past proposals on linearization that makes reference to labels, an option that must be excluded in the theory of syntax without labeling/projection. Incidentally, it seems that there have been so far no successful label-free proposals concerning the mechanics of linearization in the literature: past proposals, including any form of directionality parameter (Chomsky 1981, Epstein et al. 1998, Saito and Fukui 1998, Fox and Pesetsky 2005 etc.), the LCA of any sort and Fukui and Takano's (1998) symmetry principle, were all dependent on labeled phrase structure. However, this by no means proves that theorists of projection-free syntax can never approach the problem of linearization. In what follows, I will try to lay out my own proposal regarding the issue of linearization that makes no recourse to labeling.

Admittedly, the account to be explicated below is relatively weak in its predictive force. Especially, it does not provide any prediction of antisymmetry—it doesn't even aim at doing so. I do not regard this as a bad result, though. As I noted above, we can conjec-

ture that the theory of syntax does not necessarily bear the burden of accounting for weak antisymmetry, given that countless other complex factors figure in the process of externalization, and some of those factors, including processing efficiency, limitations on memory and attention, etc., which may well be shared across language users, would presumably figure in the account of weak antisymmetry. Moreover, Spell-Out of linguistic structures is in itself a very complicated task, too, involving massive violations of inclusiveness and no-tampering, but this may be just a result of its being subsidiary to the core architecture of FL, which presumably evolved saltationally in the species optimized primarily to interface with CI (Chomsky 2007a,b, 2008).

Before laying out the proposal, it may be worth pointing out first that Uriagereka's LCA is simple enough to reformulate without resorting to labels and c-command. For example, the following proposition suffices to capture the essential effect of Uriagereka's LCA, where H$\rightarrow\alpha$ denotes 'H precedes α':

(52)   A revamped version of Uriagereka's LCA:

An SO of the form {H, α} is mapped by Spell-Out to H$\rightarrow\alpha$.

Provided that each of the SOs subjected to Transfer unambiguously takes the form of unidirectional H-α branching (53), pretty much keeping to the line envisaged by Uriagereka, the unidirectional mapping of each SO {H, α} to H$\rightarrow\alpha$ is rather straightforward.

(53)   a.   {X, {Y, {Z, . . . }}}

   b.
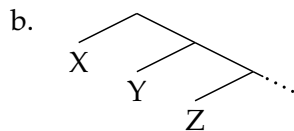
However, note that there is no principled reason to expect that the relevant relation established between H and α must be restricted to 'precedence' $\rightarrow$. That is, nothing in the

minimal theory of UG can prevent a comparable 'postcedence' relation from being established for particular pairs of H and α. That is, the following should be as simple a procedure of linearization as (52), despite Kayne's (1994, 2010) attempts to argue for the contrary.

(54)   An SO of the form {H, α} is mapped by Spell-Out to $\alpha \rightarrow$H.

Indeed, my claim to be articulated is that both (52) and (54) are available options for linearization—that is, each SO {H, α} can be mapped to either H$\rightarrow\alpha$ or $\alpha\rightarrow$H, depending on the lexical features of H.

    To lay out the proposal, recall the hypothesis touched upon in §3.4 and to be explored in Chapter 5: the interpretation assigned to an SO is predominantly determined by the properties of its head identified by MHD (55), basically the LI H in {H, α}.

(55)   ***Minimal head detection (MHD)*** :

    The head of an SO Σ is the most prominent LI within Σ.

If heads and their lexical properties prominently figure in the mapping to SEM, as I will argue in Chapter 5, the null hypothesis is that the same applies to the mapping to PHON too. Specifically for linearization, then, it is reasonable to assume that linearization procedures will also strongly hinge on heads and their lexical properties. So to speak, the head of each H-α structure serves as the locus of the above-mentioned linearization procedures. Moreover, if linearization really assumes the optionality of assigning precedence or postcedence to H and α in {H, α}, it is reasonable to suppose that the optionality will hinge on lexical features (categorial features in particular) of H, not α. Taking these into consideration, I would like to propose that phonological rules of each language will specify whether each LI H will be assigned precedence or postcedence in relation to its sister SO α, depending on featural specifications of H. To use some informal but helpful terminology, Spell-Out may determine whether each H will be 'prefixed' or 'suffixed' to α at PHON. Various 'redundancy rules' may sometimes achieve partial to total uniformities regarding prefixation v.s.

suffixation directionality in some languages, while sometimes not (cf. Chomsky 1965).[18]

For example, we may assume that the following ordering rules for the Spell-Out operations of English and Japanese, where $\rightarrow$ again stands for precedence, and Lin(α) stands for the linear sequence assigned to the SO α.

(56)   *English*: Spell-Out

    a.   Lin({D, α}) = D $\rightarrow$ Lin(α).

    b.   Lin({K, α}) = K $\rightarrow$ Lin(α).

    c.   Lin({V, α}) = V $\rightarrow$ Lin(α).

    d.   Lin({v, α}) = $v\rightarrow$ Lin(α).

    e.   Lin({T, α}) = T $\rightarrow$ Lin(α).

    f.   Lin({C, α}) = C $\rightarrow$ Lin(α).

(57)   *Japanese*: Spell-Out

    a.   Lin({D, α}) = Lin(α) $\rightarrow$ D.

    b.   Lin({K, α}) = Lin(α) $\rightarrow$ K.

    c.   Lin({V, α}) = Lin(α) $\rightarrow$ V.

    d.   Lin({v, α}) = Lin(α) $\rightarrow$ v.

    e.   Lin({T, α}) = Lin(α) $\rightarrow$ T.

    f.   Lin({C, α}) = Lin(α) $\rightarrow$ C.

These two sets of language-specific phonological rules constitute mirror images of each other. This is a reasonable hypothesis, given that these two languages are among the so-called 'harmonic' head-initial and head-final languages, and as the Spell-Out rules in (56-57) suggest, each instance of H in {H, α} is mapped to a unidirectional sequential order, H $\rightarrow$ α in English, and α $\rightarrow$ H in Japanese. English LIs are uniformly 'prefixal',

---

[18]Boeckx (2011) suggests that there may be a principle that sets a tendentious preference for parametric value consistency among similar lexical parameters (he calls this principle the 'Superset Bias'). Maybe such a principle is operative in the acquisition of prefixal/suffixal directionality.

whereas Japanese ones are more or less uniformly 'suffixal', so to speak.

Moreover, as the notation involving $\text{Lin}(\alpha)$ indicates, I will also specifically assume that the Spell-Out rules in (56)/(57) apply recursively from bottom-up in phase structures subjected to Spell-Out, just like recursive Merge applies invariably from bottom-up.[19] For example, there should be no 'intermediate' entities like $[v^* \to V] \to \{K, \{D, N\}\}$, where the primarily phonological relation $\to$ relates a linear sequence $[v^* \to V]$ of a phonological sort and a set-theoretic object $\{K, \{D, N\}\}$ of a syntactic one. On the other hand, the mapping from an SO $\{D, N\}$ to a linear sequence of LIs, say, $D \to N$ by (56a) is a legitimate operation, and so is the inductive mapping from $\{K, \{D, N\}\}$ to $K \to \text{Lin}(\{D, N\}) = K \to D \to N$ by (56b), as I will assume. Thus, for the present discussion, I will generally assume that $\text{Lin}(X) = X$ for each LI X, which will serve as the 'base step' for inductive linearization. This 'bottom-to-top' directionality of linearization will turn out to be an important hypothesis of distinct empirical import, as we will see shortly.

Let us go through the derivation of *the boy will eat the apple* described in (43)–(45) of Chapter 2, and its counterpart in Japanese:

(58)   *English*:

(that/if/whether/...) the boy will eat the apple.
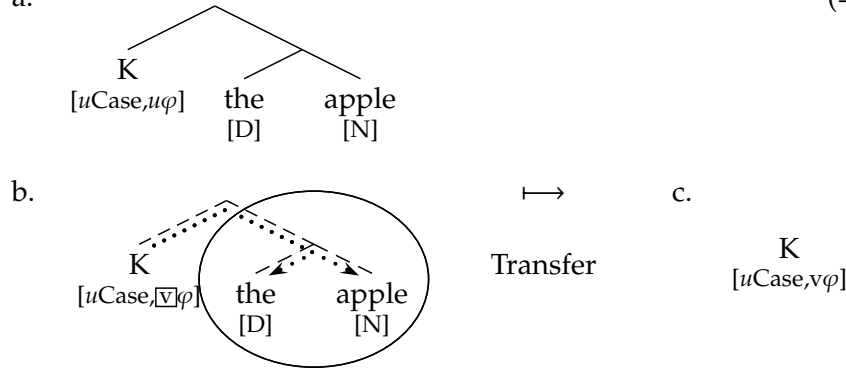
(59)   *Japanese*:

otokonoko-ga ringo-o   tabe-ru (yo/ne/no?...).
boy-NOM   apple-ACC eat-will SFP

First, applying to the object K-phase structure in (60), the English Spell-Out rule (56a) will first produce the sequence $the_D \to apple_N$, and then (56b) will attach K to the sequence, producing $K \to the_D \to apple_N$. Applying in the same order, Japanese Spell-Out rules (57)

---

[19]This claim is rooted in my assumption that linear sequences based on $\to$-relations are entities of a different category from SOs generated by narrow syntax. Such a distinction is sometimes blurred in the literature, but nevertheless is important, as effectively pointed out by de Vries (2009).

will produce $ringo_N \rightarrow \varnothing_D \rightarrow o_K$ '(the/a) apple (accusative)'.[20,21]

(60)   a.   (= (43), Chapter 2)

K
[uCase,$u\varphi$]   the   apple
          [D]   [N]

b. $\longmapsto$ c.

K   K
[uCase,$\boxed{v}\varphi$]   the   apple   Transfer   [uCase,vφ]
              [D]   [N]

We assumed that Transfer severs K from the interior SO {D, N}, so to speak, and leaves only K for later computation. However, the hypothesis was that Transfer sends not only the interior Σ itself but also all the structural relations established with regard to Σ, including the 'constitute' relation that Σ has established with the phase-head LI H for {H, Σ} (recall §2.5.1).

(61) α and β *constitute* γ (or conversely, γ *consists of* α and β) iff γ is the output of Merge(α,

        β).
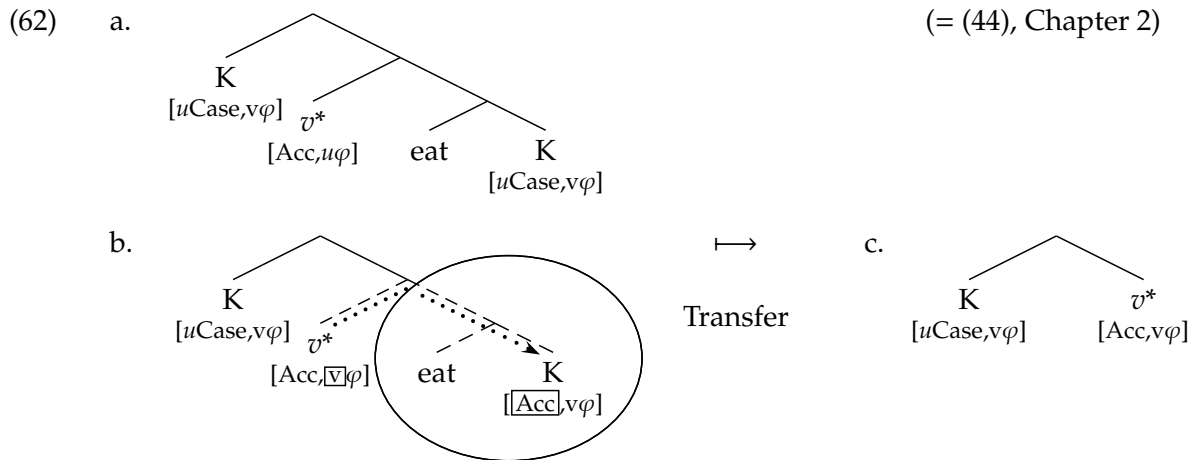
Thus, the constitute relation for {K, {D, N}} between K and {D, N} is subjected to Spell-Out at the K-phase-level Transfer, which sufficiently guarantees that K will precede or follow the linear sequence corresponding to {D, N}, as I will assume. The same consideration

---

[20]Japanese D is morphologically null most of the time, but we may regard, e.g., human definite plural suffixes *-tati*, *-ra*, *-domo*, etc., as instances of overt D in Japanese: consider, e.g., *kodomo-tati* 'the children (child-DEF.PL)'. The relation between plurality and definiteness in Japanese has been pointed out by Kawasaki (1989) and Kurafuji (1999). Cf. Watanabe (2006, 2008) for cartographic descriptions of Japanese nominals.

[21]Incidentally, it is not unreasonable to suppose that Japanese demonstratives *ko* 'this', *so* 'that', *a* 'that' and *do* 'which' are instances of a *bona fide* 'specifier/adjunct' D that somehow are lexically marked as instantiating a rather exceptional instance of 'head-initial' order D → α in Japanese. This option is available in our projection-free syntax, where no structural notions like 'specifier-of' or 'adjunct-of' are available. This simplistic approach might be sufficient to meet the problem regarding Japanese demonstratives raised by Hajime Hoji (p.c.), which concerns how such demonstratives (appearing in a genitive-marked 'spec'-like position) crucially determine the anaphoric properties of the KP. Of course, a question remains as to why these particular instances of D deviate from the otherwise uniform head-final word order in Japanese, but this may be just as puzzling as all the other instances of occasional deviations of directionality harmonization, such as the head-initial complementizer in otherwise head-final Germanic languages.

applies to the subject [K *the*$_D$ *boy*$_N$] in English and the corresponding [*otokonoko*$_N$ $\varnothing_D$ -*ga*$_K$] in Japanese. Incidentally, various considerations suggest that Japanese has no φ-feature agreement operations (Fukui 1986/1995, 1988, 1995, Fukui and Speas 1986, Kuroda 1988, Fukui and Sakai 2003), and hence no [$u\varphi$]'s are distributed over the relevant LIs in the Japanese counterparts of (60)–(63), but the procedure for linearization is essentially the same.

At the next phase with v^*, Transfer eliminates {*eat*$_V$, K} and assigns precedence to v^*, eat$_V$ and object K. In the same vein, the English Spell-Out rule (56c) maps {*eat*$_V$, K} to a linear sequence *eat*$_V$ → K, and combines it with the previously established sequence K → *the*$_D$ → *boy*$_N$, yielding *eat*$_V$ → K → *the*$_D$ → *apple*$_N$. (56d) then attaches v^* to the sequence, yielding v^* → *eat*$_V$ → K → *the*$_D$ → *apple*$_N$. Through the same processes, the Japanese Spell-Out rule will yield *ringo*$_N$ → $\varnothing_D$ → *o*$_K$ → *tabe*$_V$ → v^* 'eat (the/a) apple (accusative)'.

(62)   a.   (= (44), Chapter 2)

K
[uCase,vφ] v^*
[Acc,$u\varphi$]   eat   K
[uCase,vφ]

b. $\longmapsto$ c.

K   Transfer   K v^*
[uCase,vφ] v^* [uCase,vφ]   [Acc,vφ]
[Acc,$\boxed{v}\varphi$]   eat   K
[$\boxed{\text{Acc}}$,vφ]

Note that MHD (55) may single out either *eat*$_V$ or K as the head of {*eat*$_V$, K}. However, it is presumably the case that only the choice of V will lead to a legitimate interpretation at PHON, probably because the affixal morphology of the light verb v requires it to be affixed to the immediate left or right of V. I will also assume that the θ-role assignment from V to K at SEM indirectly requires the choice of V as the head of {V, K}. Therefore, it is only affixation of V that can continue cyclic linearization.

Finally, let us consider the C-phase-level linearization, where a couple of inter-

esting problems arise. The standard analysis holds that the subject in English and some
other languages undergoes the so-called 'EPP-driven' A-movement to 'Spec-T'. In §2.5.2,
I specifically argued that the movement results in the 'multi-rooted' structure in (63b),
because EM of C applies before IM of K, the latter being in simultaneity with Transfer.

(63) a.   (= (45), Chapter 2)
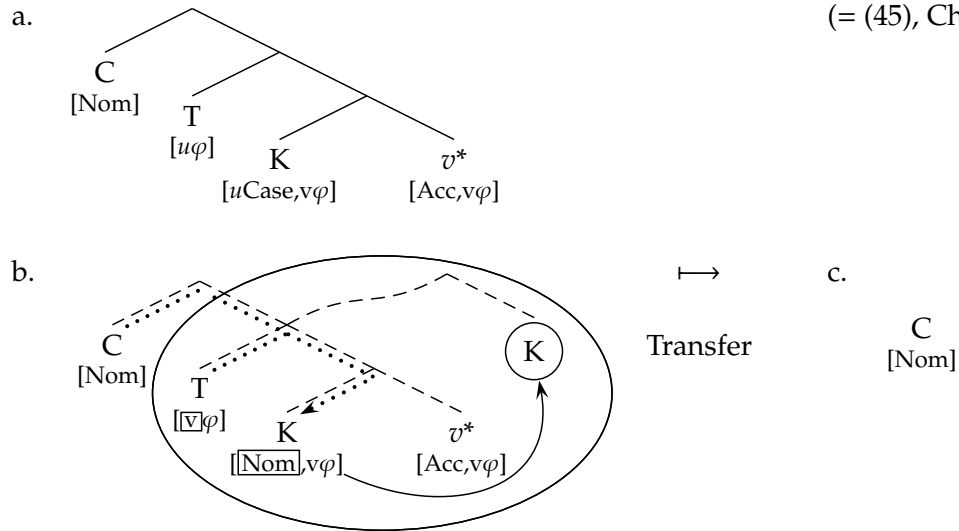
How does Spell-Out deal with such a multi-rooted structure? This problem seems nontriv-
ial, but note that the 'TP-node' root SO will be immediately subjected to Spell-Out/Transfer,
which applies simultaneously with the very application of IM that creates that SO. There-
fore, there is a sense in which linearization of elements internal to the TP-root precedes,
in rule ordering, linearization applying to C. I assume so, and propose that the mapping
of {C, α} to C $\rightarrow$ α will be postponed until Spell-Out of the TP-root determines the linear
sequence corresponding to α. I will return to this problem shortly.

Linearization of the TP-internal elements is also not trivial. One of the problems
is how Spell-Out can deal with the lowest constituent {K, v^*}. We assumed that the two LIs
contained in {K, v^*} are phase-heads that have already subjected their interiors to Transfer:
thus, K is already half-embedded in the linear sequence K $\rightarrow$ *the*$_D$ $\rightarrow$ *boy*$_N$, and v^* in v^* $\rightarrow$
eat$_V$ $\rightarrow$ K $\rightarrow$ *the*$_D$ $\rightarrow$ *apple*$_N$. No obvious instruction is given in (56) to achieve the ordering

of these pre-established two linear sequence of LIs.

However, what is fortunate in the TP-structure in (63b) is that the subject K moves to the 'EPP' position, resulting in {K, {T, {t_K, v*}}}, leaving a copy of K. Then, the lower copy of K is presumably deleted at PHON, due to the well-established generalization in (64), which is shared by virtually every proposal (cf. Sheehan's (33d)):

(64)   If IM creates {X, α},

      a.   the lower copy of X within α is deleted.

      b.   {X, α} is mapped to X → Lin(α).

Presumably, (64) contributes to the reduction of computational load of Spell-Out, and thus its emergence may be required by the principle of computational efficiency: if multiple copies are spelled out, each of the copies has to go through language-specific phonological processes over and over again, which will necessarily increase the burden of computation at Spell-Out. The computational load can be significantly reduced once Spell-Out is allowed to delete all but one copy.[22]

Moreover, it is reasonable to assume that there is a strong preference for movement by IM to apply leftward (64b). (64b) presumably has its root in the parsing preference for forward (left-to-right) filler-gap (antecedent-trace) dependency over backward depen-

---

[22]It may be that (64a) should be understood not as an exceptionless 'hard' principle but rather as a softer preference that may be overridden by some other phonological considerations. See, e.g., Bobaljik (1995), Bošković (2002b), Kato (2004), Narita (2007a,b, 2008) among many others for the proposal that certain instances of covert movement in various linguistic phenomena may be achieved by deleting the higher copy of the moved element, constituting minor departures from (64a) of limited sorts. For example, Bošković (2002b) proposes that certain instances of covert *wh*-movement in multiple WH-fronting languages, exemplified in (i), results from respecting a PF-constraint against consecutive homophonous *wh*-phrases (example (i) is drawn from Serbo-Croatian).

(i)   *Serbo-Croatian*: (Bošković 2002b)

    Šta   ~~šta~~$_i$ uslovljava **šta**$_i$?
    what   conditions what

The fundamental insight that Bošković and others pursue is that the deletion of a higher copy is a "last resort" strategy of Spell-Out that is undertaken in order to salvage otherwise legitimate syntactic structures from violations of PF-constraints. I will put aside the study of such phenomena. See also Nissenbaum (2000), Marušič (2009) and others.

dency (see Ackema and Neeleman 2002, Abels and Neeleman 2009, Cecchetto 2007, to appear for quite relevant discussions). Given these considerations, I maintain that (64a) and (64b) are actutally integrated as concrete rules of Spell-Out.

Then, the ordering problem for $\{K, v^*\}$, say in English, is effectively resolved by movement of K to Spec-T. The occurrence of K in $\{K, v^*\}$ is deleted by (64a), and thus $\{K, v^*\}$ will be simply mapped to the linear sequence established at the v^*-phase, $v^* \rightarrow eat_V \rightarrow K \rightarrow the_D \rightarrow apple_N$.[23] $\{will_T, \{t_K, v^*\}\}$ will be then mapped by (56e) to $will_T \rightarrow v^* \rightarrow eat_V \rightarrow K \rightarrow the_D \rightarrow apple_N$.

A technical question remains as to how the linear sequence achieved at the K-phase-level Spell-Out, namely $K \rightarrow the_D \rightarrow boy_N$, is carried along with the moved K. Let's call this linear sequence L1. Notice that (64b), formulated as such, only guarantees that K precedes the linear sequence corresponding to $\{will_T, \{t_K, v^*\}\}$, namely $will_T \rightarrow v^* \rightarrow eat_V \rightarrow K \rightarrow the_D \rightarrow apple_N$. Then, it will yield $K \rightarrow will_T \rightarrow v^* \rightarrow eat_V \rightarrow K \rightarrow the_D \rightarrow apple_N$. Call this would-be sequence L2. Then, a literal application of (64b) would generate two linear sequences that are still not fully integrated with each other, namely L1 and L2.

(65)   a.   L1: $K \rightarrow the_D \rightarrow boy_N$

    b.   L2: $K \rightarrow will_T \rightarrow v^* \rightarrow eat_V \rightarrow K \rightarrow the_D \rightarrow apple_N$

This is clearly an undesirable result. How should we deal with this issue? To tackle this problem, let us note that the generation of L1 is in a certain sense 'earlier' than that of L2, in that L1 is already established at the end of the K-phase-level Spell-Out. In contrast, the instructions that would yield L2 are given only at the C-phase. Taking this into consideration, I would like to propose the following linearization convention for the mapping of '$\rightarrow$'-relations to full linear sequences of LIs.

---

[23]Incidentally, we may also speculate that the very ordering problem of $\{K, v^*\}$ provides a partial rationale for the EPP: thus, EPP movement arises in some languages like English as a trick to overcome the partial undecidability of the ordering of K and v^*. See also note 17 of Chapter 3.

(66)   a.   If the sequence $\alpha \rightarrow \beta$ is generated prior to the sequence $\alpha \rightarrow \gamma$, then map these

sequences to $\alpha \rightarrow \beta \rightarrow \gamma$.

b.   If the sequence $\alpha \rightarrow \beta$ is generated prior to the sequence $\gamma \rightarrow \beta$, then map these

sequences to $\gamma \rightarrow \alpha \rightarrow \beta$.

Thanks to (66), then, L1 is sustained as such, and L2 minus the leftmost K will be attached to

the immediate right of L1. The result of the C-phase-level Spell-Out is the linear sequence

(67), the target result.

(67) $K \rightarrow the_D \rightarrow boy_N \rightarrow will_T \rightarrow v^* \rightarrow eat_V \rightarrow K \rightarrow the_D \rightarrow apple_N$

It should be also noted that the same convention in (66) will presumably be

responsible for the later linearization of C with respect to the 'T'-node', too. Notice that the

SO $\{T, \{K, v^*\}\}$ has two sisters, the externally merged C and the internally merged subject K,

whose proper linearization would invoke the same sort of integration problem. Then, just

like the case in (65), linearization of C will produce an again unintegrated linear sequence

L3:

(68)   L3: $C \rightarrow will_T \rightarrow v^* \rightarrow eat_V \rightarrow K \rightarrow the_D \rightarrow apple_N$

However, notice that C is left in the active workspace for later computation, and therefore

the ultimate attachment of C to Lin($\{T, \{K, v^*\}\}$), producing L3, will have to await Spell-Out

at the next phase level. This means that the integration of L1 and L2, yielding (67), precedes

the production of L3. Thus, (67) and L3 are integrated by the convention (66), correctly

producing (69).[24]

---

[24] Incidentally, things might become different when we adopt Chomsky's (2004) hypothesis that Transfer, if applied to the root clause CP, will subject the entire CP, including C, to Spell-Out (a hypothesis that the definition of Transfer in reached in Chapter 2 at least makes room for; see §2.5.1). If we pursue this hypothesis, there is no obvious sense in which the generation of L3 follows the generation of L1 and L2 in (65). This may be a true case where the convention in (66) will fail to achieve total ordering, yielding undecidability of precedence between C and the moved K with respect to Lin($\{T, \{K, v^*\}\}$). Some phonological means should then be employed to overcome this problem. One may speculate that assignment of a null phonetic content to C might be a simple way to circumvent the problem, as in English root clauses. Moreover, alignment of

(69) $(that/if/whether/\oslash)_C \rightarrow K \rightarrow the_D \rightarrow boy_N \rightarrow will_T \rightarrow v^* \rightarrow eat_V \rightarrow K \rightarrow the_D \rightarrow apple_N$

Let us finally go through the C-phase in Japanese. As we saw above, there is evidence that Japanese subjects can stay in situ at the edge of v^* (Fukui 1986/1995, Kuroda 1988 and Kato 2006).[25] In particular, recall the lack of CED effects for subjects in Japanese, which shows in our account that KP in {KP, v^*} can stay in situ (see §3.3.2). Thus, consider the following structure (linear order irrelevant in syntax; as noted above, there is no evidence for the presence of $[u\varphi]/[v\varphi]$ in Japanese):

(70)   a.

b. $\longmapsto$ c.

Transfer
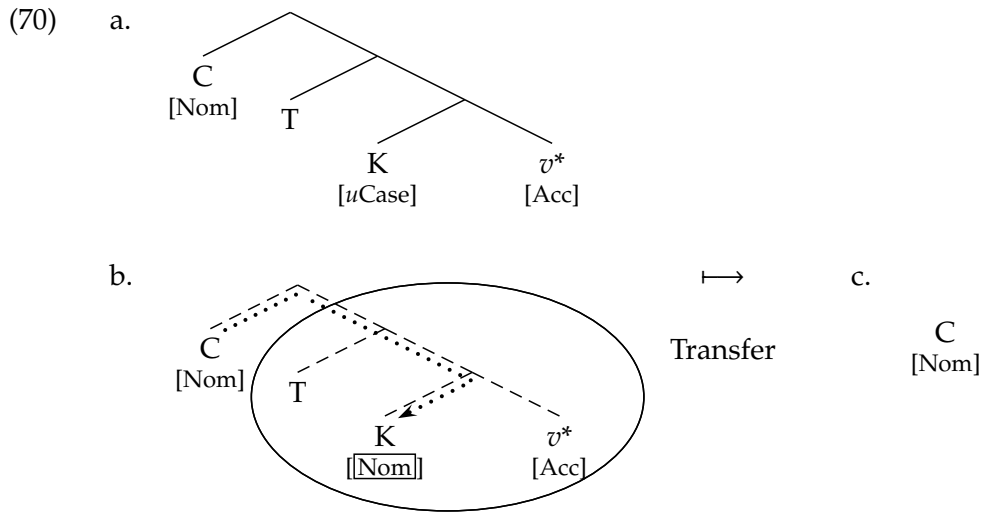
Note that, unlike the derivation for English in (63), the occurrence of K in (70b) will not undergo deletion, since no IM applies to it. Then, {K, v^*} should be mapped to some sort of linear sequence. Recall that Spell-Out appling to the K-phase and the v^*-phase produce

C to the right of Lin({T, {K, v^*}}) may also salvage the structure from linearization undecidability, an option presumably instantiated by sentence-final particles in Japanese and other languages. Such speculations may provide a partial explanation for why sentence-*initial* complementizers are almost always crosslinguistically null in root-clauses, whereas there are numerous examples of sentence-final complementizer-like particles attested crosslinguistically, predominantly appearing in root clauses. I will leave this interesting problem for future research, but it is significant that our label-free approach to the linearization problem paves the way for providing a partial account for such a well-established typological finding.

[25]The edge of T can be optionally utilized by scrambling. See Saito (1985, 1992, 2003), Kuroda (1988), Fukui (1986/1995, 1988), Fukui and Speas (1986), Saito and Fukui (1998), Miyagawa (1997, 2003) among many others. But see also Ueyama (1998), Fukui and Kasai (2004).

the following two independent sequences:

(71)   a.   *otokonoko*$_N$ → ∅$_D$ → *ga*$_K$ '(the/a) boy (nominative)'.

       b.   *ringo*$_N$ → ∅$_D$ → *o*$_K$ → *tabe*$_V$ → v^* 'eat (the/a) apple (accusative)'.

Now, how can Spell-Out combine them in accordance with the Japanese-specific rules in (57)? I would like to maintain that the same convention in (66) can be adopted to account for this problem, too. I will suggest in Chapter 5 that a certain θ-theoretic consideration requires that v^*, not K, is singled out as the head of {K, v^*} by MHD (55). Assuming this much, then {K, v^*} will be mapped by (57d) to the following two yet unintegrated sequences, thanks to v^* being its head:

(72)   a.   *ringo*$_N$ → ∅$_D$ → *o*$_K$ → *tabe*$_V$ → v^* (= (71b))

       b.   *otokonoko*$_N$ → ∅$_D$ → *ga*$_K$ → v^*

That is to say, rule (57d) is employed twice, once at the end of the v^*-phase-level Spell-Out (yielding (72a)) and then at the beginning of the C-phase-level Spell-Out (yielding (72b)). Given that the relevant linearization processes apply in that order, and given that v^* is shared by both (72a) and (72b), the convention in (66) will ensure the result that *otokonoko*$_N$ → ∅$_D$ → *ga*$_K$ of (72b) is attached to the left of (72a), the target result. Further linearization of {T, {K, v^*}} and {C, {T, {K, v^*}}} by (57e) and (57f) will then yield the correct order in the final sentence.

(73)   *otokonoko*$_N$ → ∅$_D$ → *ga*$_K$ → *ringo*$_N$ → ∅$_D$ → *o*$_K$ → *tabe*$_V$ → v^* → *ru*$_T$ → (*yo/ne/no?/...*)$_C$

(59)   *Japanese*:

       otokonoko-ga ringo-o   tabe-ru (yo/ne/no?...).
       boy-NOM   apple-ACC eat-will SFP

        There is no doubt that a lot of ramifications of this proposal are left for future re-search, but it should be clear from the discussion above that no recourse to labels/projections

nor c-command is necessary in accounting for surface word orders. Thus, all the children need to learn from the external data are the rules in (56) and (57) as well as the one for IM (64). This is not a difficult task, given that the number of rules in (56)/(57) to be learned is very restricted in the first place, and that the overarching constraint of the H-α schema will severely restricts the space of hypotheses to be tested against external data (recall §3.5). Word order is something that primary linguistic data provide ample and obvious cues for, so it is plainly reasonable to let the 'second factor' of FL design (external data) take partial responsibility for the learning of linearization rules, thereby reducing the burden of the unexplained 'first factor' (genetic endowment). This is quite a desirable consequence, approaching the MP's desideratum of reducing the number of stipulations in the first factor to its bare minimum.

4.5 Concluding Remarks

In this chapter, various proposals of 'H-α schemata' are reviewed and compared with my projection-free, precedence-free conception of the H-α schema established in Chapter 3. It was observed that Uriagereka's (1999) LCA-based H-α schema, though motivated by his sound and serious criticism of label-based stipulations in Kayne's original LCA, nevertheless fails to achieve the empirical result he had in mind, namely the deduction of the CED. Sheehan's (2009, 2010, in press) unsuccessful attempt at resuscitating Uriagereka's LCA puts things backward, re-importing another set of label-based stipulations while failing to appreciate the importance of Uriagereka's criticism, and thereby incorporates fundamental inadequacies of both Kayne's and Uriagereka's LCAs. We concluded that the LCA is irrelevant to the account of CED effects and has in fact no place in projection-free syntax. While Kayne's (2010) attempt at transposing the idea of antisymmetry into the p-merge theory successfully eliminates recourse to labels/nonterminals and also the LCA, it at the same time fails to appreciate the fundamental desideratum of bare

phrase structure, namely to eradicate precedence from syntax. An alternative approach was briefly articulated in §4.4. While it has no intention of deriving strong antisymmetry (or the questionable universal word order template) from the core architecture of syntax, it is arguably the first theory of linearization that successfully eliminates recourse to labels and c-command while still relegating precedence and linear order to the postsyntactic mapping to PHON. The desideratum of bare phrase structure is thus sustained, and the formulation of Merge is kept in its simplest form, namely just set-formation free from linear order. Furthermore, it may be worth stressing that phase cyclicity plays a vital role in the proposed theory of linearization. The empirical thesis of phase-by-phase derivation is thus correspondingly corroborated as well.

Chapter 5

Minimal Search and Endocentric Interpretation

5.1 Introduction

One of the oldest intuitions in generative grammar is that linguistic expressions in human language are 'endocentric', or 'headed' by an LI within them: in general, there is a certain central LI, called the *head*, for each phrase that determines the essential properties of the phrase at SEM. Thus, a verb phrase is predicative due to properties of its head verb, referential properties of a noun phrase are determined by its head noun, and so on. This intuition can be summarized as in (1):

(1) *Endocentricity (headedness) of phrase structure* :

The interpretive properties of a phrase Σ are determined largely by the features of a prominent lexical item within Σ (referred to as the *head* of Σ).

Throughout the history of the Principles-and-Parameters approach to the study of FL, the intuition has been regarded as the core empirical support for the labeling-by-projection hypothesis, originating with the X-bar theory (Chomsky 1970, Jackendoff 1977 and others).

According to this conception of endocentricity, the head LI 'projects' its lexical features locally up to phrasal constituents (with or without bar-level notations), and that kind of projection path characterizes the centrality of the head LI for the phrases. For example, thanks to X projecting its features through the X'-node up to the topmost XP-node, the phrasal constituent in (2) counts as the maximal projection of X, inheriting the lexical properties of X.

(2)
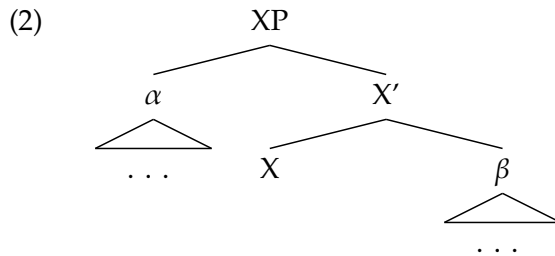
However, if we pursue the line of approach articulated in the previous chapters that syntax assumes no mechanism of projection, then as a matter of principle we can no longer attribute the notion of endocentricity/headedness to projection/labeling. In projection-free syntax, the structure corresponding to (2) is nothing more than a set-theoretic object in (3), where no representational trick makes such notions like 'head', 'complement' or 'specifier', available to the computation of narrow syntax.

(3)   a. $\{\alpha, \{X, \beta\}\}$

     b.
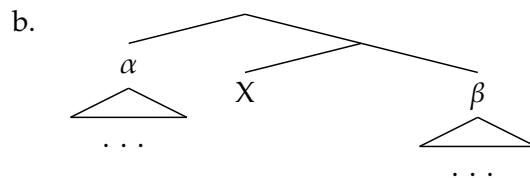
Therefore, so long as we pursue projection-free syntax, and so long as we regard the generalization of endocentricity (1) as a real fact about human lanugage, then some alternative characterization of the notion of 'head' is needed.

     As noted at the beginning of Chapter 3, the present approach to projection-free syntax is in part motivated by the desire to capture the effect of endocentricity of phrase

structure without recourse to labeling/projection. In the present chapter, I will specifically argue that the mechanism of head-detection can be reduced to minimal search of an LI for each phrase. I refer to the proposed mechanism, which is adapted from a different framework proposed by Chomsky (lectures at MIT in fall 2010), as *Minimal head detection* (MHD):

(4)   ***Minimal head detection (MHD)*** :

The head of an SO Σ is the most prominent LI within Σ.

The effect of MHD is essentially that for any SO {H, α}, where H is an LI and α an SO, H is the head of {H, α} (cf. the first clause of Chomsky's 2008 labeling algorithm; see also Piattelli-Palmarini et al. 2009:52ff). As the formulation of MHD clearly shows, no notion of labeling/projection is implied in this minimal theory of endocentricity, thus keeping to the SMT. Minimal search of LIs involved here may be an effect of computational efficiency or, in other words, the third factor of language design (Chomsky 2005), as I will argue in what follows.

The purpose of the present chapter is to explore how MHD can provide a minimal but sufficient account of the effect of what has been attributed to labeling by projection. It will be argued that the characterization of endocentricity in terms of MHD will sometimes depart rather dramatically from the traditional conception of heads and labels, most notably in the treatment of specifiers. It will also be argued that MHD will constitute an important basis for addressing the fundamental question of why the effect of endocentricity arises at all in the interpretation of SOs generated by Merge-based syntax.

5.2   Why Endocentricity?

The proper treatment of the mechanism of labeling has been one of the most widely discussed research topics in contemporary linguistics. Countless proposals have

accumulated in the literature, and a lot more are in press now. Despite the fact that labeling/projection is a departure from the SMT and hence should have no place in bare phrase structure (Chapter 3), there has been a strong impetus to defend labeling/projection, rooted in the important and justifiable conjecture that the theory of phrase structure should capture the old intuition about endocentricity (1): it is simply an ordinary fact about language that 'noun phrases' are interpreted in a 'nouny' way, while 'verb phrases' in a 'verbal' way, and likewise phrases 'headed by' P, A, C, T, etc., are all interpreted differently. Whatever theoretical analyses one may entertain to describe this intuition, the prominent effect of endocentricity on the CI-side of interpretation seems to be simply a matter of fact.

Therefore, we are correspondingly interested in asking the following *why*-question:

(5)   Why does the effect of endocentricity/headedness (1) arise at SEM?

What (5) is calling for is an *explanation* of endocentricity/headedness, not so much a description of it. As we repeatedly mentioned, countless descriptions of endocentricity have been put forward since the X-bar theory (Chomsky 1970, Jackendoff 1977 *et seq.*), making use of analytical tools such as labeling, projection, probe-goal relations, semantic- or categorial-selection, and so on. But these previous proposals have largely failed to address the *why*-question, seemingly because they have regarded labeling by projection as an axiom of syntactic analysis and thereby trivialized that very *why*-question as a matter of course. Departing from this tradition, then, what can we say about a possible explanation of endocentricity?

Importantly, the minimalist program (MP) offers a ready-made answer to such a question, namely (6):

(6)   Endocentricity/headedness arises at CI because it is the simplest mode of interpre-
         tation.

There is little doubt that this is the desideratum that we are led to defend in the pursuit of

the SMT. The question is how we can make sense of this answer.

In order to tackle this problem, let us consider the relation between syntax and CI. Syntax is the component of FL that generates SOs, and CI is the performance system that receives these SOs (via SEM) and exploits them for human thoughts, in an efficient and optimal way, so long as the SMT holds. Among other things, an important consequence of such efficient exploitation of SOs by CI should be what is often called the *strict compositionality* of semantic interpretation. Simply put, it holds that the formal semantic property of a linguistic expression is contingent on those of its parts. Captured in terms of bare phrase structure, then, it should be that the interpretation assigned to an SO $\{\alpha, \beta\}$ at SEM is determined in some way or another by the properties of α and β. This is arguably an optimal way of assigning interpretation to phrasal SOs, rooted in computational efficiency, and this should be why syntax generates structured phrases at all: otherwise, there is little reason for syntax to generate structures to begin with.

In order to assign compositional interpretation to each SO $\{\alpha, \beta\}$, then, CI will need to 'inspect' semantic properties of α and β. Note that each of α and β may be another set-theoretic object $\{\gamma, \delta\}$, whose semantic properties are then determined by γ and δ. Then, the relevant inspection may go on, recursively, all the way down to an 'atomic element' whose semantic properties are not dependent on any smaller entities. These elements are what we called *lexical items* (LIs), acquired from radically impoverished linguistic data from the surrounding environment and stored in the mental component called the *Lexicon*. The uniformly rapid and successful acquisition of LIs is in itself a quite marvelous fact about human infancy, but maybe rendered somewhat manageable by the help of such analytic means as the H-α schema (recall §3.5). In this framework, CI can start assigning interpretation to each SO Σ only after it inspects the internal constitution of Σ and finds an LI that can serve as an 'axis' for compositional interpretation. Presumably, then, such inspection of LIs should be minimized, for the sake of CI's efficient utilization of each SO.

I would like to maintain that the strongest, across-the-board minimization of the

LI-search can be efficiently achieved by the H-α schema:

(7)   ***The H-α schema***: Merge(H, α) → {H, α}.

Merge must take at least one LI as its input.

This derivational constraint, thoroughly supported by various empirical data (Chapter 3), simply follows from the properties of EFs (see §3.4), and it guarantees that every SO takes the form {H, α}, for which an LI can be efficiently detected by minimal search. Notice also that (8) is minimally required from the principle of strict compositionality:

(8)   ***Semantic features of LIs***:

Features of each LI H contain, at the very least, instructions for CI regarding how H contributes to the interpretation of the SO it is embedded into, the smallest one of which is {H, α}.

We may refer to such features of LIs as *semantic features*.[1] Then, the H-α schema (7), working in tandem with the semantic features of LIs (8), guarantees (9):

(9)   Each phrasal SO {H, α} can efficiently receive compositional interpretation at SEM via the semantic features of H.

This much is empirically necessary for any theory of SEM, strictly adhering to the SMT. And this minimal assumption is exactly what MHD articulates:

(4)   ***Minimal head detection (MHD)*** :

The head of an SO Σ is the most prominent LI within Σ.

The effect of headedness in terms of MHD was left open in the earlier chapters, but now we can give well-defined theoretical content to the notion of head in MHD: the head of

---

[1]See McGilvray (1998, 2002, 2009) and Pietroski (2005, 2008 *et seq.*) for discussion of the conception of semantic features (or 'concepts') as instructions for CI.

an SO Σ is the LI that determines compositional interpretation of Σ at SEM by means of its semantic features. Therefore, as long as syntax is under the strict government of the H-α schema, generating only SOs of the form {H, α}, all such SOs satisfy the overarching principle of Full Interpretation (FI), in the form of (9):

(10)   *Full Interpretation (FI)* :

   Every constituent of SEM and PHON contributes to interpretation.

In this line of reasoning, then, endocentricity/headedness of phrase structure arises at SEM simply as a result of optimal exploitation of SOs via strict compositionality, and further, SOs generated by syntax are all structured in such a way that they can efficiently receive compositional interpretation by minimal search of head LIs' semantic features, which is made possible by the H-α schema, serving for full satisfaction of FI.

      This line of reasoning is what I take to be a promising way of making sense of the minimalist answer (6) to the *why*-question (5): endocentricity is reduced to just a cover term for the simplest mode of compositional interpretation. Moreover, the H-α schema makes the satisfaction of strict compositionality constant and consistent, maximally respecting FI.

5.3   Phasing out Specifiers

5.3.1   Minimal Search and Specifiers

      We have seen that MHD is reducible to CI's efficient exploitation of strict compositionality, and the H-α schema paves the way for constant satisfaction of FI, maximally respecting the SMT. Then, the MP is led to pursue this line of analysis (or reanalysis) of the effect of endocentricity, which has been erroneously attributed to labeling by projection in the earlier approaches.

      Notice that the notion of complement, or the structural relation referred to as 'head-complement', is relatively easy to capture (or reformulate) in terms of our projection-

free syntax, as long as the H-α schema gives rise to uniform H-α branching of linguistic constituents. The essential property of head-complement structures, according the traditional view, lies in the asymmetry between the relevant sisters: the head is an LI, whereas the complement is a non-LI/phrase. If we keep to that much, then the H-α schema really is a constraint whereby syntax always generates 'head-complement' structures, informally speaking. Hence, MHD unambiguously singles out the relevant LI H as the head of {H, α}, whose semantic features (8) serve as the prime agent for the compositional interpretation assigned to {H, α}.[2]

The properties of head-complement structures are easily explained by MHD. Other cases are less straightforward, though. At this point, it should be noted that MHD, when combined with the H-α schema, makes some predictions that are quite different from the ones made by traditional theories of labeling by projection. One of the most notable differences will arise for cases of 'specifier'-merger: to take a representative example, consider cases where IM moves Y to the 'specifier' of X, creating an SO of the form:

(11)   a.   {Y, {X, [... t_Y ...]}}.

      b.
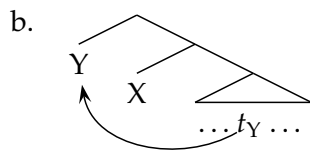
Examples are A'-moved Q and A-moved K.

(12)   a.   {Q, {C, [... t_Q ...]}}.

---

[2]One may alternatively entertain a derivational approach to the definition of head-complement, which holds that for each SO Σ, an LI H is the head of Σ if Σ is formed by Merge by utilizing the EF of H. However, for CI and SM to utilize the information that an instance of Merge utilizes the EF of a particular LI, it seems generally necessary to somehow reconstruct the entire derivational history of a phase at SEM and PHON, a large computational load that seems undesirable. The load may be eased if each application of Merge 'marks' the relevant EF in some way or another upon utilization, but such marking would be a departure from inclusiveness and also no-tampering (see Chapter 2). Incidentally, this is one of the reasons why I was forced to depart from my earlier approach to endocentricity entertained in Narita (2009a), where it was proposed that the effect of endocentricity arises as a result of Agree applying to EFs of LIs, which would involve some sort of marking on EFs upon Agree.

b.

(13)   a. $\{K, \{T, [\ldots t_K \ldots]\}\}$.

b.

For such configurations, MHD will unambiguously identify the moved element as the head of the structure. Thus, it is the WH-moved Q, not C, that will be the head of $\{Q, \{C, [\ldots t_Q \ldots]\}\}$, and it is the EPP-raised subject K, not T, that will be the head of $\{K, \{T, [\ldots t_K \ldots]\}\}$, and so on.

More generally, MHD and the H-α schema fundamentally predict that there is no notion of 'specifier' available in the theory of minimal endocentricity (cf. Hoekstra 1991, Starke 2004, Jayaseelan 2008, Chomsky 2010b,c and Lohndal in progress). Notice that the earlier concept of the X-bar-theoretic 'projection path' not only representationally signifies the centrality of the head LI for the relevant phrases, but also singles out structural notions like 'head', along with 'Spec(ifier)' and 'complement'. For example, α qualifies as the specifier of X, and β the complement of X for the labeled phrase-marker in (14).

(14)

This way, the label-based characterization of headedness immediately gives rise to structural relations like 'head', 'complement', and 'specifier', which I take to be one of the fundamental functions that has been attributed to labeling/projection:

(15)   Labeling by projection defines structural notions like 'head', 'complement' and

'specifier' for each phrasal SO.

However, if syntax assumes no mechanism of labeling by projection, the structure corresponding to (15) is nothing more than the set-theoretic object in (16), where no representational trick is immediately available to characterize such notions like 'head', 'complement' or 'specifier'.

(16)   a. $\{\alpha, \{X, \beta\}\}$

         b.
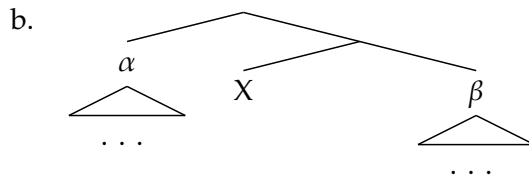
What is more, the H-α schema excludes the possibility of 'XP-XP' merger as a matter of principle. Then, it must be the case that α can never be phrasal. Therefore, a more straightforward representation of the SO corresponding to (16) would be (17), where Y represents an LI:

(17)   a. $\{Y, \{X, \beta\}\}$

         b.
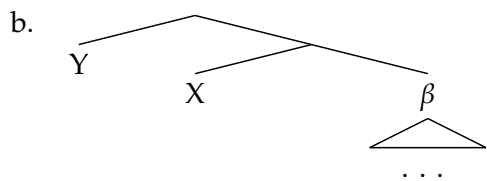
Nothing will prevent MHD from singling out Y as the head of this SO, and only stipulations can ensure that the structure is headed by X, containing Y as the specifier of X. Correspondingly, it should be concluded that the minimal theory of endocentricity provides no room for the notion of 'specifier', a conclusion that has been envisaged and investigated independently by Hoekstra (1991), Starke (2004), Jayaseelan (2008), Chomsky (2010b,c) and Lohndal (in progress) on different grounds.

It is instructive to attend closer to the way the notion of specifier has been used in traditional linguistic analysis. As far as I can see, the following can be pointed out as the major function of the 'specifier-of' relation:

(18)   *The specifier stipulation*:

The specifier of an LI H is an SO that is structurally higher than H, but still marked as subsidiary to (or 'governed by') H in the computation of SEM.

Various effects have been attributed to this hypothesis, some of the major ones being listed in (19).

(19)   If an SO Ω is the specifier of an LI H,

a.   *Pied-piping*: Ω is to be pied-piped by the attraction of H.

b.   *Selection from above*: An LI X externally merged with $\{\Omega, \{H, \ldots \}\}$ will be able to select the categorial (and sub-categorial) feature of H.

c.   *Spec-head licensing*: Some special relation is to be established between H and Ω (morphological agreement, theta-marking, etc.).

However, we may question if the specifier stipulation (18) really receives empirical support from considerations in (19).

5.3.2   Specifiers in Pied-piping

Notice that we can immediately discard the relevance of pied-piping (19a) from our discussion: recall the discussion in Chapter 3 that the H-α schema dispenses with the notion of pied-piping as well as XP-movement altogether from the theory of syntax. Their effects instead arise just as a consequence of cyclic derivation by phase. In the pursuit of the H-α schema, the analysis of pied-piping sometimes need to take advantage of covert higher phase-head LIs like K(ase) (Bittner and Hale 1996a,b, Neeleman and Weerman 1999,

Asbury 2008, Caha 2009 and references cited therein; cf. Chomsky's 2007a *n**) and Q (Cable 2007, 2010). This may at first glance strike us as a kind of worrisome cartography expansion. However, recall from §3.5 that once infants are armed with a strong analytical tool like the H-α schema, the acquisition of such LIs from external linguistic data will be a relatively easy task to accomplish: most importantly, the task is free from the complication that the previous X-bar-theoretic approach requires: the child can detect and acquire such LIs without assigning to them any extraneous notions like 'head of X' or 'specifier of X'. Various cues are available in overt evidence, and all the child needs to find is either (i) overt morphological realization of such LIs, or (ii) overt movement of the relevant (apparently phrasal) constituent. For K, the grammar of English signifies its presence by movement, whereas languages like Japanese also choose overt realization of case-particles:[3]

(20)   *English*:

    a.   [K *the boy*]$_i$ t_i hit Mary.

    b.   [K *the boy*]$_i$ was called t_i by his teacher yesterday.

(21)   *Japanese*:

    a.   [sono otokonoko -**ga**$_K$] [Mary -**o**$_K$] tataita.
        that boy   -NOM Mary -ACC hit
        'That boy hit Mary.'

    b.   [Mary -**o**$_K$]$_i$ [sono otokonoko -**ga**$_K$] t_i tataita.
        Mary -ACC that boy   -NOM   hit
        'That boy hit Mary.'

For Q, English again resorts to movement to signal its presence, whereas languages like Tlingit also assign overt morphology to it.

(22)   a.   [**Q what**]$_i$ did you see t_i?

      b.   [**Q whose mother**]$_i$ did you see t_i?

---

[3]See note 17 of Chapter 3.

       c.   [**Q at which station**]*ᵢ* did you see John *tᵢ*?

(23)   *Tlingit*: Wh-pied-piping structures (Cable 2010)

       a.   [**Aadóo** yaagú **sá**] ysiteen?
            who   boat   Q you.saw.it
            'Whose boat did you see?'

       b.   [**Aadóo** teen **sá**] yeegoot?
            who   with Q you.went
            'Who did you go with?'

       c.   [**Daakw** keitl **sá**] asháa?
            which   dog Q it.barks
            'Which dog is barking?'

       d.   [[**Wáa** kwligeyi] xáat **sá**] i   tuwáa sigóo?
            how it.is.big.ʀᴇʟ fish Q your spirit it.is.glad
            'How big a fish do you want?' (lit. 'A fish that is how big do you want?')

Thus, these functional categories are compliant with what Fukui and Sakai (2003:327; see also Thráinsson 1996) call the *visibility guideline for functional categories*, which holds that functional categories have to be visible (i.e., detectable) for learning in the primary linguistic data, simplifying the problem of learnability, and restricting the number of functional categories. To the extent that postulation of these LIs is reasonable, the problem of pied-piping of specifiers becomes less of a concern.

5.3.3 Selection and Specifiers

      The specifier stipulation was also used in earlier labeling-by-projection approaches as a way to capture the following effect on selection:

(24)   *Selection from above*: If an SO Ω is the specifier of an LI H, an LI X externally merged with $\{\Omega, \{H, \ldots \}\}$ will be able to select the categorial (and sub-categorial) feature of H.

For example, a V *touch* is lexically specified as taking a [+concrete] object as its Theme argument, and if this V is externally merged with the object [[*the government's*] *spokesman*], it was said that V could somehow 'reach' the relevant lexical feature of the N *spokesman*, even though it has an [–concrete] NP as its specifier. The effect is even more striking when the NP contains modifiers (*the government's three intelligent spokesmen*). In a similar vein, it was said that Vs like *ask, wonder, . . .* are lexically specified as selecting a [+WH]-marked CP, *think, assume, . . .* a [–WH]-marked CP, and *know* a CP without specifying [±WH] (see Pesetsky (1982) *et seq.*). More generally, selection and subcategorization used to make major recourse to (then narrowly syntactic) label- or projection-based headedness in earlier theories of UG.

However, the relevance of labeling/projection to the issue of subcategorization has been reduced, essentially since the proliferation of functional categories, such as the DP-analysis of nominals (Brame 1981, 1982, Fukui and Speas 1986, Fukui 1986/1995, Abney 1987) further refined by Grimshaw (Grimshaw 1991/2005 *et seq.*), the fine structure of the left periphery (Rizzi 1997 *et seq.*), and all the later expansion made by the Cinquean cartography project (Cinque 1999, 2002, Rizzi 2004).[4] In the modern framework, it is generally assumed that c(ategorial)-selection is reducible to s(emantic)-selection (Pesetsky 1982), and that selection (categorial or semantic) plays virtually no role in narrow syntax (Chomsky 2004:112-113). Considerations about differing acceptability judgments are really pointing to the same conclusion: occasional violation of selectional restrictions is a common fact about ordinary language use, and it often does not result in clear-cut unacceptability (see Chomsky 1955/1975, 1965, Chomsky 2004:112, Ott 2009, 2010, Boeckx to appear for much relevant discussion of the notion of degrees of grammaticalness),[5] which suggests that se-

---

[4]See also Grimshaw (1991/2005) *et seq.* for the notion of 'extended projections' that is put forward as a way to deal with the selection problem in the framework of expanded cartography. But if syntax is projection-free, it will not provide any room for extended projections, much less feature-percolation within them, as a matter of principle.

[5]Compare it with, e.g., the failure in syntactic subject-verb agreement: *you is a hero*, *they loves Mary*, and so on.

lection is by and large a matter of extra-grammatical concern. To the extent that the theory of syntax can eliminate recourse to selection, the relevance of the specifier stipulation is correspondingly undermined. Presumably, CI just retroactively assigns selectional interpretations to whatever SOs it receives from syntax, yielding various degrees of deviance.

However, I would like to remark on one prominent residue of selection from above: the C-T dependency. It is generally acknowledged that the finiteness and tense properties of T strongly depend on the properties of the C that 'selects' T from above. Thus, the form of modals in T is largely determined by the type of C (and indirectly the verb that selects such clauses), as in (25). Also, it is generally assumed that a tenseless infinitival is a clause without C, exemplified by the raising infinitival in (26b) and the ECM infinitival in (27b), and that these clauses' defective tense properties result from a lack of C that selects these Ts from above.

(25)   a.   John demands [*that*$_C$ his son *be*$_T$ brought to the emergency room]

          b.   John thought [*that*$_C$ his son *was*$_T$ sick for a while].

(26)   a.   It seems [*that*$_C$ his son *was*$_T$ brought to the emergency room]

          b.   His son seems [$_{TP}$ *to*$_T$ have been brought to the emergency room]

(27)   a.   John believes [that$_C$ his son *is*$_T$ a genius]

          b.   John believes [$_{TP}$ his son *to*$_T$ be a genius]

Consider also the distinction between control infinitivals and raising/ECM infinitivals, which has been predominantly characterized in the literature by the assumption that control infinitivals are CP whereas raising/ECM infinitivals are TP (Chomsky 1981; see also Chomsky and Lasnik 1977, Stowell 1981, Bresnan 1972, Martin 2001).

This is by no means a language-specific property of English, and there are countless other observations which seem to point to the same conclusion: T's properties strongly depend on C. This C-T relation has been characterized in numerous ways, most notably by

assuming that some special selectional relation is established between C and T (c-selection, government, etc.). Importantly, the relevant relation cannot be something that arises post-syntactically, given that it has various syntactic consequences. For example, the presence of subject-verb agreement features ($[u\varphi]$) on T and that of Nominative Case on the subject that T agrees with are determined by the presence of C. In light of these phenomena, it is not reasonable to expect that we can reduce such prominent and ubiquitous syntactic C-T relations to postsyntactic interpretive factors.

It should be noted in this context that the 'specifier' of T is prototypically occupied by the A-moved subject KP in languages like English. If we draw the traditional labeled phrase structure for the TP-configuration, then, it may first appear that such KP-raising would not interfere with the close proximity of C and T, given that the complement of C is anyway 'lableled' by T. This is in fact what has been thoroughly assumed in the past literature, hinging on the specifier stipulation.

(28)

```
        CP
       /  \
      C    TP
          /  \
        KP    T'
         ↑   /  \
         |  T    vP
         |      /  \
          \   ...t_KP ...
```
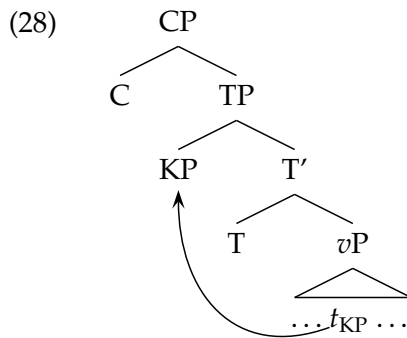
However, the theory of bare phrase structure cannot blame the C-T relation on labeling by projection anymore. Then, how can we make sense of the strong syntactic bond between C and T, without making recourse to labeling by projection or the specifier stipulation?

For that matter, recall the hypothesis established in Chapter 2 that applications of EM are required to precede IM and all the other operations applying at the phase level. Specifically, it was assumed that EM of T and C precedes the EPP-driven A-movement of the subject K. Due to the no-tampering condition (NTC), the relevant internal merger of K

yields the following two-rooted structure at the C-phase level.

(29) a.   (= (45), Chapter 2; (63), Chapter 4)

C
[Nom]   T
        [$u\varphi$]   K   v*
              [uCase,$v\varphi$]   [Acc,$v\varphi$]

b. $\longmapsto$ c.

C   K   Transfer   C
[Nom]   T   [Nom]
        [vφ]   K   v*
              [Nom,$v\varphi$]   [Acc,$v\varphi$]
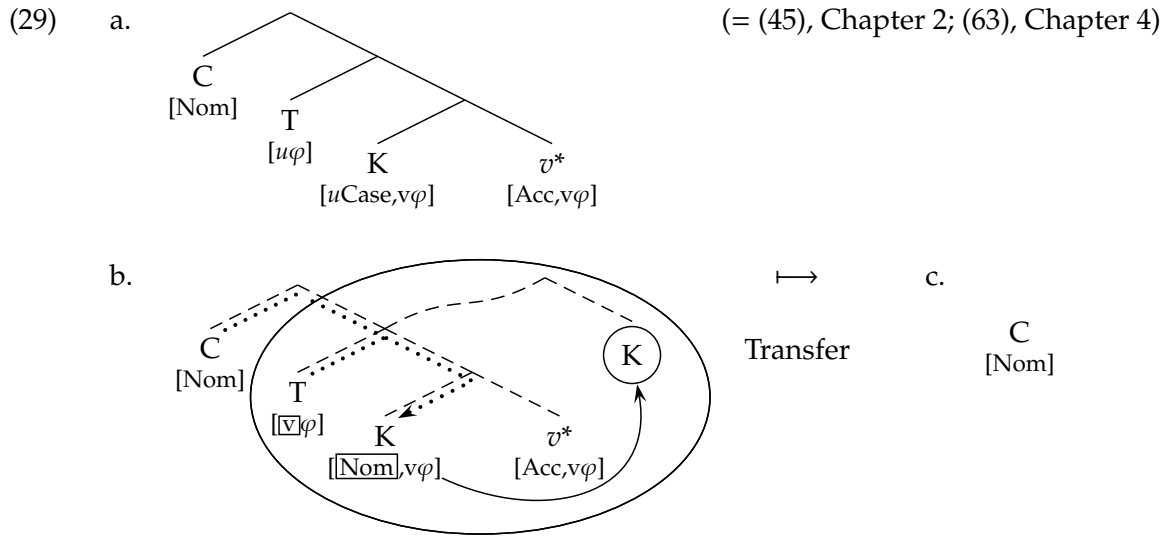
This structural representation radically departs from the traditional conception of labeled phrase structure. The internally merged 'specifier' K and the externally merged C are 'on different planes', so to speak, and they are both as close to T as one can get. Thus, the relevant syntactic relation between C and T can be naturally established by structural minimality, probably by means of 'c-selection feature-checking' or Agree.

Chomsky (2007a, 2008) and Richards (2007b) claims that the sort of structural dependency between C and T is further generalized, and that it holds between a given phase head and the next lower non-phase head in general. Among other things, they claim that the relation between *v** and V (accompanying object agreement morphemes in some languages), as well as the one between a categorizer head (*v*, *n*, *a*) and a root LI $\sqrt{root}$,[6] should be analyzed in the same vein, explaining their strong affinity. If we adopt their hypothesis, then the same account sketched above can be strengthened to unify the accounts of these c-selection-like relations, a desirable result.

---

[6] See Halle and Marantz (1993, 1994), Marantz (1996, 1997, 2007) among others for the theory of Distributed Morphology.

Given these considerations, we can safely conclude that the specifier stipulation is plainly unnecessary for the explanation of 'selection from above'. Such selectional dependencies are readily reanalyzed either as postsyntactically assigned semantic interpretations, or as a result of minimal search in cyclic derivation by phase.

5.3.4 Spec-head Licensing

Finally, let us remark on the final residue of the specifier stipulation (30)

(30) *Spec-head licensing*: If an SO Ω is the specifier of an LI H, some special relation is to be established between H and Ω (morphological agreement, theta-marking, etc.).

As far as so-called 'spec-head agreement' phenomena are concerned, evidence is accumulating for the view that such spec-head configurations should be treated as resulting from some derivationally prior probe-goal relationship. To take the φ-feature agreement between the EPP-raised subject and T as a representative example, it is now standardly assumed that this relation is established by T's $[u\varphi]$ probing the $[v\varphi]$ of the in-situ subject K, which in some languages is accompanied by A-movement of the subject to the Spec-T position, as we saw for English in (29). The apparent relation between the spec and the head of 'TP' then is just illusory, reducing to Agree with respect to φ-features and ancillary internal merger. Pushing this line of approach to its limit, Chomsky (2004:113) puts forward the empirical thesis that all apparent spec-head relations are illusory and are to be reduced to probe-goal relations involving minimal search. Taking advantage of various pieces of evidence in favor of this empirical thesis accumulated in the literature, I take it for granted that internally merged specifiers in themselves have no special relations to an LI embedded in their sisters. That is, the higher occurrence of X in {X, {H, { ... X ... }}} has no special relation to H—it is just merged with {H, { ... X ... }}, not with H (contra Kayne's 2010 treatment of specifier-merger; see §4.3.2).

However, given that the argument above holds only for internally merged specifiers, there still remains the problem of the apparent relation between a head and an externally merged specifier. One of the most prominent examples concerns the relation between v^* and the external argument KP (EA). It is assumed that v^* is the locus of the Agent or Causer θ-role (henceforth the external θ-role for short), and that v^* assigns this θ-role to the EA occupying the (innermost) specifier of v^* (Chomsky 1995, Kratzer 1996, 2000; see already Marantz 1984; see also Hale and Keyser 1993, 2002). This was relatively an easy problem in the earlier X-bar-theoretic approach, where the 'projection of v^*' governs the external argument, as in (31):

(31)
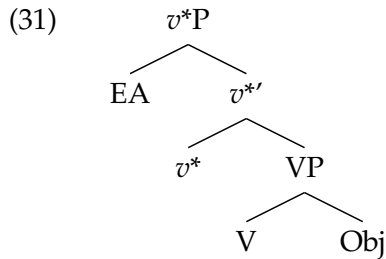
Again, however, such a representational trick with labeling/projection has no place in the theory of bare phrase structure.

However, recall the derivation of the relevant merger of the EA and v^*:

(32)   a.
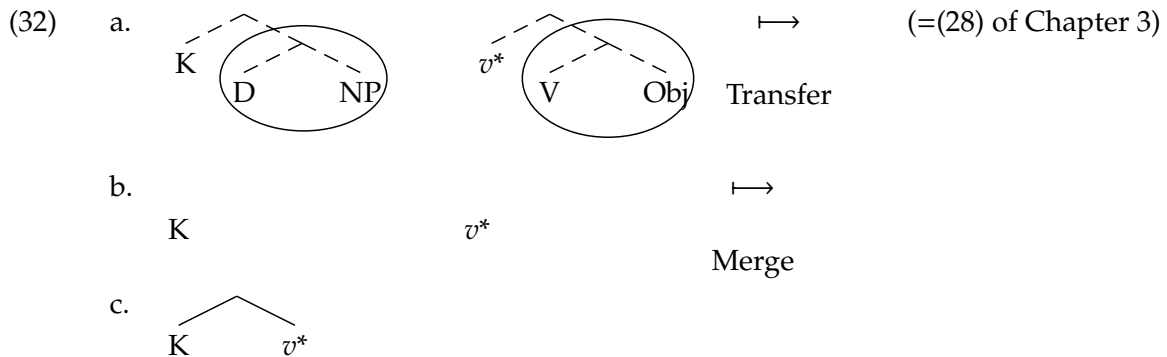
We saw in Chapter 3 that phase-by-phase Transfer effectively reduces $\{v^*, \{V, Obj\}\}$ to v^*, and thus at the point of EM (32b), v^* is just merged with the EA (or more specifically the edge of the relevant K-phase left in the active workspace). Then, in a certain sense, the EA

occupies the 'second complement' of v^*, forming {EA, v^*}.

I would like to maintain that this positioning is sufficient to guarantee the θ-role assignment from v^* to the EA. To be concrete, I propose that v^* contains the following semantic features (33):

(33)   *Semantic features of v^**:

If v^* is singled out as the head of {v^*, α},

a. v^* assigns a verbal interpretation to α in {v^*, α}.

b. v^* assigns the external θ-role to α in {v^*, α}, under the condition that α is headed by K.

The effect of (33) is generally taken for granted in the current literature. (33a) takes care of the semantic interpretation assigned to {v^*, {V, Obj}}. I assume that this feature is shared by v in general, so passive and unaccusative v also assume the semantic feature (33a). The difference between v^* and v then lies in that only v^* assumes (33b) in addition to (33a). The effect of (33b) arises when its sister SO is headed by K—basically the case where v^*, after Transferring VP, takes the EA as its second complement. Thus, provided that MHD singles out v^* as the head of {v^*, EA}, K receives the external θ-role from v^*, the target result. This way, the assumption in (33) is sufficient to capture the effect that was once attributed to the specifier stipulation.

To conclude this section, then, no considerations in (35) lend support to the specifier stipulation (34) anymore.

(34)   *The specifier stipulation*:

The specifier of an LI H is an SO that is structurally higher than H, but still marked as subsidiary to (or 'governed by') H in the computation of SEM.

(35)   If an SO Ω is the specifier of an LI H,

a.   *Pied-piping*: Ω is to be pied-piped by the attraction of H.

b.   *Selection from above*: An LI X externally merged with {Ω, {H, . . . }} will be able
     to select the categorial (and sub-categorial) feature of H.

c.   *Spec-head licensing*: Some special relation is to be established between H and
   Ω (morphological agreement, theta-marking, etc.).

Therefore, we can safely discard the specifier stipulation (34). This amounts to the elim-
ination of the notion of specifier altogether from the theory of syntax, a conclusion inde-
pendently argued for by Starke (2004), Jayaseelan (2008), Chomsky (2010b), Lohndal (in
progress) among others on different grounds.

5.4   The Bifurcation of External and Internal Merge Revisited

Now, the total elimination of 'specifier-of' paves the way for keeping to the sim-
plest possible formulation of endocentricity/headedness, namely the one based on MHD.
As MHD constitutes arguably a necessary component of our principled explanation of
endocentricity under development (see §5.2), the MP is led to welcome this result, and
examine various ramifications that MHD gives rise to.

At this point, recall that MHD, when tied to the H-α schema, makes some pre-
dictions that are quite different from the ones made by traditional theories of labeling by
projection. One of the most important differences will arise for cases of IM: since IM can
affect only LIs due to the H-α schema, any SO resulted from IM will have the moved LI as
its head. Thus, if IM moves Y to the edge of X, creating an SO of the form:

(36)   a.   {Y, {X, [. . . t_Y . . .]}}.

       b.
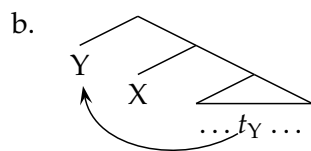
MHD will unambiguously identify Y as the head of the structure. Thus, it is the WH-moved Q, not C, that will be the head of {Q, {C, [... t_Q ...]}}, and it is the EPP-raised subject K, not T, that will be the head of {K, {T, [... t_K ...]}}, and so on.

Although this conception of heads for SOs created by IM rather dramatically departs from the traditional labeled phrase structure, no theory-internal considerations speak against this conclusion, as we saw above.[7] Probably, then, the notion of head as determined by MHD may be reduced in this approach to just a hierarchically prominent LI that primarily configures CI-interpretation of the SO in question.

In this connection, we reached in Chapter 2 the important conclusion that not only the phase-interior SO itself but also the information regarding copy-formation by IM is subjected to Interpret, via derivational simultaneity of IM with Transfer. Then, it is reasonable to suppose that each H in the constitution of {H, α} may contribute to one of two different kinds of instructions to CI, depending on whether it is introduced by EM or IM. If the head LI H is introduced into {H, α} by EM, it will primarily configure the 'deep structure interpretation' of {H, α}, including s-selection, argument-structure, predication and so on. On the other hand, if H is introduced by IM, it may make sense to regard it as constituting the locus of 'surface structure interpretation' that determines, e.g., the scope of relevant operator-variable linkages between the two occurrences of moved K or Q, among other things. All these ramifications are seemingly welcome, especially in the context of the explanation we are hoping to offer for the effect of endocentricity: the head H of an SO {H, α} is just the LI that primarily determines the compositional interpretation of {H, α} by the medium of its semantic features.

Along this line of reasoning, then, let me provide the two general descriptions of the forms of instructions that MHD-based endocentricity provides. First, as for the 'head-complement' structure formed by IM, I would like to propose the following:

---

[7]See Starke (2004), Donati (2006) and Chomsky (2008) for much relevant discussion.

(37)   ***Endocentric interpretation for IM at SEM***:

{H, α} formed by IM contributes to an operator-variable formation where H scopes over α, and the copy of H introduced by EM is interpreted as a logical variable bound by H.

In a schematic structure in (38) where H, located in the sister of another SO β, moves to the periphery of α, Interpret maps it to SEM where the lower copy of H is interpreted as a variable bound by the higher copy of H (cf. Sauerland 1998), and the semantic features of H will presumably determine the semantics of the logical operator it is mapped to. Here I use the letter Π just as a cover term for the relevant logical operator: it may range over various quantifiers and scopal operators, including, say, the universal quantifier '$\forall$' and the existential quantifier '$\exists$'.

(38)   a. $\longmapsto$ b.
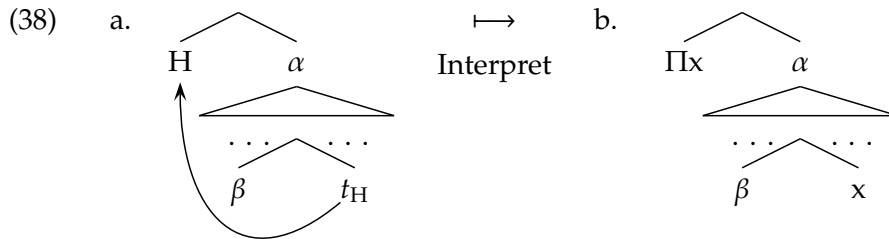
For example, if Π is a universal quantifier '$\forall$x:' and β a V, say *come*, then the interpretation corresponding to (38b) would be something like '$\forall$x: . . . come(x) . . .' If Π is an interrogative operator of the sort determined by the interrogative operator Q, then the interpretation corresponding to (38b) would be something like 'For which x: . . . come(x) . . .?' and so on. It may be that H contains no semantic features that contribute to any specific quantificational interpretation, as is often the case with the A-movement of normal nonquantificational KPs like *the boy* or *John* (although we may also follow the idea of Barwise and Cooper 1981 *et seq.* and assume that all nominal expressions are some sort of generalized quantifier), in which case the relevant interpretation may be simply something like: 'As for H (or speaking of H): . . . come(H) . . .', corresponding to topic or focus

interpretation for H, as I assume.[8]

In the case of EM, no copy-formation is involved, and hence the H-α structure signifies the genuine mode of endocentric interpretation. I propose the following as a general description of the effect of EM-based endocentricity:

(39)   ***Endocentric interpretation for EM at SEM***:

If {H, α} is formed by EM, the core semantic properties of {H, α} are determined prominently by the semantic features of H, and α is interpreted relative to H.

One of the semantic effects that (39) is meant to capture is θ-role assignment by a verbal category H to its sister α. For example, the semantic properties of {V, KP} are largely determined by the intrinsic features of V, and KP is interpreted relative to V, here constituting an integral 'thematic part' of this verbal SEM.[9] More generally, I would like to propose that the following generalization holds for θ-role assignment by verbal LIs in general (cf. Narita 2009a, Epstein 2009):

(40) θ-role assignment by H to α is achieved in the structure {H, α} created by EM.

which I claim is just one particular aspect of (39). Thus, the external θ-role is assigned by v^* to the EA in the configuration {v^*, EA} as we saw in (b) of (33), repeated here.

(33)   *Semantic features of v^**:

If v^* is singled out as the head of {v^*, α},

a. v^* assigns a verbal interpretation to α in {v^*, α}.

b. v^* assigns the external θ-role to α in {v^*, α}, under the condition that α is headed

---

[8]The notion of 'categorical judgment' proposed by Kuroda (1965, 1971, 1972, 1976, 1992, 2005) may turn out to be highly relevant in the unification of these various aspects of 'surface structure interpretation'. Kuroda's theory of judgments is explored in different ways by Raposo and Uriagereka (1995) and Uriagereka (1997, 2008) among others.

[9]Tenny (1987, 1994) expresses an important intuition that the Theme 'measures out' or delimits the event denoted by the verbal phrase. This idea is further explored in Mori (2005) and Uriagereka (2008).

by K.

We may further speculate that the two *prima facie* different instructions of v^* in (33) can be uniformly treated as following from (39). Indeed, if v^*, a variety of the 'verbalizer' category v, is the locus of verbal interpretation and its event semantics (as assumed in theories of Distributed Morphology), it makes sense to think of v^* as 'verbalizing' its complement {V, Obj} (or more precisely {$\sqrt{root}$, Obj}) by virtue of the latter being interpreted relative to v^* via (39). This explains (33a). Moreover, if the EA is further interpreted relative to v^* in {EA, v^*}, whose head v^* is already assigned the integral compositional interpretation of {v^*, {V, Obj}} at the v^*-phase-level Interpret/Transfer, this may be sufficient to give rise to the external θ-role interpretation, explaining (33b). As far as I can see, no empirical considerations speak against this simplistic approach, which follows as a null hypothesis. Thus, I would like to maintain that it is possible to provide a uniform characterization of the semantic instruction by EM as in (39).
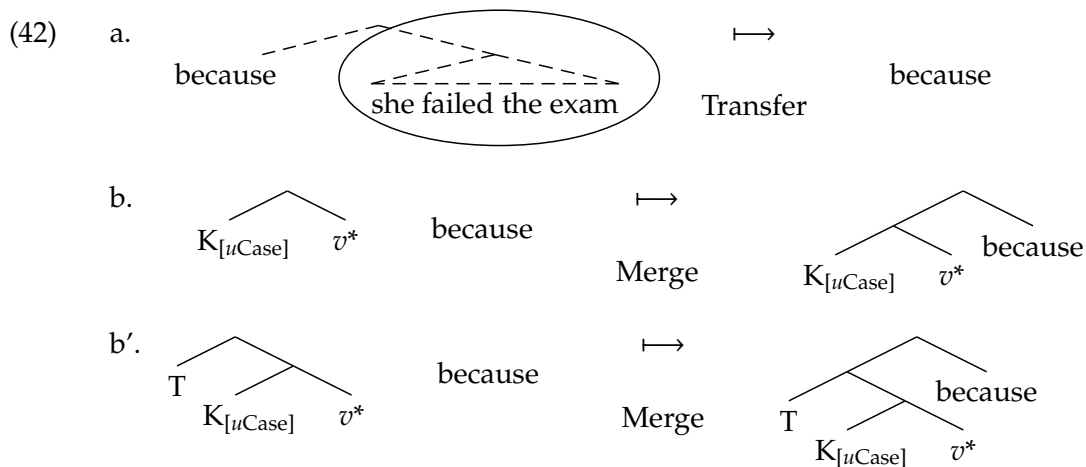
To conclude, we saw that once Interpret is allowed to exploit the bifurcation of EM and IM for the mapping to SEM, we can provide a natural characterization of the pervasive duality of semantics at CI by the minimal search-based conception of endocentricity. I maintain that the two modes of endocentric interpretation ((37) and (39)) arise simply as a result of CI's full exploitation of the bifurcation of IM and EM, which comes for free in the framework of bare phrase structure. The efficient contribution of Merge-based syntax to CI is further corroborated, a desirable result.

5.5   Remarks on Adjuncts

In relation to EM, I would like to make a brief remark on the treatment of adjuncts in the proposed framework. Recall the hypothesis established in §3.3.3 that adjunct-merger obeys the the H-α schema, just like the other cases of Merge. This hypothesis was shown to be necessary to account for the adjunct condition effect in terms of the H-α schema. For
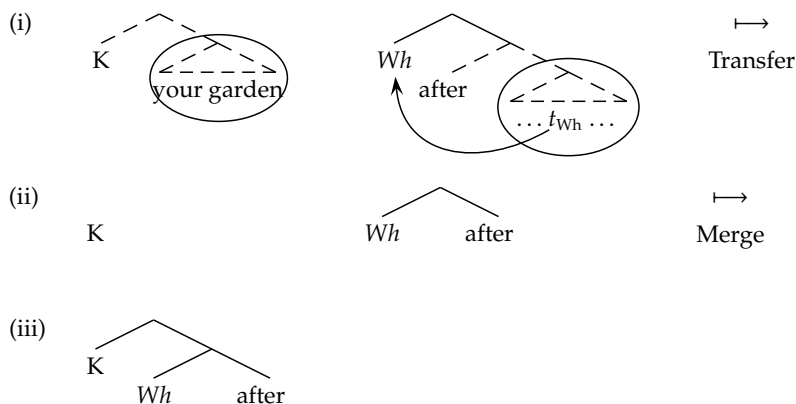
example, it was proposed that high adjuncts, such as *because*-clauses and finite temporal adverbial clauses, are always reduced to LIs by Transfer prior to adjunction, given that the presence of [uCase] on the subject makes the main clausal spine irreducibly phrasal. The relevant derivation is reproduced here.

(41)   The man criticized Mary [because she failed the exam].

(42)   a.

because   (she failed the exam) $\longmapsto$ because
                                     Transfer

b. $K_{[u\text{Case}]}$ v^* because $\longmapsto$ $K_{[u\text{Case}]}$ v^* because
                                    Merge

b'.   T $K_{[u\text{Case}]}$ v^* because $\longmapsto$ T $K_{[u\text{Case}]}$ v^* because
                                        Merge

Now, it should be noted in regard to this analysis that it is the adjoined material that occupies the position of H in {H, α}. This is always true for high adjuncts, and most cases of low adjuncts also conform to this picture, too.[10] Then, MHD makes a prediction different

---

[10]However, in contrast with high adjuncts, there are cases in which low adjuncts like locative PPs and gerundives may occupy the position of α in {H, α}, provided that the main clausal spine can be reduced to an LI H prior to adjunction. In particular, it was proposed in §3.3.3 that a low adjunct can subsume an edge for successive cyclic movement by being located in the sister of the object KP. The following derivation is repeated from (70) of Chapter 3.

(i)   K (your garden)   *Wh* after ($\ldots t_{\text{Wh}} \ldots$) $\longmapsto$ Transfer

(ii)   K   *Wh* after $\longmapsto$ Merge

(iii)   K *Wh* after

from the traditional theory of labeled phrase structure: it is the adjoined material that becomes the head of the output of adjunct-merger. In particular, the semantic rule for EM in (39), reproduced here, holds that the main clausal spine α is to be interpreted relative to the adjunct H in {H, α}.

(39)   ***Endocentric interpretation for EM at SEM***:

If {H, α} is formed by EM, the core semantic properties of {H, α} are determined prominently by the semantic features of H, and α is interpreted relative to H.

As for the example in (41), the prediction is that the matrix v*P or TP, corresponding to the event in which the man criticized Mary, is interpreted in relation to the semantics of the adjunct headed by *because*.

Is this an undesirable prediction? Maybe, or maybe not. Different hypotheses may provide different answers, but it is clear at least that this prediction is a straightforward consequence of MHD and the H-α schema. Therefore, so long as we hope to keep to our principled explanation of endocentricity sketched in §5.2, we just have to swallow this

---

In such a derivation, the low adjunct occupies the position of α in the H-α structure. (39) then predicts that the interpretation of the SO should be determined prominently by the semantic features of the object K, while the adjunct phrase is interpreted relative to K.

There is an interesting set of data that might support this prediction. Consider the following set of examples.

(iv)   A: *Which book$_i$* did John design his garden [after reading t_i]?
         B: An introduction to landscape gardening.

(v)   A: *Which book$_i$* did John design his garden [after reading t_i]?
         B: #Finnegans Wake.

Truswell (2007b) reports that subextraction of a WH-phrase from an adjunct imposes an interesting restriction concerning the assumed answer to the question. In his terms, the event described by the transparent adjunct should be connected to the matrix event by what he calls the *enablement* relation: for instance, the form of question in (iv)-(v) induces the speaker's assumption that reading the relevant book(s) is an event such that it in some way or another enables John to design his garden. This assumption is compatible with, e.g., the answer in (iv), namely *an introduction to landscape gardening*, while it makes answers like *Finnegans Wake* infelicitous. I would like to speculate that this state of affairs may alternatively be characterized as follows, in line with the proposed framework: since the adjunction structure is generated by EM, it is subject to the rule in (39), which requires that the adjunct phrase must be interpreted relative to the object LI K. Then, it is not unreasonable to suppose that the event described by the adjunct is thereby semantically related to the object *his garden*, which may explain the semantic coherency of the matrix event and the adjunct. Though this line of approach obviously requires more elaboration and refinement, I would like to note that the MHD-based theory of endocentric interpretation may be further extended to this *prima facie* semantic observation (cf. Truswell's Single Event Condition discussed in §3.3.3).

conclusion. Note that irrespective of the presence of one or more adjuncts within a phase SO Σ, the endocentric interpretation of Σ is anyway determined prominently by its phase head. For example, the matrix clause is headed by C, so propositionality, finiteness, tense and force of the matrix clause is anyway determined prominently by the phase head C, whether or not it contains one or more adjuncts. Similarly, K, P, v^* and other phase heads also determine endocentric interpretations of their own phases. The fact that a phase contains an intermediate structure headed by an adjunct would thus pose little problem at Interpret, which applies at the phase level. Moreover, roughly speaking, adjuncts are elements that provide background descriptions of the constituent they adjoin to, so it is not obviously unreasonable to state that the interpretation of the matrix is relativized to the semantics of adjoined materials, in line with the broad characterization in (39). All in all, it is not obvious whether we need to withdraw the conclusion forced by the H-α schema and MHD, namely that adjuncts prototypically determine endocentricity.

However, it is of course possible that further empirical considerations will eventually force us to depart from such a null hypothesis, going back to, say, the old idea that adjuncts are invisible for head-detection. One way to approach this problem, while keeping to the empirical strength of the H-α schema, is to suppose that interpretive asymmetry between the main structure and adjuncts is captured by a different source of structural asymmetry other than the LI v.s. phrase distinction. For example, we may adopt Chomsky's (2004) proposal that adjunction is an operation that generates an ordered pair of the two input SOs, $\langle \alpha, \beta \rangle$. Chomsky refers to this kind of Merge as *pair-Merge*, to distinguish it from the ordinary variety of Merge as set-formation (*set-Merge* in his terminology). Pair-Merge creates SOs with a built-in asymmetry encoded by order, which may be sufficient to capture the relevant distinction between the main structure and the adjunct. Again, all we need to assume to sustain the account of adjunct condition effects (discussed in §3.3.3) is that pair-Merge also obeys the H-α schema, presumably because its application is also contingent on EFs.

For the lack of strong empirical evidence for or against them, I will refrain from adopting such additional assumptions, although it is certain that much more research is needed to determine the proper treatment of adjuncts on a finer-grained basis.

5.6 The Strength of Full Interpretation

5.6.1 Strong and Weak Full Interpretation

It should be acknowledged that the MHD-based conception of endocentricity was first effectively proposed by Chomsky (2008) as one of the components of his 'labeling algorithm'. Here I reproduce Chomsky's original formulation of the relevant mechanism of labeling:

(43)   Chomsky's (2008:145,(2)-(3)) labeling algorithm:

    a.   In {H, α}, H an LI, H is the label.

    b.   If α is internally merged to β, forming {α, β} then the label of β is the label of
        {α, β}.

Although it was still misleadingly called the labeling algorithm, it is clear that this conception of endocentricity is meant to make no recourse to labeling by projection. It is clear that (43a) is essentially the effect that MHD is meant to capture.

(4)   *Minimal head detection (MHD)* :

    The head of an SO Σ is the most prominent LI within Σ.

The second clause of Chomsky's labeling algorithm was added to guarantee that in each case of IM the 'target' of movement 'projects', a traditional idea borrowed from the X-bar theory, which has an earlier root in Emonds' (1970, 1976) Structure Preservation hypothesis. But it simply doesn't follow from minimal search in the way (43a) arguably does. Then, so long as the two very different definitions of 'label'/head are incorporated into the theory

of syntax, we will lose the plausible line of explanation for endocentricity sketched in §5.2, namely that endocentricity arises as a result of CI's simplest mode of compositional interpretation, involving only minimal search. This is presumably why Chomsky (p.c., lectures at MIT in fall 2010) departs from (43) and suggests that (43b) should be eliminated from the labeling algorithm, reducing the mechanism of head-detection to minimal search of LIs (see also Piattelli-Palmarini et al. 2009:52ff).

Importantly, Chomsky still assumes in his class lectures that there are instances of 'XP-XP' merger, departing from the H-α schema. As we saw in Chapter 3, it was assumed in the theory of syntax proposed by Chomsky (2008) that edge-features (EFs) of LIs percolate up to phrasal SOs, and thereby SOs can be freely combined by unconstrained Merge. The hypothesis of EF-percolation immediately gives rise to the possibility of XP-XP merger, but I argued in Chapter 3 that such hypotheses of projection and feature-percolation should be eradicated from the theory of inclusive syntax. But as a reply, Chomsky suggests in personal communication that the concept of edge-feature as defined in Chomsky (2008) may be a dispensable notion after all (see also Fukui forthcoming), and that Merge should be truly unconstrained, admitting instances of XP-XP merger.

Now, it should be noted in this context that if there are any SOs of the form {XP, XP}, MHD will fail to determine the head of such structures, and thus it would fail to assign to them any endocentric interpretation at SEM. Then, if we stick to the assumption that endocentric interpretation is the only mode of compositional interpretation that CI can manage to offer, then any such {XP, XP} structure would violate the principle of Full Interpretation (FI), which requires among other things that every constituent structure receive interpretation at CI:

(44)   *Full Interpretation (FI)* :

Every constituent of SEM and PHON contributes to interpretation.

Chomsky actually admits this prediction as such, regarding occasional violations of FI as defined in (44) as unavoidable. He suggests that FI should probably be replaced with somewhat weaker version of it, which we may refer to as *Weak Full Interpretation*:

(45)   ***Weak Full Interpretation*** :

Every constituent whose head can be determined by minimal search (MHD) receives interpretation.

He argues that the possible generation of {XP, XP} structures as well as their occasional failures to receive interpretation should be regarded as a matter of fact. For example, intermediate copies of successive cyclic WH-movement and EPP-raised subjects are regarded by him as *prima facie* instances of such SOs that apparently receive no obvious interpretation at CI.

(46)   The {XP, XP} structure formed by successive cyclic WH-movement:
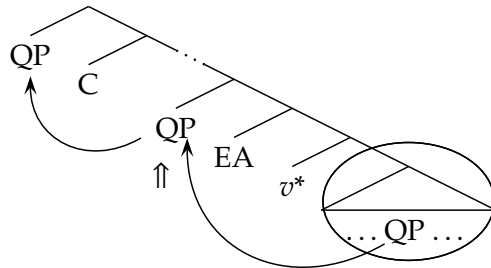
(47)   The {XP, XP} structure formed by EPP-driven A-movement of the subject KP (whether it is represented by 'tucking-in' (a) or by a multi-rooted structure (b)):
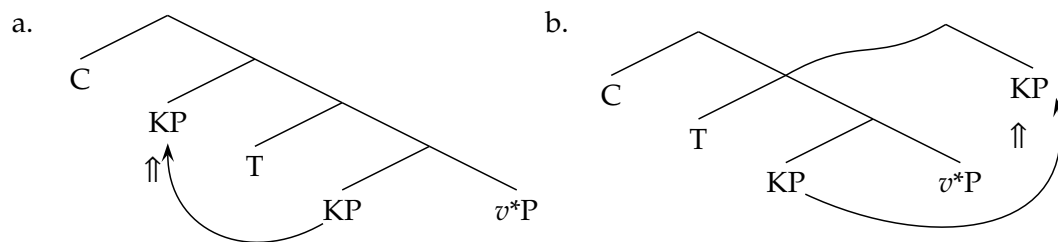
In this line of approach suggested by Chomsky, then, the strength of FI must be correspondingly weakened as in (45).[11]

However, it is not clear whether we really have to withdraw the simplest and hence strongest interpretation of FI as defined in (44) in the face of the examples in (46)-(47). Successive cyclicity of A'-movement arises simply as a result of cyclic derivation by phase. To take the example in (46), it is clear that the intermediate copy of QP is created by IM in order not to subject QP to the v*-phase-level Transfer, and keep it for the computation at CP. Then, the intermediate copy of such movement presumably contributes to 'bridging' the operator-variable relation between the highest and the lowest copies of QP. Moreover, A-movement of the subject KP also apparently contributes to various sorts of discourse-related interpretation, such as 'subjecthood', 'aboutness', discoursal scope, topicality, and so on, though such effects are sometimes blurred and hard to detect (see Ortega-Santos 2008 for some relevant discussion).

Notice further that I argued in §3.4 and §5.2 that once we adopt the H-α schema, it simply follows that syntax *cannot* generate any interface-illegitimate objects of the form {XP, XP} for which MHD fails to determine heads. Syntactic computation governed by the H-α schema is thus 'failure-proof' so to speak, and SOs generated thereby constantly satisfy FI of the strongest sort (44). For the cases in question, I thereby maintain that such instances of IM are not departures from FI (44), and they contribute to the establishment of 'surface structure interpretation' as dictated by (37).

---

[11] Chomsky (p.c.) suggests that there are several ways to salvage XP-XP structures from their 'unlabelable' nature. For example, he suggests that traces of movement are invisible for head-detection, thus an XP-XP structure can be salvaged for head-detection if one of the XPs moves. He speculates that this hypothesis might provide a way to reformulate Moro's (2000) idea of dynamic antisymmetry, according to which symmetric {XP, XP} small clause structures are salvaged by moving one of the XPs in a similar vein. Since the primary purpose of this thesis is to explore the H-α schema and its contribution to FI of the stronger form (44), it will not concern any further the investigation into the ramifications of Chomsky's Weak Full Interpretation and the trace-invisibility hypothesis.

(37)   ***Endocentric interpretation for IM at SEM***:

{H, α} formed by IM contributes to an operator-variable formation where H scopes over α, and the copy of H introduced by EM is interpreted as a logical variable bound by H.

This formulation is straightforward for A-movement of the type in (47). As for successive cyclic A′-movement, which can span over multiple phases, I claim that the intermediate copies of such movement still contribute (though indirectly) to bridging the eventual operator-variable relation between the highest and the lowest copies of the relevant LI, hence not violating the stronger FI. Therefore, the H-α schema, whose empirical import is thoroughly corroborated by the data discussed in Chapter 3, can pave the way for strict adherence to the stronger formulation of FI.

5.6.2   *{*t, t*}

Recall our hypothesis from §3.4 and §5.2 that the H-α schema can be motivated as a derivational constraint that contributes to constant satisfaction of FI (15): so long as syntax generates only recursive H-α structures, MHD can consistently detect the head of each SO by means of minimal search, satisfying the requirement imposed by FI. However, let me at this point note that it is not always the case that the H-α schema alone can ensure the full satisfaction of FI. In this section, I would like to maintain that the following structure is one prominent example of uninterpretable SOs that are excluded by FI (in the strong sense) but not by the H-α schema:

(48)   *{*t, t*}:

CI cannot assign a legitimate interpretation to an SO whose two members are both copies of some moved elements.

If I am allowed to use some informal terminology borrowed from the earlier trace theory of movement (cf. Chomsky 1973, 1975 *et seq.*), the claim in (48) can be summarized as follows: no trace can be a sister of another trace. I claim that it can be explained as a direct consequence of FI (in the strong sense).

Recall the hypothesis that an SO {H, α} constructed by IM always signifies the following instruction for Interpret.

(37)   ***Endocentric interpretation for IM at SEM***:

{H, α} formed by IM contributes to an operator-variable formation where H scopes over α, and the copy of H introduced by EM is interpreted as a logical variable bound by H.

Consider again a structure where H is first externally merged with β and then moves to the periphery of α. The structure is schematically shown in (49a). According to (37), Interpret maps the structure to SEM where the lower copy of H is interpreted as a logical variable bound by the higher copy of H, as in (49b).

(49)   a.
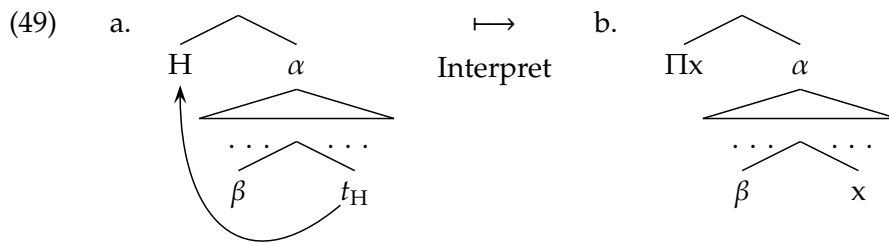
Given this mode of CI-interpretation, it is reasonable to speculate that the following interface condition holds at SEM:[12]

(50)   If copies of H are created by IM, the semantic features of H are interpreted at the highest occurrence of H, where the scope of the logical operator is determined.

---

[12]The effect of this hypothesis is shared by Chomsky's theory of the weaker version of FI (45) (p.c., class lectures at MIT in fall 2010), when he suggests that traces of movement are thence rendered invisible for MHD. See also note 11.

If we assume that much, then (51) follows:

(51)   If H in {H, β} is not the highest occurrence of H created by IM, H cannot define the
         head of {H, β}.

As the notion of headedness arises just as a result of minimal search of LIs' semantic
features, MHD fails to single out traces of IM as heads of any structure, given that their
semantic features are left uninterpreted at trace-positions.

Note that (51) does not necessarily cause problems for intermediate traces of
successive cyclic movement, such as the one in (46) (indicated by ⇑). Recall the conclusion
from §5.6.1 that such traces are only there to 'bridge' the operator-variable linkage between
the highest and the lowest occurrences of the relevant LI H. Then, their presence at edges
of successive phases is still contributing to interpretation and hence does not violate FI.
However, it would become potentially problematic when the relevant trace is the lowest
occurrence of H, which is by definition introduced by EM and converted to a logical variable
at SEM, according to (37).

It is reasonable to suppose that by virtue of being interpreted as a logical variable,
the lowest copy of H is required to sit in a position where it can constitute a logical argument
of some predicative category. Specifically, then, the occurrence of H should be located in a
position where it can receive a θ-role. Recall further that the θ-role of an LI H is assigned
to an SO Σ by EM, forming {H, Σ}, as a consequence of the rule of Interpret for EM (39):

(39)   ***Endocentric interpretation for EM at SEM***:
         If {H, α} is formed by EM, the core semantic properties of {H, α} are determined
   prominently by the semantic features of H, and α is interpreted relative to H.

(40) θ-role assignment by H to α is achieved in the structure {H, α} created by EM.

For the SO in (49a), then, it should be that β is a predicative LI (verb, adjective, etc.) that can be interpreted as θ-marking the occurrence of H it is merged with. In a nutshell, by virture of H being subjected to IM and hence to be interpreted as a logical variable, $\{H, \beta\}$ must be interpreted as β, an LI, θ-marking (being predicated of) a logical variable of H. That is, β should be singled out as the head of $\{H, \beta\}$ in such a structure, invoking the Interpret rule (39). This amounts to the conclusion that β cannot become a trace of movement in addition to H, since such application of IM would render the semantic features of β in $\{H, \beta\}$ invisible for MHD. Thus, it can never be the case that both H and β move. Therefore, *$\{t, t\}$ is derived.

I would like to maintain that the data from the freezing effect lends important support to *$\{t, t\}$. First of all, recall that one of the direct consequences of the H-α schema is that all moved ph(r)ases exhibit the freezing effect (Culicover and Wexler 1980, Uriagereka 1999 and many others; see §3.3.1):

(52)   A moved phase constitutes an island for extraction.

This is simply because the H-α schema predicts that only LIs can undergo IM, and thus that all cases of apparent 'XP'-movement are reduced to instances of internal merger of a phase-head LI that has subjected its phase-interior domain to Transfer prior to IM.

(53)
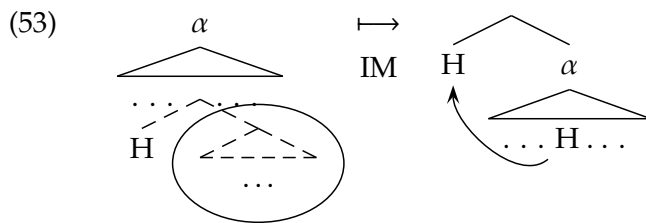
Various data are attributed to the freezing effect (52), some of which are reproduced below.

(54)   a.   *Which candidate$_i$* were there [posters of t_i] all over the town?

         b.   **Which candidate$_i$* were [posters of t_i]$_j$ t_j all over the town?

(55)   a.   *Who$_i$* did you see [pictures of t_i]?

b. *Who$_i$ did you believe [pictures of t_i]$_j$ to be t_j the cause of the riot?

(56)   a.   *Who$_i$ is it obvious [that John likes t_i]?

        b.   *Who$_i$ is [that John likes you]$_i$ t_i obvious?
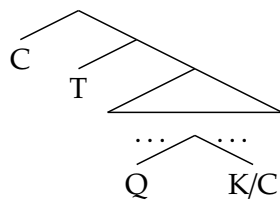
In general, it has been claimed that the freezing effect in these examples arises as a result of EPP-driven A-movement of the subject KP/CP.

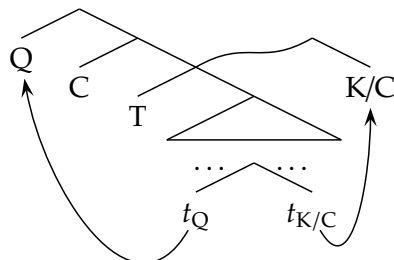   However, the freezing effect alone cannot exclude the following derivation.

(57)   a.   Successive cyclic A′-movement of Q to the edge of K/C at the level of the K/C-phase:
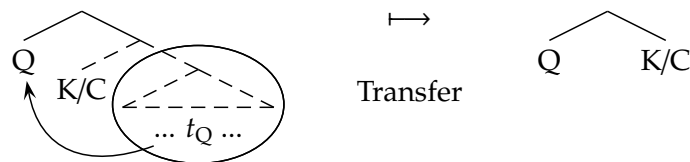
        b.   Construction of the the C-phase by EM:

        c.   A-movement of K/C and A′-movement of Q apply independently at the C-phase level:

As far as we subscribe to the theory of unbounded Merge (or of undeletable EFs), IM applies freely, and nothing prevents the KP-internal Q from escaping to the edge of the K-phase (57a). If this happens, Q is kept accessible for further operation even after the K/C-phase-level Transfer. Q and K/C then can act as dissociated LIs, and at the relevant C-phase-level, they can independently move to their designated positions, as in (57c). The derivation in (57) appears to be perfectly legitimate, as far as the H-α schema is concerned.

It is true to say that the H-α schema predicts freezing effects on any moved subject KP, *as long as Q stays within the interior of K/C*. However, the theory of free Merge itself cannot force anything to stay in situ, given that IM is (as) costless (as EM). Then, unless we find some independent means to restrict the applicability of such 'vacate-phase' movement as the one in (57a), it would be predicted that anything can freely move and escape the effect of the PIC, a worrisome conclusion that virtually nullifies the prediction of freezing effects.

Instead of stipulating *ad hoc* constraints on the applicability of IM, I would like to maintain that the prediction of freezing effects should be kept as such, and derivations like the one in (57) should be ruled out by some independent constraint. I would like to claim that what is at stake in (57) is specifically *{t, t} (48). Note crucially that the SO {t_Q, $t_{K/C}$} in (57c) violates this constraint. Specifically, by virtue of being subjected to A-movement, the trace of K/C becomes unable to head the SO {t_Q, $t_{K/C}$}. {t_Q, $t_{K/C}$} crucially occupies the configuration to which the θ-role is assigned, and thus K's inability to head this SO violates the CI requirement that each K be assigned one and only one θ-role (the Θ-criterion, Chomsky 1981 *et seq.*). This way, the unacceptability of the relevant examples of freezing effects is accounted for by the combination of the freezing effect (52) and *{t, t} (48): on the one hand, if Q stays inside KP/CP, then the PIC-based proposition in (52) accounts for the freezing effect. On the other hand, if Q moves to the edge of the K/C-phase, forming {Q, K/C} after Transfer, further movement of both Q and K/C from such a structure is ruled out

by *{*t*, *t*}.

Let me note that the effect of *{*t*, *t*} arises only when K/C eventually moves by IM. Thus, cases involving an in-situ object K/C are not excluded by *{*t*, *t*}, as, e.g., the following contrast readily suggests:

(58)   a.   *Who$_i$* did you see [pictures of t_i]?

b.   **Who$_i$* did you believe [pictures of t_i]$_j$ to be t_j the cause of the riot?

(59)   a.   *Who$_i$* did everybody believe [that John kissed t_i]?

b.   **Who$_i$* was [that John kissed t_i]$_j$ t_j believed by everybody?

If we schematically describe the derivation for the acceptable examples in (58a)/(59a), it would look like (60):
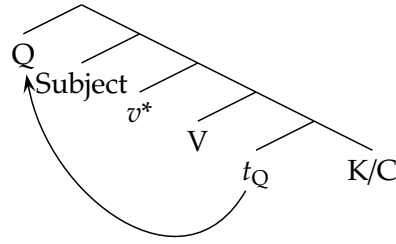
(60)   a.   Successive cyclic A'-movement of Q to the edge of K/C at the level of the K/C-

phase:
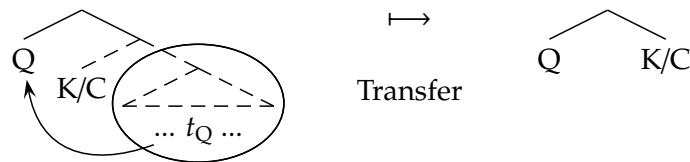
b.   Construction of the next phase with v^* by EM:
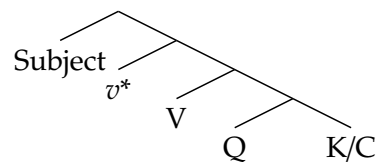
c.   A'-movement of Q at the v^*-phase level:

The lowest constituent in (60c), namely $\{t_Q, K/C\}$, is interpretable, given that K/C stays in situ and hence it can provide its semantic features for endocentric interpretation of $\{t_Q, K/C\}$. V assigns its θ-role to $\{t_Q, K/C\}$, and K/C can receive it via headedness. This way, *$\{t, t\}$ and the freezing effect readily rule in this structure.

Notice crucially that the H-α schema alone cannot exclude any structure of the form $\{t, t\}$: as long as these two traces are copies of simplex LIs, IM of such LIs routinely satisfies the H-α schema. Thus, it may be that the configuration *$\{t, t\}$ is one prominent example that justifies the independent force of FI in the strong sense: such a configuration only violates FI, due to its failure to provide any endocentric interpretation via semantic features. Thus, we may conclude that the empirical generalization *$\{t, t\}$, which is anyway necessary for our account of the freezing effect, can be seen as following only from FI in its strong sense. Crucially, Weak FI will fail to rule out *$\{t, t\}$, since it holds that failures of head-detection are tolerated by CI: it simply neglects such headless structures. Then, speculative though it may look, the empirical generalization *$\{t, t\}$ may be seen as a piece of evidence in favor of the strong version of FI.[13]

---

[13]vP-fronting may constitute a piece of apparent counterevidence to *$\{t, t\}$.

(i)   [$_{vP}$ criticize himself]$_i$, John$_j$ really did [t_j t_i].

However, it appears that vP is not interpreted as a logical operator for such a structure. It is known that fronted vPs somehow obligatorily 'reconstruct', so to speak (Heycock 1995, Takano 1995), probably due to their predicative nature. Presumably, then, the fronted v is just phonologically pronounced in the higher position, signaling some sort of discourse prominence, while semantically interpreted at the lower position, θ-marking t_j I have to leave the study of the mechanism of reconstruction for future research.

5.7   Concluding Remarks

This chapter is dedicated to the following question: why does the effect of endocentricity arise the way it does? This fundamental question has been more or less masked by the dominant labeling-by-projection hypothesis in most of the current literature, but the fact that all kinds of mechanisms and algorithms have been put forward for the description of endocentricity is a suggestive sign for the importance of the *why*-question. The MP provides a ready-made answer to such a question, namely that endocentricity arises since it is the simplest mode of SEM-interpretation. I attempted at making sense of such an answer by claiming that MHD provides the shortest path to detect semantic features of LIs and thereby assign compositional interpretation to each SO. The H-α schema is further rationalized as contributing to constant and failure-proof detection of heads via MHD. The strength of FI is correspondingly corroborated.

Chapter 6

Conclusion

6.1 Reviewing the *Why*-questions

As a final remark, let me go through the set of several *why*-questions that the present thesis was dedicated to. These questions and problems which they raise constitute the major underlying themes of the present study. I do not claim that these questions are new by any means, nor that I have managed to offer fully satisfactory answers to any of them. Nevertheless, the specific combination of these considerations and the visions for possible answers to them constitutes the basis of the conclusions reached in the present thesis.

(1)   *Why is linguistic computation always structure-dependent? (Why are there no structure-independent (e.g., linear-dependent) rules?)*

Throughout the thesis, I tried to articulate the answer envisaged by Berwick and Chomsky (2008): syntactic computation is always structure-dependent since syntax only generates compositional structure via Merge. No notion of linear order or precedence is available in syntax, and hence it admits no linear-dependent computation. The present thesis was devoted to the exploration of the fundamental ramification of the theory of bare phrase

structure, which is the full dissociation of the three kinds of linguistic information that were once compounded into phrase structure rules (PSRs) and/or the X-bar schema: compositional structure, labeling and precedence. It was also claimed that the notions of labeling and projection are entirely dispensable. Compositional structuring by means of recursive Merge, along with certain other minimal search operations for Agree and MHD, is essentially the only mode of structure generation.

(2)   ***Why are applications of insertion and tampering severely restricted in the syntax of natural language?***

The answer suggested in Chapter 2 and explored throughout the present thesis is this: insertion and tampering are disallowed because there are no such operations in syntax as a matter of principle. Syntax assumes Merge as the only generative device, hence it cannot insert any new features or modify any internal constitution of already created SOs in the course of derivation. Strict adherence to the Inclusiveness Condition (IC) and the No-tampering Condition (NTC) is achieved with the help of phase cyclicity. Various departures from the IC and the NTC are shown to be unwarranted and simply dispensable: referential indices and distinctness marking, Numeration and lexical (sub-)arrays, bar-level indices of the X-bar-theory, labels and nonterminals, projection and feature-percolation, are all eliminated in favor of the minimal theory of Merge-based derivation by phase. So long as we stick to this simplest answer to the *why*-question, we really should depart from the PSR-based hypothesis that labeling by projection is part of syntax, as we concluded through the present study.

(3)   ***Why does syntactic computation obey cyclic derivation by phase, and enforce locality constraints in this way?***

Syntactic computation, optimized for CI's efficient exploitation of SEM, naturally benefits from adopting cyclic reduction of computational load and active derivational workspace,

and the most efficient means of such computational cyclicity is achieved by Transfer, which terminates further access to SOs upon subjecting them to Interpret and Spell-Out, as soon as they become convergent. Such elimination upon interpretation by Transfer constitutes the rationale for the effect of what is usually called the Phase Impenetrability Condition (PIC), which derives a number of empirically grounded locality constraints. Moreover, given that the single formulation of Merge immediately gives rise to two modes of compositional structuring, external Merge (EM) and internal Merge (IM), cyclic coordination of Transfer and other copy-formation operations, IM and Agree, is arguably the simplest way to exploit the natural bifurcation of EM and IM for CI purposes, keeping to the strict government by the IC and the NTC. It was also speculated that the existence of uninterpretable features in human language may be rationalized as a natural device to encode phase cyclicity.

(4)   *Why are there word order variations in natural languages? Why is there no (strong) antisymmetry, i.e., no universal word order template?*

The answer is simply that linear order is not part of narrowly syntactic computation. Syntax is too simple and minimal to give rise to one and the same solution to every problem of externalization. Indeed, if precedence is an integral part of syntax, as Kayne claims, it becomes really mysterious why it does not figure in linguistic computation for most of the time, and also why 'optimal' utilization of such precedence-relations does not lead to a single uniform answer to each and every linearization problem.

(5)   *Why is there no XP-XP merger?*

The H-α schema simply follows from the fact that FL assumes no legitimate means to generate and interpret such XP-XP structures. Projection and feature-percolation are prohibited (since there are no such operations to begin with), and specifically there is no way to project edge-features (EFs) of LIs up to phrasal nodes. The lack of labeling/projection also predicts that there are no 'label-dependent' operations in Merge-based minimal syn-

tax. The mechanism of cyclic Transfer naturally works in service of recursive structure embedding governed by the H-α schema. As long as endocentricity reduces to minimal search of LIs (as MHD holds), recursive H-α structuring paves the way for constant and failure-proof satisfaction of endocentric interpretation, which provides a further rationale for the H-α schema.

(6)   ***Why is the acquisition of LIs so uniformly rapid and successful?***

The H-α schema, an independently motivated principle, provides a strong analytical tool for the acquisition of LIs. As suggested by Yang and Gambell's work (see Yang 2002, 2004, Gambell and Yang 2003 among others), the general mechanism of statistical data analysis (presumably a constituent of the third factor of FL design) provides a reasonable first-cut segmentation of words in primary linguistic data, when it works in tandem with the principle (presumably determined by UG) that each phonological word bears a single primary accent. As acquisition proceeds, the H-α schema helps both the 'bottom-up' synthesis of such LIs into clausal syntax, and also the 'top-down' analysis and reanalysis of phonological word structures, leading to constant revisions of the Lexicon of provisionally analyzed LIs, each of which is assigned, among other properties, undeletable edge-features, intrinsic semantic features and also instructions for phonological linearization and morphological constitutions. Various cues from the primary linguistic data (distributional or semantic) may be taken as evidence for such decomposition, while the number of functional categories is also severely restricted by the interplay of the H-α schema and such cues from the external data.

(7)   ***Why does the effect of endocentricity arise so prominently in linguistic computation***
         of SEM and PHON?

The present thesis is devoted to making sense of the following answer: endocentricity arises because it is the simplest possible form of compositional interpretation. No extra-

neous representational tricks such as labeling and projection are available in the minimal theory of bare phrase structure, and thus CI just seeks to determine compositional interpretation of each SO by minimal search. The theory of phase cyclicity and the eradication of projection and feature-percolation pave the way for consistent and efficient satisfaction of this CI-requirement, which amounts to incorporating the derivational constraint of the H-α schema.

6.2 Full Interpretation All the Way Through

What do we have as a result of this minimalist inquiry, in the end? Syntax assumes only the simplest form of Merge (set-formation) for compositional structuring, and it generates, via cyclic derivation by phase, only recursive H-α structures, each of which can be efficiently assigned compositional interpretation by CI (and SM) via minimal search. No extraneous representations or operations are introduced during the course of syntactic derivation, dispensing with referential indices and distinctness marking, Numeration and lexical (sub-)arrays, labels and nonterminals, projection and percolation of features, specifier-of relations, etc., not only because Merge cannot generate them but also, presumably, because they just don't contribute to any interpretation at CI. Full and consistent satisfaction of Full Interpretation (FI) is thereby achieved, maximally corroborating the desideratum of the Strong Minimalist Thesis (SMT).

(8)   *Strong Minimalist Thesis (SMT)* :

FL is an optimal solution to the conditions imposed by CI and SM.

There is absolutely no doubt that countless many problems are left for future research, but I nevertheless hope that the theoretical considerations and empirical results that the present thesis offers may constitute a partial and modest step toward the eventual demonstration of the SMT.

Bibliography

Abels, Klaus. 2001. Successive cyclicity, anti-locality, and adposition-stranding. Doctoral Dissertation, University of Connecticut, Storrs.

Abels, Klaus. 2007. Towards a restrictive theory of (remnant) movement. *Linguistic Variation Yearbook* 7:53–120.

Abels, Klaus. 2009. Some implications of improper movement for cartography. In *Alternatives to cartography*, ed. J. van Craenenbroek, 325–359. Berlin: Mouton de Gruyter.

Abels, Klaus, and Ad Neeleman. 2009. Universal 20 without the LCA. In *Merging features: Computation, interpretation and acquisition*, ed. Josep M. Brucart, Anna Gavarró, and Jaume Solà, 60–80. Oxford: Oxford University Press.

Abney, Steven Paul. 1987. The noun phrases in its sentential aspect. Doctoral Dissertation, MIT.

Ackema, Peter, and Ad Neeleman. 2002. Effects of short-term storage in processing rightward movement. In *Storage and computation in the langauge faculty: Studies in theoretical psycholinguistics*, ed. Sieb Nooteboom, Fred Weerman, and Frank Wijnen, 219–256. Kluwer.

Agbayani, Brian, Chris Golston, and Toru Ishii. to appear. Prosodic scrambling. In *Proceedings of the 5th Formal Approaches to Japanese Linguistics Conference (FAJL5)*. MITWPL.

Asbury, Anna. 2008. *The morphosyntax of case and adpositions*. Utrecht: LOT.

Barwise, Jon, and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4:159–209.

Belletti, Adriana. 2001. 'inversion' as focalization. In *Subject inversion in Romance and the theory of universal grammar*, ed. Aafke Hulk and Jean-Yves Pollock, 60–90. Oxford: Oxford University Press.

Berwick, Robert C., and Noam Chomsky. 2008. 'poverty of the stimulus' revisited: Recent challenges reconsidered. ms. MIT. A short abstract appeared in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Washington, DC, p. 383.

den Besten, Hans, and Gert Webelhuth. 1990. Stranding. In *Scrambling and barriers*, ed. Günther Grewendorfand and Wolfgang Sternefeld, 77–92. New York: Academic Press.

Bever, Thomas G., and David Poeppel. 2010. Analysis by synthesis: A (re-)emerging program of research for language and vision. *Biolinguistics* 4:174–200.

Biberauer, Theresa, Anders Holmberg, and Ian Roberts. 2008. Structure and linearization in disharmonic word orders. In *Proceedings of WCCFL26*, 96–104. Somerville, MA.: Cascadilla Press.

Bittner, Maria, and Ken Hale. 1996a. The structural determination of Case and agreement. *Linguistic Inquiry* 27:1–68.

Bittner, Maria, and Ken Hale. 1996b. Ergativity: Toward a theory of a heterogeneous class. *Linguistic Inquiry* 27:531–604.

Bobaljik, Jonathan David. 1995. Morphosyntax: The syntax of verbal inflection. Doctoral Dissertation, MIT.

Boeckx, Cedric. 2003. *Islands and chains*. Amsterdam: John, Benjamins.

Boeckx, Cedric. 2008a. *Bare syntax*. Oxford: Oxford University Press.

Boeckx, Cedric. 2008b. A note on object EPP. ms. Harvard University.

Boeckx, Cedric. 2008c. *Understanding minimalist syntax*. Oxford: Blackwell.

Boeckx, Cedric. 2009. On the locus of asymmetry in UG. *Catalan Journal of Linguistics* 8:41–53.

Boeckx, Cedric. 2010a. Defeating lexicocentrism. ms. ICREA/UAB. Downloadable at ling-Buzz/001130.

Boeckx, Cedric. 2010b. Elementary syntactic structures: a minimalist inquiry. ms. ICREA/UAB.

Boeckx, Cedric. 2011. Approaching parameters from below. In *The biolinguistic enterprise new perspectives on the evolution and nature of the human language faculty*, ed. Anna Maria Di Sciullo and Cedric Boeckx, 205–221. Cambridge, MA: Oxford University Press.

Boeckx, Cedric. to appear . A tale of two minimalisms: Reflections on the plausibility of crash-proof syntax, and its free-merge alternative. In *Derivational simplicity in minimalist syntax*, ed. Michael Putnam. Amsterdam: John Benjamins.

Boeckx, Cedric, and Kleanthes K. Grohmann. 2007. Putting phases into perspective. *Syntax* 10:204–222.

Boeckx, Cedric, and Juan Uriagereka. 2007. Minimalism. In *The oxford handbook of linguistic interfaces*, ed. Gillian Ramchand and Charles Reiss, 541–573. Oxford: Oxford University Press.

Bošković, Željko. 1997. *The syntax of nonfinite complementation: an economy approach*. Cambridge, MA: MIT Press.

Bošković, Željko. 2002a. A-movement and the EPP. *Syntax* 5:167–218.

Bošković, Željko. 2002b. On multiple wh-fronting. *Linguistic Inquiry* 33:351–383.

Brame, Michael. 1981. The general theory of binding and fusion. *Linguistic Analysis* 7.3:277–325.

Brame, Michael. 1982. The head-selector theory of lexical specifications and the nonexistence of coarse categories. *Linguistic Analysis* 10.4:321–325.

Bresnan, John W. 1970. On complementizers: Toward a syntactic theory of complement types. *Foundations of Language* 6:297–321.

Bresnan, John W. 1972. Theory of complementation in English syntax. Doctoral Dissertation, MIT.

Cable, Seth. 2007. The grammar of Q. Doctoral Dissertation, MIT.

Cable, Seth. 2010. *The grammar of Q: Q-particles, wh-movement, and pied-piping*. Oxford:

Oxford University Press.

Caha, Pavel. 2009. The nanosyntax of case. Doctoral Dissertation, University of Toromsø.

Carnie, Andrew. 2002. *Syntax: A generative introduciton*. Blackwell.

Carnie, Andrew. 2008. *Constituent structure*. Oxford: Oxford University Press.

Carstens, Vicki. 2000. Concord in minimalist theory. *Linguistic Inquiry* 31:319–355.

Carstens, Vicki. 2003. Rethinking complementizer agreement: Agree with a Case checked goal. *Linguistic Inquiry* 34:393–412.

Cattell, Ray. 1976. Constraints on movement rules. *Language* 52:18–50.

Cecchetto, Carlo. 2007. Some preliminary remarks on a 'weak' theory of linearization. *Annali Online di Ferrara Lettere* 1:1–13. http://eprints.unife.it/annali/lettere/2007vol1/indice.htm.

Cecchetto, Carlo. to appear. Backward dependencies must be short: A unified account of the Final-over-Final and the Right Roof Constraints and its consequences for the syntax/morphology interface. In *Challenges to Linearisation*, ed. Theresa Biberauer and Ian Roberts. Berlin: Mouton.

Ceplová, Markéta. 2001. Minimalist islands—restricting P-features. ms. MIT.

Chametzky, Robert A. 2000. *Phrase structure: from GB to minimalism*. Oxford: Blackwell.

Chomsky, Noam. 1955/1975. *The logical structure of linguistic theory*. ms. Harvard University, 1955. Published in part in 1975, New York: Plenum.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton. 2nd edition, (2002).

Chomsky, Noam. 1959. On certain formal properties of grammars. *Information and Control* 2:137–167. Reprinted in R. D. Luce, R. Bush, and E. Galanter, eds. (1965),.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1966. *Cartesian linguistics: A chapter in the history of rationalist thought*. New York: Harper & Row. The 2nd edition with an introduction by James McGilvray is published in 2002 by Christchurch, New Zealand: Cybereditions; the 3rd edition to appear in 2009, supplemented with a new introduction by James McGilvray, to be

published by Cambridge University Press.

Chomsky, Noam. 1966/2002. *Cartesian linguistics: A chapter in the history of rationalist thought*. Christchurch, New Zealand: Cybereditions, 2nd edition edition. Supplemented with an introduction by James McGilvray is published in 2002.

Chomsky, Noam. 1966/2009. *Cartesian linguistics: A chapter in the history of rationalist thought*. Cambridge: Cambridge University Press, 3rd edition edition. Supplemented with a new introduction by James McGilvray.

Chomsky, Noam. 1968. *Language and mind*. New York: Harcourt, Brace, Jovanovitch.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English transformational grammar*, ed. Roderick A. Jacobs and Peter S. Rosenbaum, 184–221. Waltham, MA: Ginn.

Chomsky, Noam. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, ed. Stephen Anderson and Paul Kiparsky, 232–286. New York: Rinehart and Winston.

Chomsky, Noam. 1975. *Reflections on language*. New York: Pantheon.

Chomsky, Noam. 1977. *Essays on form and interpretation*. New York: North-Holland.

Chomsky, Noam. 1980a. On binding. *Linguistic Inquiry* 11:1–46.

Chomsky, Noam. 1980b. On cognitive structures and their development. In *Language and learning, the debate between jean piaget and noam chomsky*, ed. M. Piattelli-Palmarini, 35–52. Cambridge, MA: Harvard University Press. (Talks held at Centre Royaumont, Paris, October 1975.).

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1982. *Some concepts and consequences of the Government and Binding Theory*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986a. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986b. *Knowledge of language*. New York: Praeger.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Ken Hale and Samuel J. Keyser, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 1994. Bare phrase structure. *MIT Occasional Papers in Linguistics* 5:Department of Linguistics and Philosophy, MIT. Reprinted in G. Webelhuth, ed. (1995), Malden: Blackwell, *Government and Binding Theory and the Minimalist Program*, 383-439.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and beyond: The cartography of syntactic structures*, ed. Adriana Belletti, volume 3, 104–131. New York: Oxford University Press.

Chomsky, Noam. 2005. Three factors in the language design. *Linguistic Inquiry* 36:1–22.

Chomsky, Noam. 2007a. Approaching UG from below. In *Interfaces + recursion = language? Chomsky's minimalism and the view from semantics*, ed. U. Sauerland and H.-M. Gärtner, 1 –29. Berlin and New York: Mouton de Gruyter.

Chomsky, Noam. 2007b. Biolinguistic explorations: design, development, evolution. *International Journal of Philosophical Studies* 15(1):1–21.

Chomsky, Noam. 2007c. Of minds and language. *Biolinguistics* 1:9–27.

Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory*, ed. Robert Freidin, Carlos Otero, and Maria Luisa Zubizarreta, 133–166. Cambridge, MA: MIT Press.

Chomsky, Noam. 2010a. Introduction to the collection ed. by Naoki Fukui. ms. MIT.

Chomsky, Noam. 2010b. Poverty of stimulus: Unfinished business. Lecture presented at Johannes-Gutenberg University Mainz, March 24, 2010.

Chomsky, Noam. 2010c. Restricting stipulations: consequences and challenges. Talk given at the University of Stuttgart, March 24.

Chomsky, Noam, and Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8:425–504.

Chomsky, Noam, and Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research*, ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann, volume Vol. 1, 506–569. Walter de Gruyter. Reprinted in N. Chomsky (1995),.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Cinque, Guglielmo. 2002. *Functional structure in DP and IP: The cartography of syntactic structures, vol. 1*. Oxford: Oxford University Press.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36:315–332.

Cinque, Guglielmo, and Luigi Rizzi, ed. 2010. *Mapping spatial PPs: The cartography of syntactic structures, Vol.6*. New York: Oxford University Press.

Collins, Chris. 2002. Eliminating labels. In *Derivation and explanation in the minimalist program*, ed. Samuel David Epstein and T. Daniel Seely, 42–64. Oxford: Blackwell.

Corver, Norbert. 2006. Freezing effects. In *The Blackwell companion to syntax II*, ed. M. Everaert and H. van Riemsdijk, 383–406. Oxford: Blackwell.

Crain, Stephen, and Mineharu Nakayama. 1987. Structure dependence in grammar formation. *Language* 63:522–543.

Culicover, Peter W. 1982. *Syntax*. New York/London: Academic Press.

Culicover, Peter W., and Kenneth Wexler. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Donati, Caterina. 2006. On wh-head-movement. In *Wh-movement: Moving on*, ed. Lisa L.-S. Cheng and N. Corver, 21–46. Cambridge, MA: MIT Press.

Embick, David, and Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39:1–53.

Emonds, Joseph. 2009. Valuing V features and N features: What adjuncts tell us about case, agreement, and syntax in general. In *Merging features: Computation, interpretation and acquisition*, ed. Josep M. Brucart, Anna Gavarró, and Jaume Solà, 194–214. Oxford: Oxford University Press.

Emonds, Joseph E. 1970. Root and structure-preserving transformations. Doctoral Dissertation, MIT.

Emonds, Joseph E. 1976. *A transformational approach to English syntax: Root, structure-preserving, and local transformations*. New York, NY: Academic Press.

Endo, Yoshio. 1996. Right dislocation. In *MIT working papers in linguistics 29: Formal approaches to Japanese linguistics 2*, 1–20. MITWPL.

Epstein, Samuel David. 1999. Un-principled syntax: The derivation of syntactic relations. In *Working minimalism*, ed. Samuel D. Epstein and Norbert Hornstein, 317–345. Cambridge, MA: MIT Press.

Epstein, Samuel David. 2009. The unification of theta relations: How TRANSFER renders SpecvP a theta marked complement. ms. University of Michigan.

Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima, and Hisatsugu Kitahra. 1998. *A derivational approach to syntactic relations*. Oxford: Oxford University Press.

Epstein, Samuel David, Hisatsugu Kitahara, and T. Daniel Seely. 2009. Structure building that can't be. ms. University of Michigan, Keio University and Eastern Michigan University. Presented at the Ways of Structure Building conference, Vitoria-Gasteiz, November 13-14, 2008.

Epstein, Samuel David, and T. Daniel Seely. 2002. Rule applications as cycles in a level-free syntax. In *Derivation and explanation in the minimalist program*, ed. Samuel D. Epstein and T. Daniel Seely, 65–89. Oxford: Blackwell.

Epstein, Samuel David, and T. Daniel Seely. 2005. *Transformations and derivations*. Cambridge: Cambridge University Press.

Fiengo, Robert, and Robert May. 1994. *Indices and identity*. Cambridge, MA: MIT Press.

Fox, Danny, and David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31:1–45.

Frampton, John, and Sam Gutmann. 2000. Agreement is feature sharing. ms. Northeastern University.

Freidin, Robert, and Howard Lasnik. 1981. Disjoint refernece and wh-trace. *Linguistic Inquiry* 12:39–53.

Fujita, Koji. 2007. Facing the logical problem of language evolution. *English Linguistics* 24(1):78–108.

Fujita, Koji. 2009. A prospect for evolutionary adequacy: Merge and the evolution and development of human language. *Biolinguistics* 3:128–153.

Fukui, Naoki. 1986/1995. A theory of category projection and its applications. Doctoral Dissertation, MIT. Published in 1995 with revisions as *Theory of projection in syntax*, Kurosio Publishers and CSLI publications.

Fukui, Naoki. 1988. Deriving the differences between English and Japanese: A case study in parametric syntax. *English Linguistics* 5:249–270.

Fukui, Naoki. 1993. Parameters and optionality. *Linguistic Inquiry* 24:399–420. Reprinted in Fukui (2006).

Fukui, Naoki. 1995. The principles-and-parameters approach: A comparative syntax of English and Japanese. In *Approaches to language typology*, ed. Masayoshi Shibatani and T. Bynon, 327–371. Oxford: Oxford University Press. Reprinted in Fukui (2006).

Fukui, Naoki. 1996. On the nature of economy in language. *Cognitive Studies* Bulletin of the Japanese Cognitive Science Society, 3:51–71. Reprinted in Fukui (2006).

Fukui, Naoki. 1999. An A-over-A perspective on locality. In *Linguistics: In search of the human mind - a festschrift for kazuko inoue*, ed. M. Muraki and E. Iwamoto, 109–127. Tokyo: Kaitakusha. Reprinted in Fukui (2006).

Fukui, Naoki. 2001. Phrase structure. In *The handbook of contemporary syntactic theory*, ed. Mark Baltin and Chris Collins, 374–406. Oxford: Blackwell. Reprinted in Fukui (2006).

Fukui, Naoki. 2006. *Theoretical comparative syntax: Studies in macroparameters*. Dordrecht: Routledge.

Fukui, Naoki. 2008. Gengo-no kihonenzan-o meguru oboegaki [A note concerning basic operations in language]. In *Gengokenkyuu-no genzai: keishiki-to imi-no intaafeisu [The state of the art in linguistic research: the syntax-semantics interface]*, ed. Yoshiaki Kaneko, Akira Kikuchi, and Daiko Takahashi, 1–21. Tokyo: Kaitakusha.

Fukui, Naoki. forthcoming. Merge and bare phrase structure. In *The Oxford handbook of linguistic minimalism*, ed. Cedric Boeckx. Oxford: Oxford University Press.

Fukui, Naoki, and Hironobu Kasai. 2004. Spelling-Out scrambling. *Linguistic Variation Yearbook* 4.1:109–141.

Fukui, Naoki, and Hiromu Sakai. 2003. The Visibility Guideline for Functional Categories: Verb raising in Japanese and related issues. *Lingua* 113:321–375. Reprinted in Fukui (2006).

Fukui, Naoki, and Margaret Speas. 1986. Specifiers and projection. *MIT Working Papers in Linguistics* 8:128–172. Reprinted in Fukui (2006).

Fukui, Naoki, and Yuji Takano. 1998. Symmetry in syntax: Merge and Demerge. *Journal of East Asian Linguistics* 7:27–86. Reprinted in Fukui (2006).

Fukui, Naoki, and Yuji Takano. 2000. Nominal structure: an extension of the Symmetry Principle. In *The derivation of VO and OV*, ed. Peter Svenonius, 219–254. Amsterdam: John Benjamins. Reprinted in Fukui (2006).

Gallego, Ángel J. 2007. Phase theory and parametric variation. Doctoral Dissertation, Universitat Autònoma de Barcelona.

Gallego, Ángel J. 2009. 3 questions about phase theory. Paper presented at Harvard University, February 2009.

Gallego, Ángel J. 2010a. A framework of phases for linguistic theory. Under review for Ángel J. Gallego, ed., *Phases: developing the framework*.

Gallego, Ángel J. 2010b. *Phase theory*. Amsterdam: John Benjamins.

Gambell, Timothy, and Charles D. Yang. 2003. Scope and limits of statistical learning in word segmentation. In *Proceedings of NELS 34*, 29–30.

Goto, Nobu. 2010. Some consequences of feature inheritance. ms. Tohoku Gakuin University.

Grimshaw, Jane B. 1991/2005. Extended projection. In *Words and structure*, 1–73 (final version of the manuscript written in 1991). CSLI Publications.

Groat, Erich. 1999. Raising the case of expletives. In *Working minimalism*, ed. Samuel D. Epstein and Norbert Hornstein, 27–44. Cambridge, MA: MIT Press.

Groat, Erich, and John O'Neil. 1996. Spell-Out at the LF interface: Achieving a unified syntactic computational system in the minimalist framework. In *Minimal ideas: Syntactic studies in the minimalist framework*, ed. Werner Abraham, Samuel D. Epstein, Höskuldur Thráinsson, and Jan-Wouter Zwart, 113–139. Amsterdam: John Benjamins.

Guimarães, Maximiliano. 2000. In defense of vacuous projections in bare phrase structure. *University Of Maryland Working Papes In Linguistics* 9:90–115.

Hagstrom, Paul. 1998. Decomposing questions. Doctoral Dissertation, MIT.

Hale, Ken, and Samuel J. Keyser. 1993. On argument structure and the lexical expression of grammatical relations. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Ken Hale and Samuel J. Keyser, 53–110. Cambridge, MA: MIT Press.

Hale, Ken, and Samuel J. Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Ken Hale and Samuel J. Keyser, 111–176. Cambridge, MA: MIT Press.

Halle, Morris, and Alec Marantz. 1994. Some key features of Distributed Morphology. *MIT Working Papers in Linguistics* 21:275–288.

Halle, Morris, and Kenneth Stevens. 1959. Analysis by synthesis. In *Proceedings of Seminar on Speech Compression and Processing, Vol. 2*, ed. W. Wathen-Dunn and L. E. Woods, paper

D7.

Halle, Morris, and Kenneth Stevens. 1963. Speech recognition: A model and a program for research. *IRE Transactions on Information Theory* 8:155–159. [Reprinted in Morris Halle. 2002. *From memory to speech and back*. Berlin: Mouton de Gruyter.].

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.

Heycock, Caroline. 1995. Asymmetries in reconstruction. *Linguistic Inquiry* 26:547–570.

Hinzen, Wolfram. 2006. *Mind design and minimal syntax*. Oxford: Oxford University Press.

Hinzen, Wolfram. 2007. *An essay on names and truth*. Oxford: Oxford University Press.

Hiraiwa, Ken. 2005. Dimensions of symmetry in syntax: Agreement and clausal architecture. Doctoral Dissertation, MIT.

Hiraiwa, Ken, and Shinichiro Ishihara. 2002. Missing links: Cleft, sluicing and 'no da' construction in Japanese. In *Proceedings of HUMIT 2001. MITWPL #43*, 35–54. Cambridge, MA: MITWPL.

Hiraiwa, Ken, and Shinichiro Ishihara. to appear. Syntactic metamorphosis. *Syntax* .

Hoekstra, Eric. 1991. Licensing conditions on phrase structure. Doctoral Dissertation, University of Groningen.

Hoekstra, Jarich, and Laszlo Marácz. 1989. On the position of inflection in west-germanic. *Working Papers in Scandinavian Syntax* 44:75–88.

Hoji, Hajime. 1990. Theories of anaphora and aspects of Japanese syntax. ms. University of Southern California.

Hoji, Hajime. 2003. Falsifiability and repeatability in generative grammar: A case study of anaphora and scope dependency in Japanese. *Lingua* 113:377–446.

Holmberg, Anders. 2000. Deriving OV order in Finnish. In *The derivation of VO and OV*, ed. Peter Svenonius, 123–152. Amsterdam: John Benjamins.

Hornstein, Norbert. 2009. *A theory of syntax: Minimal operations and universal grammar*. Cambridge: Cambridge University Press.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. MIT.

Hurford, James. 2007. *The origins of meaning: Language in the light of evolution*. Oxford: Oxford University Press.

Iatridou, Sabine. 1993. On nominative case assignment and a few related things. *MIT Working Papers in Linguistics* 19:175–196.

Irurtzun, Aritz. 2006. Focus and clause structuration in the minimalist program. In *Minimalist essays*, ed. Cedric Boeckx, 68–96. Amsterdam: John Benjamins.

Irurtzun, Aritz. 2007. The grammar of focus at the interfaces. Doctoral Dissertation, the University of the Basque Country.

Ishii, Toru. 1997. An asymmetry in the composition of phrase structure and its consequences. Doctoral Dissertation, University of California, Irvine.

Jackendoff, Ray. 1977. *X'-syntax*. Cambridge, MA: MIT Press.

Jayaseelan, Karattuparambil. 2008. Bare phrase structure and specifier-less syntax. *Biolinguistics* 2:87–106.

Kato, Takaomi. 2004. Not so overt movement. In *Proceedings of WCCFL 23*, 436–449.

Kato, Takaomi. 2006. Symmetries in coordination. Doctoral Dissertation, Harvard University.

Kato, Takaomi. 2007. On the nature of the left branch condition: Syntactic or phonological? In *Proceedings of the 9th Seoul International Conference on Generative Grammar*.

Kawasaki, Noriko. 1989. Jibun-tachi and non-coreferential anaphora. In *Papers in quantification*, NFS Grant BNS 8719999, Principal investigators: Emmon Bach, Angelika Kratzer and Barbara Partee, Department of Linguistics, University of Massachusetts, Amherst.

Kayne, Richard S. 1981. ECP extensions. *Linguistic Inquiry* 12:93–133.

Kayne, Richard S. 1983. Connectedness. *Linguistic Inquiry* 14:223–249.

Kayne, Richard S. 1984. *Connectedness and binary branching*. Dordrecht: Foris.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kayne, Richard S. 2004a. Antisymmetry and Japanese. In *Variation and universals in biolin-*

guistics, ed. L. Jenkins, 3–35. Elsevier B. V.

Kayne, Richard S. 2004b. Prepositions as probes. In *Structures and beyond: The cartography of syntactic structures*, ed. Adriana Belletti, volume 3, 192–212. Reprinted in Richard S. Kayne, *Movement and silence*, 2005, pp. 85–105. New York: Oxford University Press.

Kayne, Richard S. 2009. Antisymmetry and the lexicon. *Linguistic Variation Yearbook* 8:1–31. Reprinted in Anna Maria Di Sciullo and Cedric Boeckx (eds.), *The biolinguistic entreprise: New perspectives on the evolution and nature of the human language faculty*. Oxford: Oxford University Press.

Kayne, Richard S. 2010. Why are there no directionality parameters? ms. NYU.

Kishimoto, Hideki. 2005. Wh-in-situ and movement in Sinhala questions. *Natural Language and Linguistic Theory* 23:1–51.

Kitahara, Hisatsugu. 1997. *Elementary operations and optimal derivations*. Cambridge, MA: MIT Press.

Koopman, Hilda, and Dominique Sportiche. 1983. Variables and the bijection principle. *The Linguistic Review* 2:139–160.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase structure and the lexicon*, ed. Johan Rooryck and Laurie Zaring, 109–137. Dordrecht: Kluwer.

Kratzer, Angelika. 2000. The event argument and the semantics of verbs. ms. University of Massachusetts, Amherst.

Kuno, Masakazu. 2003. Scrambling in Japanese as pure Merge. *Linguistic Research: Working Papers in English Linguistics* Tokyo University English Linguistics Association:45–78.

Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.

Kuno, Susumu. 1978. *Danwa-no bunpo*. Tokyo: Taishukan Shoten.

Kurafuji, Takeo. 1999. Japanese pronouns in dynamic semantics: The null/overt contrast. Doctoral Dissertation, Rutgers University.

Kuroda, S.-Y. 1965. Generative grammatical studies in the Japanese language. Doctoral Dissertation, MIT.

Kuroda, S.-Y. 1971. Remarks on the notion of subject. *Journal of Japanese Linguistics* 11.

Kuroda, S.-Y. 1972. Categorical and thetic judgments: Evidence from Japanese syntax. *Foundations of Language* 9:1–37.

Kuroda, S.-Y. 1976. Subject. In *Japanese generative grammar*, ed. Masayoshi Shibatani, 1–16. New York: Academic Press.

Kuroda, S.-Y. 1988. Whether we agree or not: A comparative syntax of English and Japanese. In *Papers from the second international workshop on japanese syntax*, ed. W. J. Poser, 103–143. Stanford, Calif.: CSLI Publications. Reprinted in S.-Y. Kuroda (1992).

Kuroda, S.-Y. 1992. *Japanese syntax and semantics: collected papers*. Dordrecht: Kluwer.

Kuroda, S.-Y. 1999. Bunpo riron to tetsugakuteki shizenshugi [grammatical theory and philosophical naturalism]. In *Gengo to shiko [language and thought]*, ed. Masayuki Oishi, ed., trans. Noam Chomsky, 95–134. Tokyo: Shohakusha.

Kuroda, S.-Y. 2005. Focusing on the matter of topic: A study of *wa* and *ga* in Japanese. *Journal of East Asian Linguistics* 14:1–58.

Lasnik, Howard. 1990. *Essays on restrictiveness and learnability*. Dordrecht: Kluwer.

Lasnik, Howard, Marcela Depiante, and Arthur Stepanov. 2000. *Syntactic Structures revisited: Contemporary lectures on classic transformational theory*. Cambridge, MA: MIT Press.

Lasnik, Howard, and M.-K. Park. 2003. The EPP and the Subject Condition under Sluicing. *Linguistic Inquiry* 34:649–660.

Lasnik, Howard, and Mamoru Saito. 1991. On the subject of infinitives. In *Papers from the 27th regional meeting of Chicago Linguistic Society 1991. Part 1: The General Session*, ed. L. K. Dobrin, L. Nichols, and R. M. Rodriguez, 324–343. Chicago, IL: Chicago Linguistics Society. Reprinted in H. Lasnik (1999), Minimalist Analysis, Blackwell.

Lasnik, Howard, and Mamoru Saito. 1992. *Move α: Conditions on its applications and outputs*. Cambridge, MA: MIT Press.

Lasnik, Howard, Juan Uriagereka, and Cedric Boeckx. 2005. *A course in minimalist syntax: Foundations and prospects*. Oxford: Blackwell.

Legate, Julie Anne. 2003. Some interface properties of the phase. *Linguistic Inquiry* 34:506–516.

Legate, Julie Anne. 2005. Phases and cyclic agreement. In *Perspectives on phases: MITWPL 49*, ed. Martha McGinnis and Norvin Richards, 147–156. Cambridge, MA: MITWPL.

Lohndal, Terje. in progress. Without specifiers: Phrase structure and events. Doctoral Dissertation, University of Maryland, College Park.

Longobardi, Giuseppe. 2003. Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook* 3:101–138.

Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Marantz, Alec. 1984. *On the nature of grammatical relations*. Cambridge, MA: MIT Press.

Marantz, Alec. 1996. 'cat' as a phrasal idiom: Consequences of late insertion in Distributed Morphology. ms., MIT.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In *Proceedings of the 21st annual penn linguistics colloquium*, ed. et al. A. Dimitriadis, L. Siegel, 201–225. University of Pennsylvania Working Papersin Linguistics, Vol. 4.2.

Marantz, Alec. 2007. Phases and words. In *Phases in the theory of grammar*, ed. S.-H. Choe, 191–222. Seoul: Dong In.

Martin, Roger. 2001. Null Case and the distribution of PRO. *Linguistic Inquiry* 32:141–166.

Marušič, Franc. 2009. If non-simultaneous spell-out exists, this is what it can explain. In *Merging features*, ed. J. Brucart, A. Gavarro, and J. Sola, 175–193. Oxford: Oxford University Press.

Matushansky, Ora. 2003. Dps and phase theory. Handout. Presented at UiL OTS, January 30, 2003.

May, Robert. 1985. *Logical form: its structure and derivation*. Cambridge, MA: MIT.

McCawley, James. 1968. Concerning the base component of a transformational grammar.

Foundations of Language 4:243–269.

McGilvray, James. 1998. Meanings are syntactically individuated and found in the head. *Mind and Language* 13:225–280.

McGilvray, James. 2002. MOPs: the science of concepts. In *Belief and meaning: Essays at the interface*, ed. Wolfram Hinzen and Hans Rott, 73–103. Frankfurt: Ontos.

McGilvray, James. 2009. Introduction to the third edition. Noam Chomsky (2009) *Cartesian linguistics: A chapter in the history of rationalist thought, Third Edition*. Cambridge: Cambridge University Press, pp.1-52.

Miyagawa, Shigeru. 1997. Against optional scrambling. *Linguistic Inquiry* 28:1–25.

Miyagawa, Shigeru. 2003. A-movement scrambling and options without optionality. In *Word order and scrambling*, ed. S. Karimi. Oxford: Blackwell.

Miyagawa, Shigeru. 2010. *Why agree? why move?: Unifying agreement-based and discourse configurational languages*. Cambridge, MA: MIT Press.

Mori, Nobue. 2005. A syntactic structure of lexical verbs. Doctoral Dissertation, University of Maryland, College Park.

Moro, Andrea. 2000. *Dynamic antisymmetry*. Cambridge, MA: MIT Press.

Müller, Gereon. 1996. A constraint on remnant movement. *Natural Language and Linguistic Theory* 14:355–407.

Narita, Hiroki. 2007a. Project Both in Japanese, with a case study of head-internal relative clauses. In *Proceedings of the 8th Tokyo Conference on Psycholinguistics*, ed. Yukio Otsu, 213–237. Tokyo: Hituzi Syobo.

Narita, Hiroki. 2007b. A "Project Both" perspective on covert movement in Japanese head-internal relative clauses. MA thesis. Sophia University.

Narita, Hiroki. 2008. Two ways of pronouncing "unpronounceable" QR. In *Proceedings of TCP2008*, ed. Yukio Otsu, 151–175. Tokyo: Hituzi Syobo.

Narita, Hiroki. 2009a. Full interpretation of optimal labeling. *Biolinguistics* 3:213–254.

Narita, Hiroki. 2009b. How syntax naturalizes semantics: A review of Uriagereka (2008).

Lingua 119:1767–1775.

Narita, Hiroki. 2010. The tension between explanatory and biological adequacy: Review of Fukui (2006). *Lingua* 120:1313–1323.

Narita, Hiroki. forthcoming. Phase cycles in service of projection-free syntax. In *Phases: developing the framework*, ed. Ángel J. Gallego. Mouton de Gruyter.

Narita, Hiroki, and Bridget Samuels. 2009. The H-α schema and phonological ph(r)asing. Paper presented at the Formal Approaches to the Phonology-Morphology-Syntax Interfaces workshop at ConSOLE XVIII, Universitat Autònoma de Barcelona (UAB), December 16, 2009.

Neeleman, Ad, and Fred Weerman. 1999. *Flexible syntax: A theory of case and arguments*. Dordrecht: Kluwer.

Newmeyer, Frederick J. 2005. *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press.

Newmeyer, Frederick J. 2006. Newmeyer's rejoinder to Roberts and Holmberg on parameters. ms. University of Washington. lingBuzz/000248.

Newmeyer, Frederick J. 2010. Three approaches to exceptionality in syntactic typology. In *Expecting the unexpected: Exceptions in grammar*, ed. Horst Simon and Heike Wiese. Berlin: Mouton de Gruyter.

Nishigauchi, Taisuke. 1990a. Quantification in syntax. Doctoral Dissertation, University of Massachusetts, Amherst.

Nishigauchi, Taisuke. 1990b. *Quantification in the theory of grammar*. Dordrecht: Kluwer.

Nishigauchi, Taisuke. 1991. Construing wh. In *Logical structure and linguistic structure: Cross-linguistic perspectives*, ed. C.-T. James Huang and Robert May, 197–231. Dordrecht: Kluwer.

Nissenbaum, Jon. 2000. Investigations of covert phrase movement. Doctoral Dissertation, MIT.

Nunes, Jairo. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT

Press.

Nunes, Jairo, and Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3:20–43.

Ortega-Santos, Iván. 2008. Projecting subjects in Spanish and English. Doctoral Dissertation, University of Maryland.

Ott, Dennis. 2009. The conceptual necessity of phases: Some remarks on the minimalist enterprise. In *Explorations of phase theory: Interpretation at the interfaces*, ed. K. K. Grohmann. Mouton de Gruyter.

Ott, Dennis. 2010. Grammaticality, interfaces, and UG. In *Exploring crash-proof grammars*, ed. Michael Putnam, 89–104. Amsterdam: John Benjamins.

Pesetsky, David. 1982. Paths and categories. Doctoral Dissertation, MIT.

Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 355–426. Cambridge, MA: MIT Press.

Pesetsky, David, and Esther Torrego. 2004. Tense, Case, and the nature of syntactic categories. In *The syntax of time*, ed. J. Guéron and J. Lecarme, 495–537. Cambridge, MA: MIT Press.

Pesetsky, David, and Esther Torrego. 2006. Probes, goals, and the nature of syntactic categories. In *Proceedings of the 7th Tokyo Conference on Psycholinguistics*, ed. Yukio Otsu, 25–60. Tokyo: Hituzi Syobo.

Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and clausal architecture: Syntactic derivation and interpretation. in honor of Joseph E. Emonds*, ed. Simin Karimi, Vida Samiian, and Wendy Wilkins, 262–294. Amsterdam: John Benjamins.

Peters, Stanley P., and Robert W. Ritchie. 1973. On the generative power of transformational grammars. *Science* 6:49–83.

Petitto, Laura-Ann. 2005. How the brain begets language. In *The cambridge companion to chomsky*, ed. James McGilvray, 84–101. Cambridge: Cambridge University Press.

Piattelli-Palmarini, Massimo, Juan Uriagereka, and Pello Salaburu, ed. 2009. *Of minds and language: A dialogue with Noam Chomsky in the Basque Country*. Oxford: Oxford University Press.

Pietroski, Paul M. 2005. *Events and semantic architecture*. Oxford: Oxford University Press.

Pietroski, Paul M. 2008. Minimalist meaning, internalist interpretation. *Biolinguistics* 2:317–341.

Poeppel, David, William J. Idsardi, and Virginie van Wassenhove. 2008. Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B* 363:P1071–1086.

Postal, Paul M. 1974. *On raising: one rule of English grammar and its theoretical implications*. Cambridge, MA: MIT Press.

Raposo, Eduardo, and Juan Uriagereka. 1995. Two types of small clauses: toward a syntax of theme/rheme relations. In *Small clauses*, ed. A. Cardinaletti and M. T. Guasti, 179–206. New York: Academic Press. Reprinted in Uriagereka (2002).

Reinhart, Tanya. 1976. Syntacic domain of anaphora. Doctoral Dissertation, MIT.

Reinhart, Tanya. 1981. Definite NP anaphora and c-command domains. *Linguistic Inquiry* 12:605–635.

Reinhart, Tanya. 1983. *Anaphora and semantic interpretation*. London: Croom Helm.

Richards, Marc. forthcoming. On feature inheritance, defective phases, and the movement–morphology connection. In *Phases: developing the framework*, ed. Ángel J. Gallego. Mouton de Gruyter.

Richards, Marc D. 2004. Object shift, scrambling, and symmetrical syntax. Doctoral Dissertation, University of Cambridge.

Richards, Marc D. 2007a. Dynamic linearization and the shape of phases. *Linguistic Analysis* 33:209–237.

Richards, Marc D. 2007b. On feature inheritance: An argument from the phase impenetrability condition. *Linguistic Inquiry* 38:563–572.

Richards, Norvin. 1997. What moves where when in which language? Doctoral Dissertation, MIT.

Richards, Norvin. 1999. Featural cyclicity and the ordering of multiple specifiers. In *Working minimalism*, ed. Samuel David Epstein and Nobert Hornstein, 127–158. Cambridge, MA: MIT Press.

Richards, Norvin. 2001. *Movement in language: interactions and architectures*. Oxford: Oxford University Press.

van Riemsdijk, Henk. 1978. *A case study in syntactic markedness: The binding nature of prepositional phrases*. Dordrecht: Foris.

van Riemsdijk, Henk, and E. Williams. 1986. *Introduction to the theory of grammar*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1978. Violations of the *wh*-island constraint in Italian and the subjacency condition. In *Montreal Working Papers in Linguistics*, volume 11, 155–190. Reprinted in Luigi Rizzi, ed., 1982, Issues in Italian Syntax, 49-76, Dordrecht: Foris.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar: Handbook of generative syntax*, ed. L. Elaegeman, 281–337. Dordrecht: Kluwer.

Rizzi, Luigi. 2004. *The structure of CP and IP: The cartography of syntactic structures, vol.2*. Oxford: Oxford University Press.

Rosenbaum, Peter S. 1970. A principle governing deletion in English sentential complementation. In *Readings in English transformational grammar*, ed. Roderick A. Jacobs and Peter S. Rosenbaum, 220–9. Waltham, MA: Ginn.

Ross, J. R. 1967. Constraints on variables in syntax. Doctoral Dissertation, MIT.

Ross, John Robert. 1974. Three batons for cognitive psychology. In *Cognition and the symbolic processes*, ed. Walter B. Weimer and David Palermo. Hillsdale, NJ: Lawrence Erlbaum.

Ross, John Robert. 1986. *Infinite syntax! (language and being)*. New Jersey: Ablex Publishing Incorporation.

Saito, Mamoru. 1985. Some asymmetries in Japanese and their theoretical implications. Doctoral Dissertation, MIT.

Saito, Mamoru. 1989. Scrambling as semantically vacuous A'-movement. In *Alternative conceptions of phrase structure*, ed. Mark R. Baltin and Anthony S. Kroch, 182–200. Chicago: University of Chicago Press.

Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1:69–118.

Saito, Mamoru. 2003. A derivational approach to the interpretation of scrambling chains. *Lingua* 113:481–558.

Saito, Mamoru. 2005. On Miyagawa's analysis of A-scrambling. Paper presented at Keio University Linguistics Colloquium, November 25th, 2005.

Saito, Mamoru, and Naoki Fukui. 1998. Order in phrase structure and movement. *Linguistic Inquiry* 29:439–474. Reprinted in Fukui (2006).

Sakai, Hiromu. 2000. Predicate ellipsis and nominalization. In *Proceedings of the 2000 Seoul International Conference on Language and Computation*, 85–113. Seoul, Korea: The Linguistic Society of Korea.

Samuels, Bridget. in press. *Phonological architecture: a biolinguistic perspective*. Oxford: Oxford University Press.

Sauerland, Uli. 1998. The meaning of chains. Doctoral Dissertation, MIT.

Seely, T. Daniel. 2006. Merge, derivational c-command, and subcategorization in a label-free syntax. In *Minimalist essays*, ed. Cedric Boeckx, 182–217. Amsterdam: John Benjamins.

Selkirk, Elisabeth O. 1995. Sentence prosody: Intonation, stress, and phrasing. In *The handbook of phonological theory*, ed. John A. Goldsmith, 550–569. London: Blackwell.

Sheehan, Michelle. 2009. The final-over-final constraint as an effect of complement stranding. *Newcastle Working Papers in Linguistics* 15:104–125.

Sheehan, Michelle. 2010. Labelling, multiple spell-out and the final-over-final constraint. In *Incontro di Grammatica Generativa XXXV 2010*, ed. Vincenzo Moscati and Emilio Servidio.

Siena.

Sheehan, Michelle. in press. The resuscitation of CED. In *NELS 40 Proceedings*.

Simon, Mutsuko Endo. 1989. An analysis of the postposing construction in Japanese. Doctoral Dissertation, University of Michigan.

Sorida, Masanobu. 2011. Cyclic Transfer, chains, and the A/A' problem of Japanese scrambling. Paper presented at the Chains in Minimalism conference, Yokohama National University, February 12-13, 2011.

Sorida, Masanobu. in progress. Inclusiveness: Its implications for chain formation. Doctoral Dissertation, Sophia University, Tokyo.

Speas, Margaret. 1986. Adjunctions and projections in syntax. Doctoral Dissertation, MIT.

Sportiche, Dominique. 1988. A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19:425–449.

Starke, Michal. 2004. On the inexistence of specifiers and the nature of heads. In *Structures and beyond: The cartography of syntactic structures*, ed. Adriana Belletti, volume 3, 252–268. New York: Oxford University Press.

Stepanov, Arthur. 2007. The end of CED? minimalism and extraction domains. *Syntax* 10(1):80–126.

Stowell, Tim. 1981. Origins of phrase structure. Doctoral Dissertation, MIT.

Svenonius, Peter. 1994. C-selection as feature-checking. *Studia Linguistica* 58:133–155.

Svenonius, Peter. 2003. Limits on P: Filling in holes vs. falling in holes. *Nordlyd. Proceedings of the 19th Scandinavian Conference of Linguistics* 31:431–445.

Svenonius, Peter. 2004. On the edge. In *Peripheries: Syntactic edges and their effects*, ed. David Adger, Cécile de Cat, and George Tsulas, 261–287. Dordrecht: Kluwer Academic Publishers.

Svenonius, Peter. 2010. Spatial P in English. In *Mapping spatial PPs: The cartography of syntactic structures, Vol.6*, ed. Guglielmo Cinque and Luigi Rizzi, 127–161. New York: Oxford University Press.

Takahashi, Daiko. 1994. Minimality of movement. Doctoral Dissertation, University of Connecticut, Storrs.

Takano, Yuji. 1995. Predicate fronting and internal subjects. *Linguistic Inquiry* 26:327–340.

Takano, Yuji. 1996. Movement and parametric variation in syntax. Doctoral Dissertation, University of California, Irvine.

Takano, Yuji. 2000. Illicit remnant movement. *Linguistic Inquiry* 31:141–156.

Tanaka, Hidekazu. 2001. Right-dislocation as scrambling. *Journal of Linguistics* 37:551–579.

Tenny, Carol Lee. 1987. Grammaticalizing aspect and affectedness. Doctoral Dissertation, MIT.

Tenny, Carol Lee. 1994. *Aspectual roles and syntax-semantics interface*. Boston: Kluwer.

Thráinsson, Höskuldur. 1996. On the (non-)universality of functional categories. In *Minimal ideas: Syntactic studies in the minimalist framework*, ed. Werner Abraham, Samuel D. Epstein, Höskuldur Thráinsson, and Jan-Wouter Zwart, 253–281. Amsterdam: John Benjamins.

Toyoshima, Takashi. 1997. Derivational CED: A consequence of the bottom-up parallel-process of Merge and Attract. In *WCCFL 17 Proceedings*, 505–519.

Truswell, Robert. 2007a. Extraction from adjuncts and the structure of events. *Lingua* 117:1355–1377.

Truswell, Robert. 2007b. Locality of wh-movement and the individuation of events. Doctoral Dissertation, University College London.

Ueyama, Ayumi. 1998. Two types of dependency. Doctoral Dissertation, University of South California.

Ura, Hiroyuki. 1996. Multiple feature checking and grammatical function splitting. Doctoral Dissertation, MIT.

Uriagereka, Juan. 1988. On government. Doctoral Dissertation, University of Connecticut, Storrs.

Uriagereka, Juan. 1997. Formal and substantive elegance in the minimalist program (on the

emergence of some linguistic forms). In *The role of economy principles in linguistic theory*, ed. C. Wilder, H.-M. Gartner, and M. Bierwisch, 170–204. Berlin: Akademie Verlag.

Uriagereka, Juan. 1999. Multiple Spell-Out. In *Working minimalism*, ed. Samuel David Epstein and Norbert Hornstein, 251–282. Cambridge, MA: MIT Press.

Uriagereka, Juan. 2002. *Derivations: Exploring the dynamics of syntax*. London: Routledge.

Uriagereka, Juan. 2008. *Syntactic anchors: On semantic structuring*. Cambridge: Cambridge University Press.

Uriagereka, Juan. 2009. Spell-bound: Dynamical frustration and the accordion model. ms. University of Maryland.

de Vries, Mark. 2009. On multidominance and linearization. *Biolinguistics* 3.4:344–403.

Watanabe, Akira. 1992a. Subjacency and S-structure movement of wh-in-situ. *Journal of East Asian Linguistics* 1:255–291.

Watanabe, Akira. 1992b. Wh-in-situ, Subjacency, and Chain Formation. *MIT Occasional Papers in Linguistics* #2.

Watanabe, Akira. 1996. *Case absorption and wh-agreement*. Dordrecht: Kluwer.

Watanabe, Akira. 2000. Feature copying and binding. *Syntax* 3:159–181.

Watanabe, Akira. 2006. Functional projections of nominals in Japanese: Syntax of classifiers. *Natural Language and Linguistic Theory* 24:241–306.

Watanabe, Akira. 2008. The structure of DP. In *The Oxford Handbook of Japanese linguistics*, ed. Shigeru Miyagawa and Mamoru Saito. Oxford: Oxford University Press.

Yang, Charles D. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yang, Charles D. 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8:451–456.

Zwart, Jan-Wouter. 1993. Dutch syntax: a minimalist approach. Doctoral Dissertation, University of Groningen.

Zwart, Jan-Wouter. forthcoming. Structure and order: asymmetric merge. In *The Oxford*

handbook of linguistic minimalism, ed. Cedric Boeckx. Oxford: Oxford University Press.