TOWARD BETTER DESIGN
OF THE LANGUAGE FACULTY

HIROYUKI TANAKA
Kwansei Gakuin University*

Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik,
edited by Roger Martin, David Michaels and Juan Uriagereka, MIT

Keywords: the Strong Minimalist Thesis, computational complexity,
phase, cyclicity

1. Introduction

Since the inception of the Minimalist Program (Chomsky (1993)), a
new set of deep questions about language have come to the fore. In
addition to seeking the “right theory” of the human language faculty
FL, we can now ask how “perfect” FL is. One of our tasks, then, is
to spell out the vague notion of “perfection,” and to test the theses so
formed against empirical facts. Research has since been done in both
respects, and the book under review nicely represents the state of the
art of such enterprises.

The book under review is a collection of papers on minimalist syntax
dedicated to Howard Lasnik, celebrating his twenty-fifth anniversary at
the University of Connecticut in 1997. It contains twelve chapters
(plus Introduction), each an independent study, contributed by his stu-
dents and colleagues.

The volume includes a paper by Noam Chomsky, which is an attempt

* I would like to thank two anonymous EL reviewers for their critical comments.
Part of the proposal I explored in this article was presented at a workshop of the
18th meeting of the English Linguistic Society of Japan held at Konan University on
November 18, 2000.
to make clearer what "perfection" means for the theory of language and to implement ideas that build on earlier works (especially Chomsky (1995)). All the other papers contained in the book are samples of research in the Minimalist Program or are based upon some of its important consequences. All papers are on syntax but many of them inevitably touch upon semantics-related issues such as \( \theta \)-relations, quantifier scope, and interpretation of \( wh \)-questions. Given the wide range of the topics covered, it would miss the point to take up just a few topics among them. Rather, in this review article I will try to set Chomsky's paper as a reference point for the discussion of all the other papers, since it contains a number of suggestions that depart considerably from the assumptions of Chomsky (1995), which many of the other papers take for granted. I will also critically review some of the core notions proposed in Chomsky's paper and suggest a possible move to make that fits better with the goal of the Minimalist Program.

This article is organized as follows: In section 2 I will briefly summarize the points made in each paper except Chomsky's, with brief comments where appropriate. In section 3.1 I will discuss the "Strong Minimalist Thesis" Chomsky holds and some of its derivatives, and discuss their implications for the other papers. In sections 3.2 and 3.3 I will also examine the conceptual and empirical validity of the already influential notion of phase as an illustration of an implementation of the Minimalist ideas, and suggest a possible refinement. Section 4 is allotted to the conclusion.

2. Synopses and Brief Comments

The book begins with an introduction by Roger Martin and Juan Uriagereka, who first clarify what the Minimalist Program is all about, and discuss some specific proposals made in this program in the wider context of natural science. As their discussion is mostly based on Chomsky's paper in the same volume, which I will take up in some detail in section 3, I leave it for now and turn first to the following chapters. (The title and author name(s) for each chapter appear in the subsection titles below.)

1 I refer the reader to Lotfi (2001), a review of the same volume with compact summaries and critical comments that sometimes diverge substantially from mine.
2.1. Chapter 1: “Minimalism and Asymmetric Wh-Interpretation” by Andrew Barss

Barss’s paper explores the interpretation of D-linked wh-phrases in situ. He argues that there are two strategies needed for interpreting wh-phrases that remain in situ at surface structure: one via LF movement plus Absorption (in the sense of Higginbotham and May (1981)), and one via insertion of a choice function operator (in line with Reinhart (1995)). His assumptions are: (i) the operation of Absorption optionally applies to wh-phrases that occupy the specifiers of a single interrogative C at LF, yielding a multi-pair (i.e. pair-list) reading rather than a uni-pair reading. (ii) Wh-phrases that stay in situ at LF are interpreted by a choice function operator (Reinhart (1995)), yielding a uni-pair reading. (iii) D-linked wh-phrases may or may not have a [+wh] feature, which is the formal device that makes attraction by an interrogative C possible. (iv) A minimality condition (which he calls the Shortest Movement Condition; SMC) bans wh-movement over another wh-phrase with a [+wh] feature intervening in the movement path. Under this setting, the interpretation of wh-in-situ in (1a) is predicted to differ from that in (1b) (WH₁ and WH₂ are D-linked wh-phrases; both (1a) and (1b) represent the structure at the point of Spell-Out).

(1) a. WH₁ ... t₁ ... WH₂
    b. WH₂ ... WH₁ ... t₂

In (1a) WH₂ may or may not have [+wh]. If it does, it moves covertly to Spec,CP and Absorption applies to WH₁ and WH₂ and a multi-pair reading will result. If it does not, WH₂ is interpreted in situ and Absorption never applies; instead, a choice function operator is inserted and a uni-pair reading will be yielded. Thus, we expect that (1a) displays ambiguity with respect to the interpretation of WH₂. In (1b), on the other hand, WH₁ should not have [+wh], since if it did, movement of WH₂ would be impossible due to the SMC. WH₁ therefore cannot raise even at LF, inevitably being subject to the in-situ interpretation strategy. This ensures that (1b) does not have the sort of ambiguity that (1a) has.

Barss claims that just this kind of asymmetry is observable in at least four distinct cases. To take the most representative case for example, (2a) below is ambiguous between the multi-pair reading and the uni-pair reading, while (2b) can only have the uni-pair reading.

(2) a. Which man do you think t helped which woman yester-
day?

b. Which woman do you think which man helped yesterday?

This directly reflects the mechanism; which man in (2b), unlike which woman in (2a), cannot covertly move to the matrix Spec,CP, and Absorption is inapplicable. More or less the same effect is observed in the availability of "crossed" variable binding of pronouns, the scope interaction between wh-operators and quantifiers, and the availability of what Szabolcsi (1986) terms the "comparative superlative" interpretation.

Although the judgments seem less than clear (see, for example, disagreement by Pesetsky (2000: 102, n. 46; 107, n. 72) on the judgment of (2)), the logic is clear and the theoretical implications of the proposal merit attention. As Barss himself stresses, his is a hybrid theory of wh-in-situ interpretation, in that the LF-movement and non-movement theories are not in competition but in a complementary relation.

2.2. Chapter 2: “Sometimes in [Spec, CP], Sometimes in Situ” by Željko Bošković

In Chapter 2 Bošković takes up French wh-questions with in-situ wh-phrases. A representative paradigm is (3). (3b) has the same interpretation as (3a) without moving qui ‘who.’

(3) a. Qui as-tu vu?
   who have-you seen
   ‘Who did you see?’

b. Tu as vu qui?

Rather than ascribing the apparent optionality of wh-movement to the optionality of a strong feature of C, Bošković argues that the apparent optionality comes from an interesting theoretical possibility allowed in the system of Chomsky (1995) he is adopting: that a lexical item with a strong feature and without phonological content can be merged at the root of the structure after Spell-Out. In his analysis French interrogative C always has a strong [+wh] feature but lacks phonological content. Being phonologically null, it may be inserted into the structure after Spell-Out, and then have its strong feature checked. (3b) instantiates this option, with the matrix C absent at the point of Spell-Out, and qui is covertly attracted to Spec,CP by the LF-inserted C. (3a) results if this option is not chosen.

Independently, Merge is restricted so that it may only apply in such a
way as to expand the structure under construction (Chomsky’s (1993) Extension Condition). This correctly excludes the derivation in which C is inserted into an embedded clause at LF to yield an embedded question. Another interesting consequence is that, given that the LF “wh-movement” of qui in (3b) is actually featural movement that adjoins the features of qui to C, the movement is subject to the Head Movement Constraint (HMC; in the sense of Roberts (1992)) subsumed under Relativized Minimality. The relevant feature movement to C in (3b) is to an “A’-head” position, and cannot cross another “A’-head” such as C and Neg. That bars an in-situ strategy which extends beyond a finite clause boundary or a negation.

Though Bošković’s system insightfully captures the peculiar nature of the French construction, some points remain unclear. First, in his system the crucial factor that allows a language L to have the option represented in (3b) is whether C in L is phonologically null or not. Bošković presents some evidence that the correlation does hold, but in some (important) cases the facts are obscure; thus, if the English counterpart to (3b) (You saw who?) is unacceptable as a non-echo question, it is attributed to the fact that English interrogative C, which apparently has no phonological content on its own (apart from if/whether), is in fact phonologically non-null (a “phonological affix”), which is reflected in the possibility of Subject-Aux Inversion and Do-support. But then what makes us conclude that the French interrogative C is not a phonological affix though it triggers subject-verb inversion? At least the notion of “phonological content” should be made clearer in order for his theory to be testable.

Secondly, resorting to the HMC raises a lot of empirical and conceptual questions. For example, what does his theory say about the LF feature movement across an infinitival C, rather than a finite C? If the intervention of C blocks long-distance feature movement, then why

---

2 Bošković notes that the grammatical status of You saw who? is not clear, as an EL reviewer pointed out. If it is acceptable as a true, non-echo question, then English as well as French has a phonologically null C, yielding the matrix wh-in-situ construction. The claim that follows in the text is made on the assumption that the sentence is not acceptable as a non-echo question.

3 Here the question arises as to what determines the inapplicability of SAI in English embedded questions. Bošković himself is aware of this (p. 60), but he leaves it as a mystery.
is successive-cyclic feature movement not available, just as categorial movement can make use of Spec,CP as an escape hatch? More importantly, how does A/A'-distinction (now extended to head positions) fit into the Minimalist Program? I hope both data and theoretical considerations will be added in this respect.

At this point let me note an interesting conflict between Bošković's and Barss's analyses of wh-in-situ. Unlike the wh-phrase in (3b), which is a single wh-question, those that stay in situ in multiple wh-questions do allow intervention of Neg and C between them and their scope-taking positions. (4) is from Bošković and Lasnik (1999).

(4) a. Qui croit que Marie a vu qui?
   who believes that Marie has seen whom
   'Who believes that Marie saw whom?'

b. Qui ne mange pas quoi?
   who NEG eats NEG what
   'Who doesn't eat what?'

These sentences are acceptable on the pair-list reading. This makes Bošković predict that the in-situ wh-phrases in (4) do not undergo feature movement to the matrix C and are interpreted by some device like unselective binding that does not presuppose any movement relation in syntax. On the other hand, in Barss's system (covert) syntactic movement of some sort is a prerequisite for the pair-list interpretation. (4) then forces Barss to conclude that the wh-phrases in situ in (4) do raise at LF, in a direct conflict with Bošković.

2.3. Chapter 4: “The Primitives of Temporal Relations” by Hamida Demirdache and Myriam Uribe-Etxebarria

Demirdache and Uribe-Etxebarria try to unify the syntax and semantics of Tense with those of (Grammatical) Aspect. Their proposal is summarized in (5) (their (7)), with the basic phrase structure illustrated in (6) (their (8)). (UT-T is the time of utterance, AST-T the assertion time, i.e. “the time for which an assertion is made or to which the assertion is confined (Klein (1995)),” and EV-T the event time, i.e. the time at which the event (or state) denoted by the VP occurs (or holds).)

(5) Both Tense and Aspect are dyadic spatiotemporal ordering predicates taking time-denoting phrases as arguments. The external argument of Aspect (Asp0) is a reference time (the AST-T); its internal argument is the time of the event de-
noted by the VP (EV-T). The external argument of Tense
(T0) is a reference time (UT-T); its internal argument is the
AST-T.

Postulating the three kinds of temporal arguments and relating them
to the temporal predicates (T and Asp) enables us to gain the right
semantics through the same interpretive procedure. For example,
Asp0 in the progressive aspect situates AST-T within EV-T; T in the past
tense situates UT-T after AST-T. The combination yields the interpreta-
tion of past progressive sentences. Another interesting theoretical im-
pliation is the way aspect recursion is constrained. Demirdache and
Uribe-Etxebarria suggest that the occurrence of AspP is allowed only
when it affects the interpretation, very much in line with Fox’s (2000)
study on quantifier raising. The logically possible but ungrammatical
combinations of aspects (such as the Progressive of the Perfect (e.g.,
*Rosa is having read Move a)) and the Progressive of the Progressive
(e.g., *Rosa is being reading Move a) are thus predicted to be ungram-
matical.

I think that Demirdache and Uribe-Etxebarria’s contribution is not
so much to syntax as to semantics; their system makes it possible to
understand aspect and tense under the same set of semantic primitives,
in contrast to the system of Reichenbach (1947). As for syntax, it is
not clear why the system with syntactic temporal arguments (and
semantics based on that) fares better than the one that encodes the
temporal relations directly into the Asp/T head, without recourse to
temporal arguments. Even in the latter system semantic rules can be
devised that yield the same interpretation and are of the same order of
generality as in Demirdache and Uribe-Etxebarria’s system. As far as
I can tell, there is no syntactic evidence offered which shows that the
temporal arguments occupy the specifiers as in (6). Moreover, it considerably departs from the standardly assumed theory of predicate-argument configurations to say that the internal argument of a predicate is allowed to be in the specifier of the predicate’s complement (see the structural relations between $T^0$ and $\text{Ast-T}$ and between $\text{Asp}^0$ and $\text{Ev-T}$ in (6)), a conceptually undesirable non-locality that should be eliminated if possible. (A related issue is that $T^0$ and $\text{Asp}^0$ both take $\text{Ast-T}$ as their argument; it is not uncontroversial whether such “argument sharing” is allowed by UG.)

2.4. Chapter 5: “How Far Will Quantifiers Go?” by Kyle Johnson

Johnson examines the nature of Quantifier Raising (QR) in English and assimilates it to what is called “scrambling” in German and Dutch. The first half of his paper is devoted to pointing out the difficulties with the treatment of quantifier scope advanced by Hornstein (1995) and Kitahara (1996). He then turns to the possibility of collapsing QR with scrambling. He shows that adjuncts can undergo QR, that QR can move an object out of an embedded infinitival clause even in a non-restructuring context, and that QR can take a quantifier out of a QNP that dominates it, bringing about so-called “inverse linking.” All these cases, none of which can be handled in Hornstein’s approach, have overt counterparts in German/Dutch scrambling. He further goes on to argue that the extraposition-from-NP phenomena in English are the overt manifestation of scrambling and QR.

It is interesting to note that we can see here a tension between the attempt to better describe the data and the attempt to construct a better theory. Hornstein (1995) tries to “reduce QR to one of the movement operations we independently need (p. 205),” and to identify the operation with A-movement for Case checking. Johnson identifies it with scrambling. Though Hornstein’s approach suffers many empirical problems, the rationale for the LF movement is clear. Johnson’s system overcomes the empirical problems, but lacks a theoretical basis for

---

4 They only suggest that their view supports the claim that there can be more than one specifier to a single head (cf. Koizumi (1994) and Ura (1994)), for there must be extra specifiers to host DP and other arguments in addition to the temporal arguments.

5 Johnson draws mostly on Hornstein (1995) as a representative. For a critical discussion of Hornstein’s ideas, see Kennedy (1997).
the analysis. He does not discuss why QR and scrambling in two different languages behave the same way in relevant respects. Specifically, we need at least to understand what the two operations in question have in common, and why QR in English is necessarily covert while Dutch/German scrambling is overt. Unless these questions are answered, his analysis does not go beyond pointing out the parallels between the boundedness of scrambling and that of QR.6


Murasugi discusses the structure of “relative clauses” in Japanese based on the antisymmetry theory of phrase structure (Kayne (1994)). Her conclusion is that Japanese totally lacks relative clauses in the usual sense of the term; instead, so-called relative clauses are in fact sentential modifiers (more specifically, IPs) which involve no movement relation between the head noun and the “gap” (if any) in the clause, and the overall structure of the noun phrase containing the “relative clause” is a pure complex NP as illustrated in (7).

(7) [DP [IP ...] [D’ D [NP ... [N’ N ...]]]]

She first presents a historical sketch of the studies on Japanese relatives and shows that no movement relation is involved in their formation. What appears to be a movement trace in a “relative clause” is a pro, which Japanese makes extensive use of. This non-movement property is the cause of a peculiar locality of Japanese relativization that would not be expected if movement were involved. The analysis is supported by a careful examination of acquisition data, especially showing that the relative clause is an IP, not a CP.

A theoretical problem that any “non-movement” approach to Japanese relative clauses has to solve, and hence one not restricted to Murasugi’s paper, is how to determine the semantic interpretation of relative clauses without movement relations indicated by syntax. In some cases there is no element in the relative clause that can be related to the head noun (in which case the notion of “aboutness” comes into play), and in some cases the head noun is related to a pro (in Murasugi’s terms).

6 Johnson’s assimilation of QR to scrambling and extraposition stands in an interesting contrast with Baltin’s (1987) dissociation of the two types of operations.
As for the status of relative clauses as IPs, let me note that there is some independent research, especially that concerned with nominative/genitive conversion, which relies heavily on the presence of C(P) in relative clauses in Japanese (Watanabe (1996), Hiraiwa (2001a)). Interestingly, while Hiraiwa attributes the special verb form used in Japanese relatives to the property of the relative C, Murasugi focuses on the less reduced character of the verb form in Japanese compared with N-final relatives in other languages, and speculates that this is related to the fact that the Japanese relative clause is an IP, which is a marked option (the unmarked option is a CP).

2.6. Chapter 7: “A Conspiracy Theory of Case and Agreement” by Javier Ormazabal

Ormazabal proposes a radical refinement of the theory of movement in Chomsky (1995). Ormazabal claims that it is the need for the checking of an animacy feature [+animate] that triggers what we normally construe as Case-driven movement of arguments out of vP. In his theory only [+animate] arguments have to undergo movement for animacy checking. He assumes a two-layered VP-shell for the ditransitive construction as in (8) (his (3)), where the landing site for the direct object (DO) is the outer Spec of vP and that for the indirect object (IO) somewhere outside of vP (Spec,XP).

\[
\begin{align*}
XP & \quad \downarrow \\
\mathcal{V}P & \quad \downarrow \\
DO & \quad \downarrow v' \\
\mathcal{V} & \quad \downarrow \\
\text{Subj} & \quad \downarrow v' \\
\mathcal{V} \mathcal{P} & \quad \downarrow \\
\text{IO} & \quad \downarrow V' \\
& \quad \downarrow V \downarrow t
\end{align*}
\]

The two works cited differ in several important ways with respect to the analysis of nominative/genitive conversion, though.
The Minimal Link Condition (MLC; Chomsky (1995)) then plays a crucial role in restricting the movement of arguments. If either Subj or IO is missing from (8), the equidistance mechanism ensures the checking of the two remaining arguments. But in (8), where Subj and IO are both generated and DO has moved to Spec,vP, the movement of IO to Spec,XP is blocked by the MLC because DO and Subj are not in the same minimal domain of IO or the target. It follows that the only way for the ditransitive structure to be derived is to exempt DO from moving to Spec,vP; i.e., the derivation can converge only when DO does not have a [+animate] feature. This makes it possible to account for the well-known constraint on clitic/agreement clusters, first discussed by Perlmutter (1971) and adjusted by Ormazabal, that the presence of a dative agreement/clitic and of a nominative/ergative agreement/clitic blocks the occurrence of a non-third person accusative/absolutive agreement/clitic.

Ormazabal claims that his “animacy-based” theory of movement differs from the standard Case-based theory in that an animacy feature is +Interpretable on the argument while a Case feature is -Interpretable on both the attractor and the attractee, so that his theory allows us to view feature checking as uniformly asymmetric; one of the features entering into a checking relation is +Interpretable, and survives to the interface level. A question that arises here is, what is the status of (structural) Case after all? Although Case seems to be an “imperfection,” being uninterpretable at both interfaces, Case obviously is an element of morphosyntax that cannot be denied its existence, and its systematic behavior must be captured somehow (that is what Case theory is all about). Clarification is needed about this point.

Apart from conceptual issues, Ormazabal’s theory gives rise to some technical and empirical problems to solve. For example, given that some functional head attracts [+animate] because of its uninterpretable animacy feature, what happens if there is no argument to check it off, especially in the case where the direct object does not move up in (8)? Here the characterization of movement as caused by Greed rather than as an attraction fits well into the picture, but no suggestion seems to be made. Secondly, how the system works for arguments other than DO is not discussed. It is mandatory to explore the consequences for what has been believed to be Case-driven movement, which should be far-reaching.
Saito and Hoshi present a novel analysis of the so-called light verb construction in Japanese like (9).

(9) Mary-ga John-ni/-e [NP toti-no zyooto]-o sita.
    Mary-Nom John-to/-to land-Gen giving-Acc did
    ‘Mary gave a piece of land to John.’

In the classical analysis by Grimshaw and Mester (1988), some of the arguments of the θ-role assigning noun (zyooto in (9)) are transferred to the light verb su, and the transferred θ-roles are realized at the clause level, not inside the noun phrase headed by the θ-role-assigning noun. Some interesting properties of this construction are summarized by Grimshaw and Mester as follows:

(10) a. At least one internal θ-role of the noun must be assigned to an argument outside the NP.
    b. If a θ-role T is assigned outside the NP, then all θ-roles that are higher than T in the thematic hierarchy must also be assigned outside the NP.

While accepting this generalization, Saito and Hoshi reject on conceptual grounds Grimshaw and Mester’s mechanism of argument transfer coupled with the assumption that all θ-roles are assigned at D-structure. They instead propose the “LF incorporation analysis,” in which the θ-role-assigning noun discharges some of its θ-roles after it covertly raises to the position of the light verb su. Such a derivation is possible because the Projection Principle is eliminated together with the internal representational level of D-/S-structure in the Minimalist Program. With some other auxiliary assumptions, their system makes correct predictions about the weakness of the Double-o constraint in this construction, Grimshaw and Mester’s generalization (10), and the ergativity constraint.8

8 Miyagawa (1989) and Tsujimura (1990) note that ergative nouns cannot serve as the predicate in the light verb construction as in (i), in contrast to (ii) where the noun overtly “incorporates” into the light verb.

(i) ?*Ya-ga mato-ni [NP meityuu]-o sita.
    arrow-Nom target-to strike-Acc did
    ‘The arrow struck the target.’

(ii) Ya-ga mato-ni [v meityuu-sita].
    arrow-Nom target-to strike-did
    ‘The arrow struck the target.’
Saito and Hoshi’s analysis is no doubt progress in that they dispense with the mysterious stipulations about the nature of the argument transfer operation (along with the operation itself). This research is a good example in which elimination of notions not supported by the interface requirements leads to a better understanding of empirical facts. However, it must be admitted that their theoretical assumptions are not entirely uncontroversial. The mode of $\theta$-role assignment they adopt necessarily implies, as they note, that $\theta$-roles are formal features (at least on the part of the assigner, following Lasnik (1995); this is necessary in order to subject LF incorporation to the Last Resort condition). Assimilation of $\theta$-role assignment to feature checking has far-reaching effects on the locality of $\theta$-role assignment, which is standardly regarded as distinct from feature checking configurations. Furthermore, in order to deduce (10a), it is necessary to assume that nouns need not discharge their external $\theta$-roles. Although the optionality of external $\theta$-role assignment by nouns must somehow be captured, Saito and Hoshi’s implementation makes such external $\theta$-roles unique in that they are present throughout the syntactic derivation but somehow may or may not be checked. Though what Chomsky (1995) calls +Interpretable formal features have just that property, it lacks motivation to say that $\theta$-roles of a single $\theta$-assigner differ with respect to their interpretability at LF.9

2.8. Chapter 9: “Move F and Raising of Lexical and Empty DPs” by Daiko Takahashi

Takahashi applies to the analysis of Japanese raising constructions the hypothesis that empty categories should behave differently from lexical categories with respect to movement. Adopting the theory of feature movement and Generalized Pied-Piping in Chomsky (1995), Takahashi claims that phonologically null elements like PRO and pro do not undergo category movement even in overt syntax, since the driving force of overt category movement is to avoid PF crash. Empty categories are exempt from PF interpretation, so the pied-piping of

---

9 Ura (2001) discusses some other problems of Saito and Hoshi’s paper in developing an account of the light verb construction in Japanese based on the operation Agree proposed by Chomsky (the volume under review).
such categories is unnecessary and hence cannot take place due to the economy condition that the attracted feature "carries along just enough material for convergence (Chomsky (1995: 262–263))."

Takahashi’s crucial examples are (11) and (12) (his (34) and (36), respectively).

(11) pro1 [PRO1 biiru-o nomi nagara] Hanako-o sikari hajimeta
      beer-Acc drinking while Hanako-Acc scolding began
      ‘I/We/You/He/She/They began scolding Hanako while drinking beer.’

(12) pro1 [PRO1 sigure nagara] fubuki hajimeta
      raining while snowing began
      ‘It began snowing while raining.’

Though both (11) and (12) are grammatical, the nagara ‘while’ clause can be construed either with the matrix predicate hajime ‘begin’ or with the embedded predicate sikari ‘scold’ in (11), whereas it can only be construed with the lower predicate in (12). Given that PRO in the nagara clause must be c-commanded by its controller, this fact indicates that a pro functioning as a quasi-argument cannot undergo raising as in (12). The unavailability of raising in (12) is straightforwardly accounted for by the above assumptions about empty categories, supporting Takahashi’s proposal. More mysterious is the availability of the matrix-related reading of (11), in spite of the hypothesis that pro cannot raise. Takahashi argues that this is not due to the availability of raising in (11) but due to the ambiguity of verbs like hajime as raising and control verbs. Thus, when the verb is a raising predicate the embedded subject pro is forced to remain in situ (and the matrix subject position is filled with a null expletive), forcing the lower interpretation of the adjunct clause, and when the verb is a control predicate an argumental pro is generated as the subject of the control verb, with the embedded subject PRO controlled by pro, thus allowing the upper reading of the adjunct. If pro is quasi-argumental, the option of control is unavailable for θ-theoretic reasons and the downstairs reading of the adjunct is forced.

Like Bošković’s chapter, Takahashi’s is interesting in that a novel theoretical possibility found in the course of pursuing Minimalist ideas provides a new insight into empirical investigations. One important implication of Takahashi’s analysis is that the moved features (at least
in cases like (11) and (12)) are insufficient for the purpose of control. It contributes to the understanding of how much is actually moved in feature movement. Let me also note that Takahashi’s argument constitutes significant support for the existence of raising in Japanese, which is still one of the central issues in the study of this language.

2.9. Chapter 10: “On the Grammar of Polish Nominals” by Ewa Willim

Willim’s paper discusses the internal structure of Polish nominals. From several independent pieces of evidence, she concludes that there is no projection of D in Polish. She observes that there is no lexical item to fill in the D position, that extraction out of noun phrases does not target Spec,DP, and that two genitive arguments cannot be supported by a single noun. All these facts suggest that the DP layer is lacking from this language. She further rejects any contribution by D to the interpretation of the reference of Polish nominals (and hence some other means of interpretation is employed).

Willim’s argument is tightly connected to the theory of parametric variation. When some language L apparently lacks a functional category F that other languages possess, does L have F that is active in syntax but happens to be morphologically invisible, or does L indeed exclude F from syntax (in which case the function of F is performed by some other category)? Willim argues for the latter position. This issue is not limited to the syntax of nominals but is recurrent in many other areas of comparative syntax. In order to settle this problem we would need to investigate what these candidates imply for the theory of acquisition. Specifically, the former theory states that there is a universal list of functional categories with a specified set of morphological options from which the language in question chooses. The situation is less clear for the latter theory: It seems that UG provides something like a “functional category generator” whose task is to determine on the basis of morphological/syntactic input which functional categories the target language has. Anyway, we cannot draw a conclusion until the details of the latter are spelled out.

10 A representative case is the phrase structure of Japanese. See especially Fukui (1995) for a position comparable to Willim’s.
3. The Strong Minimalist Thesis

3.1. Goal and Practice

Now we are in a position to turn to Chomsky’s paper (“Minimalist Inquiries: The Framework;” henceforth MI) and explore its implications for linguistic research including other papers in this book. Beyond asking what the human language faculty is like, the Minimalist Program aims to answer why it is that way; in other words, to what extent the Strong Minimalist Thesis (SMT), succinctly stated as (13), holds.

(13) Language is an optimal solution to legibility conditions.  

In order to make this thesis clear we have to understand its two key components; what the legibility conditions imposed by the A-P and C-I interfaces are (the problem of “design specification”) and what counts as an optimal solution (the problem of “good design”). But it should be stressed that neither of these problems is well-understood and we have no choice but to speculate and gain feedback from empirical investigation.

On top of it, it is realistically impossible to conduct empirical analyses using only concepts that are supported by the SMT. This is true even in the rather technical implementation of Chomsky himself, as we will see later. It is methodologically inevitable to stipulate unwarranted devices tentatively and see how they work, and then proceed to seek a better-grounded solution. For instance, Barss’s exploitation of insertion of choice function operators and Absorption seems to violate what Chomsky calls the Inclusiveness Condition: No new features are introduced by the computational system of human language $C_{HL}$ (MI: 113). It immediately follows that these operations are not part of syntax but belong to the semantic component. Also, some structural relations that are neither “imposed by legibility conditions [n]or fall out in some natural way from the computational process (MI: 113)” play crucial roles. Thus, Murasugi relies on the Proper Binding Condition (and necessarily on the A/A’-distinction) to restrict the structure of relative clauses and on the Empty Category Principle to derive the fact that Japanese relative clauses lack C. Saito and Hoshi likewise rely on the A/A’-distinction, though more speculatively, in the discussion of the ergativity constraint. Bošković makes use of the A/A’-distinction extended to head positions to account for the impossibility of a French
wh-in-situ strategy that crosses Neg and C. These relations and devices, too, must be recast in some other way if we are to stick strictly to the SMT.

Though we do not know much about what legibility conditions are, what was said above is rather obvious, as long as we take the SMT seriously. Still less known is what the “good design” is. The mechanisms Chomsky rejects as being complex, for instance, include feature movement (in favor of Agree; MI) and equidistance which defines closeness (in favor of simpler, c-command-based closeness; Chomsky (2001a)). The former is the core concept of Takahashi’s analysis, and the latter a crucial assumption for Ormazabal. Here, too, we see conflicts that might have implications worthy of exploration. In addition to the complexity of mechanisms themselves, the “good design” problem is also concerned with the complexity of computation performed by C_{HL}. In the next section I will critically discuss the notion phase as an example of a tentative solution to the problem.

3.2. Phase: A Case of Well-Designedness?

As noted above, one of the representative issues of the “good design” problem is that of computational complexity. Roughly put, the idea is that C_{HL} should be designed so as to minimize its operative load. Chomsky (MI) implements this idea by introducing the notion phase, which he claims is supported by considerations of computational complexity. The mechanism can be described as follows:

\[(14)\]

\[
\text{<workspace subarray}_i, \text{SO}_i \rightarrow [\text{phase} \ldots] \\
\text{<workspace subarray}_{i+1}, \text{SO}_{i+1} \rightarrow [\text{phase} \ldots] \\
\ldots
\]

(14) schematically illustrates the flow of derivation by phase. The lexical array from which C_{HL} selects the lexical items (LIs) to construct a sentence (the initial lexical array) is divided into some subarrays of LIs, and the derivation proceeds accessing “one subarray at a time,” i.e., while working on a certain subarray A, it does not access LIs of other subarrays until it finishes with A. (To “access” or “work on” LIs means to execute operations that integrate them into the structure or
delete their uninterpretable features.) SO in (14) represents the set of syntactic objects already constructed. The operations of $C_{HL}$ work on the pair of the lexical subarray and SO to build a new syntactic object. The syntactic object so formed is called a *phase*, and can serve as an input to the construction of a larger phase. In MI a phase is supposed to be a CP or a vP.

The empirical motivation for introducing phases is perhaps to handle examples like (15) (his (7b)).

\[(15) \text{ there is a possibility } [\text{that proofs will be discovered}]\]

Granted that movement of an argument into Spec,TP is blocked by Merge of an expletive, it had been a mystery why the expletive *there* does not block the raising of *proofs* from the object position of *discover* to the embedded Spec,TP in (15). Now that $\alpha = \text{CP}$ is a phase and *there* and *proofs* belong to different lexical subarrays, the comparison between Move and Merge does not arise and *proofs* can successfully move. The notion of phase is, in effect, introduced in order to make appropriate portions of sentences a “closed system (p. 103).”

So far the necessity of phases seems to be supported on both conceptual and empirical grounds. But upon closer scrutiny a number of problems arise.

First of all, it should be noted that the introduction of phases clearly goes against “methodological minimalism” in Martin and Uriagereka’s terms, or, more generally, Occam’s razor: The smaller the number of axioms, the better. The question, then, is whether there is sufficient support to justify the new theoretical construct. In particular, the Minimalist Program asks whether the move gains support from the SMT (13).

Since the conceptual motivation for phases is to reduce computational complexity, it is not surprising that phases are not supported by any requirement imposed by the interfaces. Then, let us ask if phases really contribute to the reduction of complexity. In my view, the reduction is not so significant. When the earliest models of minimalist syntax were criticized for requiring too much computational burden caused by the need for trans-derivational comparisons (global economy), dispensing with global economy in favor of local economy (cf. Collins (1997)) was a serious move towards avoiding exponential blow-up.\[11\]

\[\text{11 Under a certain interpretation of global economy, the complexity in the re-}\]
Now we are comparing a theory with phase and a theory without it. Given that global economy is totally eliminated, the computational burden is evaluated in terms of the capacity of the “workspace” needed to determine which operation to apply at each step in the derivation. In this situation, the theory with phases surely has fewer lexical items to choose from at each step than the theory without phases, but even in the latter the number of LIs to be kept in the workspace is at most $n$, where $n$ is the number of LIs initially selected for the derivation. (That is, the complexity in the latter is only proportional to $n$ to the first power.) Thus, the introduction of phases does not contribute so much to reducing the computational complexity as the shift from global to local economy did. Chomsky appears to be aware of this, but says that having to do more might be considered “poor design (pp. 99–100).” As noted by Chomsky and also by Martin and Uriagereka, one can only guess what good design is, so what Chomsky says in this regard is not a necessary consequence, though it is interesting if something along these lines is true. (In the next subsection I will nonetheless pursue the basic idea of phase in a radically different direction from Chomsky’s.)

Chomsky suggests that a phase makes invisible the elements inside it except for its specifiers and head. He calls this the Phase-Impenetrability Condition (PIC), stated as in (16), where the domain of $H$ is its complement and the edge its specifier(s).

\begin{equation}
\text{In phase } \alpha \text{ with head } H, \text{ the domain of } H \text{ is not accessible to operations outside } \alpha, \text{ only } H \text{ and its edge are accessible to such operations.} \quad (MI: 108)
\end{equation}

Although PIC may fall out as a natural consequence of the closednesslevant sense grows in proportion to $n!$, in fact faster than the exponential function $2^n$ (where $n$ is the number of LIs initially selected for the derivation; see Johnson and Lappin (1999)).

\footnote{An \textit{EL} reviewer objects to this conclusion, claiming that “if there are $n$ [LIs] in the workspace, then there are $n!$ possible orders for selecting [the LIs] one at a time” and “[e]liminating global economy does not affect this calculation.” Though it is true that there are $n!$ possible orderings for $n$ items, local economy does not require that we perform all the $(n!)$ possible paths of derivation and pick out the best one, which is exactly what global economy requires. Under local economy one LI is chosen from the $n$ LIs at a time and the choice is fixed once and for all, so there is no need to list all the possible derivations.}
of phases, the exception of the head and the specifiers of a phase is nothing but a stipulation, required for empirical reasons. Note also that the PIC is only a “suggestion,” extending the closedness of phases. It is equally conceivable that the domain of H is accessible to operations outside α and still the derivation proceeds phase by phase; the validity of the PIC and that of the phase itself are independent.

Moreover, there is evidence that the PIC is empirically suspect when the Case/agreement systems of languages parametrically different from English are considered. The PIC predicts that nothing contained in the domain of C or v can be related to a head outside the phase, hence there should be no Case assignment (executed through the operation Agree) across the boundaries of TP/VP selected by C/v, respectively, or movement across such boundaries motivated by Agree. The existence of overt super-/hyper-raising (Ura (1994)), in which arguments inside a finite clause move out of it (without landing in the specifier position of the finite CP), is straightforward counterevidence. The Case system of Japanese also provides evidence against PIC. Let me present some representative examples. (17) is what Ura (1994) analyses as an instance of “covert super-raising.”

(17) boku-wa [CP [TP John-ni piano-o hik-e-ru] to] I-Top John-Dat piano-Acc play-can-Pres Comp
    omow-u
    think-Pres
    ‘I think that John can play the piano.’

In (17) the accusative Case of the embedded object piano ‘piano’ cannot be assigned by the embedded verb hik ‘play’ but rather must come from the matrix verb omow ‘think,’ since the embedded complex predicate itself cannot assign accusative Case, as seen in (18a), and (17) is degraded if the matrix verb is passivized, as seen in (18b).

(18) a. *John-ni piano-o hik-e-ru
    John-Dat piano-Acc play-can-Pres
    ‘(int.) John can play the piano.’

---

13 I will henceforth assume the following basic clause structure for active transitive sentences in Japanese (the effects of movements are omitted):

( i )  [CP [TP [vP AGENT [vP THEME V] v] T] C]
Recast in the current terminology, the matrix verb triggers Agree with the embedded object to assign accusative Case in (17). In this case Agree operates across the domain of C.

“ECM into CP” (Kuno (1976)) also provides similar evidence. In (19a) the verb omow ‘think’ assigns accusative Case to Bill inside the finite TP selected by C. That Bill is assigned Case by the matrix verb is confirmed by the impossibility of passivization of the matrix verb as shown in (19b).14

John-Nom Bill-Acc genius be. Pres Comp think-Past
‘John thought that Bill was a genius.’

b. *[CP [TP Bill-o tensai da] to] omow-are-ta
Bill-Acc genius be. Pres Comp think-Pass-Past

One might object that in (19a) Bill raises to the edge of CP (attracted by the P-feature of C; see below) and is assigned Case from the matrix verb, avoiding a violation of the PIC. It is therefore necessary to confirm that the embedded subject is allowed not to raise out of the embedded clause. (20), cited from Hiraiwa (2001b), shows that the ECM subject is allowed to stay inside the embedded clause (as indicated by the precedence of the adverb mada ‘still,’ which modifies the predicate of the embedded clause) and is assigned accusative Case.

(20) John-ga [CP [TP mada Mary-o kodomo da] to]
John-Nom still Mary-Acc child be. Pres Comp omow-ta
think-Past
‘John thought that Mary was still a child.’

This suffices to show that Agree can operate across the domain of C.15

---

14 (19b) is acceptable under the irrelevant reading as an indirect passive with null arguments.
15 Thanks to the EL reviewers for pointing this out for me. There has been much debate as to this point, in different theoretical contexts. See Sakai (1996) and references cited therein. Sakai (1996) argues that the accusative-marked
Japanese also exhibits Case-assignment across the domain of \( \nu \). Consider (21).

\[
\begin{align*}
(21) ~ (\text{?}) & \text{John-ga } [_{VP} [_{VP} [_{SC} \text{Bill-o atama-ga yo-ku}]] -\text{ta}\text{omow}] -\text{ta}^{16} \\
& \text{John-Nom Bill-Acc head-Nom good-Inf} \\
& \text{think-Past} \\
& \text{‘John thought Bill very clever.’}
\end{align*}
\]

In (21) the matrix finite \( T \) assigns nominative Case to the small clause constituent \( \text{atama} \) ‘head,’ there being no other source of nominative Case in the structure. Thus (21) shows that Agree can apply across VP, the domain of \( \nu \), contrary to what the PIC predicts.

In addition to empirical problems, the PIC forces us to make an undesirable assumption. The PIC restricts feature-driven movement to the internal domain of a phase, so any movement out of a phase (typically long-distance \( \lambda \)-movement) must make use of one of the phase’s specifiers as an “escape hatch” (that is why the exception is added to the PIC). However, at the time this intermediate step of long-distance movement takes place, the structure above the phase being constructed is not yet present and the motivation for the movement is lacking. Given that no look-ahead is allowed for complexity considerations, the only way to justify the movement is by optionally assigning a feature (which Chomsky calls a \textit{P-feature}, “P” for “periphery,” related to topic, focus, etc.) to the phase head. This is obviously a violation of the Inclusiveness Condition, which prohibits introduction of new features not present in the lexical items in the course of the derivation.\(^{17}\) The sug-

\(^{16}\) Though some informants (including one of the \textit{EL} reviewers) find (21) less than perfect, it is far better than (i), which is totally unacceptable for any informant I have consulted.

\[
\begin{align*}
\text{(i) } & \text{John-ga } [_{VP} [_{SC} \text{Bill-ga atama-o yo-ku}]] -\text{ta} \\
& \text{John-Nom Bill-Acc head-Nom good-Inf think-Past} \\
& \text{‘(int.) John thought Bill clever.’}
\end{align*}
\]

See Tanaka (2000) for a discussion of this construction.

\(^{17}\) One possible solution to this problem suggested by an \textit{EL} reviewer is to make \textit{P}-feature an optional property of a phase head whose presence or absence is determined before the derivation starts.
gestion that a QU-feature, motivating QR, may be postulated suffers the same problem.\textsuperscript{18} Thus, the PIC, even if it is conceptually motivated, leads to postulation of a conceptually unmotivated theoretical construct.\textsuperscript{19}

Another ironical consequence (this time more practical) of adopting phases is that it has invoked research, including Chomsky’s subsequent work (Chomsky (2001a)), that point to a more representational characterization of movement minimality: Chomsky (2001a), for example, suggests that the minimality of movement is evaluated after each phase is completed, the steps during the construction of a phase being ignored with respect to minimality. Here we must ask what this “minimality evaluation” is. The Minimal Link Condition of Chomsky (1995) was part of the operation triggering movement and minimality was a consequence of applying one operation at a time. In Chomsky’s (2001a) system we must have a mechanism to scan a representation and check whether minimality is respected. While syntactic operations are restricted to those essential for structure building and convergence, this kind of checking mechanism lacks support either in terms of legibility conditions or in terms of efficiency of computation. Though the idea that a derivation proceeds cyclically on the basis of phases seems at first glance to be in harmony with the derivational approach to \( C_{\text{HL}} \) with much emphasis on the reduction of computational complexity, the newly proposed approach to minimality totally goes against that orientation.

3.3. Cyclicity of \( C_{\text{HL}} \) Operations as a Consequence of Computational Complexity Reduction

As summarized in the last subsection, Chomsky (\textit{MI}) takes phases to be CP and \( \nu P \), giving some possible semantic motivations (i.e. to represent a proposition). But this assumption, too, raises some questions. At the very least, we do not have an answer to the question why these categories (propositions) are selected as phases. Furthermore, if we

\textsuperscript{18} Though Johnson’s paper may be reconsidered from that perspective.
\textsuperscript{19} In addition, if the P-feature has semantic import, it is immediately predicted that all and only the phases in the path of such long-distance movement have some semantic property attributed to this feature, but this point is not discussed in \textit{MI} at all.
want efficient computation, with as small a workspace as possible, the number of LIs in the lexical subarray for constructing a phase should be much smaller. When we push it to the extreme, the number should reduce to just one. In this section I explore the possibility that this might in fact be the case: The workspace is so limited that $C_{HL}$ can work on only one lexical item at a time, and this provides a natural basis for the type of cyclicity which Chomsky independently suggests: that the operations of $C_{HL}$ are applied cyclically.

In $MI$, two types of cyclicity are operative. One is based on the phase, and the other is based on LIs, stated as in (22) (=his (53)):

\[(22) \text{ Properties of the probe/selector $\alpha$ must be satisfied before new elements of the lexical subarray are accessed to drive further operations.}\]

That is, when an LI triggering an operation (probe/selector) is accessed, no other LI can be accessed until all the properties of $\alpha$ are satisfied. Chomsky speculates that (22) derives from the need to reduce operative complexity that stems from searching the probe/selector. Though both types of cyclicity are derived from complexity considerations, the two are of course independent from each other. Now that phases as they are formulated in $MI$ is problematic in many ways as discussed so far, there are two options to choose from: either to eliminate the notion of phase altogether, or to leave the essential motivation as it is and elaborate on it to an extent expected under the Minimalist Program. It is the latter approach that I would like to take.

Let me spell out the idea more clearly. I basically adopt the general style of derivation illustrated in (14), according to which the initial lexical array is divided into subarrays of LIs and the computation attends only to the subarray it is working on at each point in the derivation. It does not go back to the items in subarrays it has already processed, nor does it look ahead at the items in subarrays yet to be accessed. Unlike in $MI$, however, I take the need for a smaller workspace to restrict $C_{HL}$’s access to lexical items to one, eliminating the labor of determining which LI to process. That is, every subarray in (14) contains exactly one LI taken from the initial lexical array, in contrast to the $MI$ system where a subarray is an arbitrary set of LIs taken from the initial array containing exactly one $v$ or C. On the other hand, in order to integrate lexical items into the structure, the workspace must be large enough to hold the syntactic objects already constructed, i.e.,
it must contain SO as in (14). Limiting the workspace to just these elements, $C_{HL}$ cannot access the features of other LIs until it finishes with the LI it is working on, which is equivalent to (22). Thus, a radical strengthening of the idea underlying phase converges with the independently motivated mechanism of cyclicity.

In short, we abandon the phase in Chomsky’s sense in favor of much narrower partition of derivation, which forces the computation to proceed one LI at a time. We do not have the PIC, either, which, as I pointed out in 3.2, inevitably calls for arbitrary exception (for the edge and the head) and a device (like a P-feature) to ensure that a certain element within a phase can move out of it.

The empirical effect of this result is not trivial. According to my proposal (and equivalently by (22)), the computational system accesses one LI at a time, so the configuration illustrated in (23) never arises at any point in the derivation, in which X has some uninterpretable feature undeleted and Y is merged with XP to project YP.

(23) $[^{YP} Y [^{XP} X(F_{-}interpret) ...]]$

This does not only derive the strict cyclicity effect for overt movements\(^{20}\) but it also requires that “covert” operations (i.e., those that do not induce displacement) be necessarily cyclic; even if X in (23) does not have the property of overtly attracting something into its Spec and Y does have such a property, operations satisfying X’s uninterpretable features must be executed before Y is merged into the structure.

Let me illustrate one case in which this mechanism plays a crucial role. Consider the ungrammatical (24) involving an inverse assignment of nominative and accusative Cases in an active transitive sentence in Japanese.

(24) *John-o booru-ga ket-ta
J.-Acc ball-Nom kick-Past
‘(int.) John kicked the ball.’

---

\(^{20}\) The Path Containment Condition effect (Pesetsky (1982)) is a classical example.

(i) $[^{CP} \text{Who}_i \text{ do you know } [^{CP} \text{what books}_j \text{ to persuade } t_i \text{ to read } t_j]]$?
Without any mechanism that ensures cyclicity, (i) can be generated without violating superiority/the Minimal Link Condition by moving who to the matrix CP Spec and then raising what books to the lower CP Spec. The version of cyclicity proposed here straightforwardly excludes such a derivation; the derivation cannot proceed to build up the matrix clause without first satisfying the requirement of the embedded $C$. 
Suppose we do not have the mechanism to ensure the cyclicity defined above and a step like (23) is allowed. Then there is no clear way of blocking (24) under the present assumptions. The “wrong derivation” proceeds as in (25) (head-movement of V to v is omitted):

(25)  

After the logical subject and object are projected within vP, the V-v complex overtly raises to T, without v having induced Agree at this point (step <1>). v then induces Agree from the T-adjoined position, searching for the closest DP in its c-command domain, which happens to be the logical subject John, and assigns accusative to it (step <2>). Now T’s EPP-feature attracts John, the closest element to T, to Spec,TP (step <3>), and then T Agrees with the object ball to assign nominative, since at this point nothing blocks Agree between T and the object, the subject having moved up to Spec,TP (step <4>). Thus the derivation converges and (24) can be wrongly generated. Though it is easy to reject the possibility of this derivation by altering some of the auxiliary assumptions mentioned in notes 21-23 (overt V-raising, the definition of c-command, and invisibility of A-movement traces), it is by no means a principled solution. (And all of them are empirically well supported.) In contrast, under our version of cyclicity, (24) is blocked because the logical subject generated in the Spec,vP, not being within the search (c-command) domain of v, has no chance to Agree with v, and v must finish Agree before T is merged with vP. That

---

21 See Koizumi (2000) for evidence that the verb in Japanese undergoes overt raising.
22 Assuming that the upper segment of T⁰ does not prevent v from c-commanding John.
23 I share the assumption with MI that A-movement traces are invisible for the operations of C_HL.
is why the external argument is never marked with accusative in a simple clause. Thus, the mechanism motivated by computational complexity reduction provides a natural account for a phenomenon which would otherwise require arbitrary technical modifications.\textsuperscript{24} Note also that a covert operation (Agree triggered by $\nu$) must precede an overt one (Move triggered by $T$), which is impossible in a model that distinguishes overt syntax from covert syntax.

It might be objected that the wrong derivation depicted above is not a problem if we adopt the phase system and the PIC of $MI$, as the nominative Case assignment to the logical object in (25) crosses the $vP$ phase. Apart from conceptual and empirical problem of phases and the PIC I mentioned in 3.2, the necessary postulation of a P-feature makes it possible to overgenerate (24) via a different derivation, where the P-feature assigned to $\nu$ attracts the object to its edge and Agree between $T$ and the object can be executed without violating the PIC.

As mentioned above, PIC is inoperative in the system I am proposing here, so that it has no empirical effects that would come from Chomsky's implementation of phases, and my critical discussion of it in section 3.2 remains intact. Note that I have to admit that my proposal wipes out the empirical effect of the phase in $MI$ in that I have no way of ruling in (15), but that is a problem only if we stick to the assumption that Merge pre-empts Move. This assumption has certainly been playing an important role in developing the theory of economy, but it is also true that it is not a conceptual necessity. (See, among others, Chomsky's (2001b) own conjecture that it is false.) As long as it is not fatal for us to drop this assumption, we should do away with the phase in the sense of $MI$ and seek a different solution for (15) and related data.

To sum up section 3, I have taken up one influential proposal made in $MI$, i.e. that of the phase, and critically reviewed its status in the context of the Strong Minimalist Thesis. I also pointed out some pieces of empirical evidence that argue against the postulation of

\textsuperscript{24} In Tanaka (2000) I argued that the same mechanism correctly excludes a counter-cyclic derivation of (i) of note 16, but the argument is too complex to reproduce here. Elsewhere (Tanaka (1998)) I also noted that the mechanism compels us to abandon the theory of Case-checking based on Agr projections in Chomsky (1993).
phases. I further suggested that the original motivation for phases (i.e. reduction of the workspace) does not in fact provide a support for the phase in Chomsky’s sense, but for an independent mechanism of cyclicity.

4. Conclusion

In this review article I have considered it necessary to evaluate the chapters in terms of the extent to which each of them keeps to the guideline of the Minimalist Program while achieving the level of empirical analysis that merits attention. As is often said, one of the methodological virtues of the Minimalist Program is that it causes a tension that drives new research: the tension between the quest for “perfection” of language and the quest for explanatory/descriptive adequacy. The tension is quite severe, as I have pointed out throughout. However, the actual results amply documented in the book under review should be taken to show how much of the Minimalist goal we do not understand and how much we are beginning to understand. The situation is promising insofar as the Minimalist questions like those taken up in section 2 and 3 here continue to be seriously considered.

REFERENCES


Chomsky, Noam (2001b) “Beyond Explanatory Adequacy.” ms., MIT.


Kurosio, Stanford, CA and Tokyo.


1-58-14 Yurinoki-dai
Sanda-shi
Hyogo 669–1324
e-mail: htanaka@kwansei.ac.jp