Proceedings
of
Sophia University
Linguistic Society
No. 16

上智大学言語学会会報
第16号

2001年

上智大学言語学会

TWO NOTES ON THE ANALYSIS OF NP-MOVEMENT IN THE MINIMALIST PROGRAM

Mamoru Saito
Nanzan University

1. Introduction

The main purpose of this workshop, as I understand it, is to explore possible ways in which empirical research on languages can be related to the Minimalist program. I share the view of the other speakers and believe that there is nothing special about the Minimalist program in this regard. It is a research program in linguistics, which is an empirical science. In this paper, I will try to substantiate this point by presenting two notes on NP-movement and related issues. It seems to me to be appropriate to take up NP-movement in this context because the efforts to explain its properties have played a fundamental role in the development of the Minimalist theory.

In the following Section, I will briefly discuss the developments in the analysis of NP-movement in the last twenty years. In the 1980's, we had a number of conditions constraining the operation with much redundancy. The efforts to eliminate the redundancy and to sharpen the explanation have led to the postulation of two major conditions in the Minimalist program: the Last Resort Principle and the Minimal Link Condition (MLC). Then, Chomsky (1995) has proposed to eliminate these conditions by deriving their effects to a large extent directly from the definition of Attract/Move. This proposal not only has clear conceptual advantage but desirable empirical consequences as well. I hope the discussion in this section will be persuasive enough to convince the skeptics that the Minimalist program is pursued in much the same way as the "earlier" theories it developed from, although the issues it raises are deeper and more general reflecting the development of the field.

The second note is presented in Section 3. There, I will discuss the improper movement phenomenon, which has resisted a satisfactory account over the years. I will first suggest a specific instantiation of the idea proposed in Chomsky 1998, 1999 that interpretation applies cyclically. (See also Epstein, et al. 1998 for much relevant discussion on the derivational approach to interpretation.) Then, I will argue that the proposal makes it possible to provide a principled account for improper movement. This is a case study in how independently motivated new proposals may solve old outstanding problems. Section 4 concludes the paper with a brief discussion of Wh-movement.

2. The Analysis of Illicit NP-Movement

In this section, I will illustrate the developments in the analysis of NP-movement in three stages: (i) the 1980's, when many conditions on the operation were entertained, (ii) the early 1990's, when most of the core data were attributed to the Last Resort Principle and the MLC, and (iii) the proposal in Chomsky 1995 to eliminate these two conditions. With the final proposal, there are no independent conditions on NP-movement, and the illicit cases are explained in terms of derivational economy and the definition of Attract/Move. This history, I believe, is a clear example of advancement in empirical science.

2.1. The Overlapping Conditions on NP-Movement

In the LGB theory, Binding Condition (A) and the ECP are the two major constraints on NP-movement. (2a) and (2b) represent the NIC and the SSC subcases of Condition (A) effects respectively.

(1) a. Mary$_i$ was praised t_i by everyone
b. Mary$_i$ is believed [$_{IP}$ t_i to be a genius]

(2) a. *Mary$_i$ is believed [$_{CP}$ (that) [$_{IP}$ t_i is a genius]]
b. *Mary$_i$ is believed [$_{CP}$ (that) [$_{IP}$ John saw t_i]]

The ECP effects are often discussed with the paradigms in (3) and (4).

(3) a. John believes [$_{IP}$ Mary to be a genius]
b. Mary$_i$ is believed [$_{IP}$ t_i to be a genius]
c.*John believes [$_{IP}$ PRO to be a genius]

(4) a.*Mary decided [$_{CP}$ [$_{IP}$ John to go to college]]
b.*John$_i$ was decided [$_{CP}$ [$_{IP}$ t_i to go to college]]
c. Mary decided [$_{CP}$ [$_{IP}$ PRO to go to college]]

Mary receives Case from *believe* in (3a). In (3b), the trace is governed and hence properly governed by the verb, satisfying the ECP. On the other hand, PRO is illicitly governed in (3c) and the example is ruled out. Given that a CP/IP pair constitutes a barrier for government and consequently for Case assignment, the opposit pattern obtains in (4).

The ECP effects on NP traces are examined further in Lasnik and Saito 1984 and Chomsky 1986a. Based on a reformulation of the definition of proper government, Lasnik and Saito argue for an ECP account for the contrast in (5).

(5) a. John$_i$ [$_{VP}$ is [$_{AP}$ likely [$_{IP}$ t_i to win the race]]]

b.*John$_i$ seems [$_{CP}$ that [$_{IP}$ it is likely [$_{IP}$ t_i to win the race]]]

According to their analysis, neither trace in (5) is lexically governed and further, antecedent government is blocked by the intervening CP/IP pair in (5b). The severe ungrammaticality of (5b) is thus attributed to the ECP.[1] On the other hand, Chomsky points out that the following example, originally due to Mark Baker, has the same status as (5b):

(6) *John$_i$ seems [$_{CP}$ that [$_{IP}$ it was told t_i [$_{CP}$ that Mary won the race]]]

As the trace is obviously lexically governed, he suggests that NP traces (and possibly all traces) must be antecedent governed.

There are other examples of illicit NP-movement that received attention in the mid-1980's. For example, Lasnik (1985) discusses cases like (7).

(7) a.*John$_i$ seems [$_{CP}$ (that) [$_{IP}$ [$_{NP}$ the proof [$_{IP}$ t_i to be guilty]] was given]]
b.*John$_i$ seems [$_{CP}$ (that) [$_{IP}$ [$_{NP}$ his$_i$ proof [$_{IP}$ t_i to be guilty]] was given]]

In these examples, the NP *John* moves from the position of the trace to the matrix subject position across a CP/IP pair. But in (7b), since the NP *his* can serve both as a local binder and as an antecedent governor, the trace violates neither Condition (A) nor the ECP. Lasnik takes examples of this kind as evidence for the Locality Condition on Chains, initially suggested in Chomsky 1981. Another important example is (8), discussed in Chomsky 1986b.

(8) *John$_i$ seems to t_i [$_{CP}$ that Mary is a genius]

Like (7), the ungrammaticality of this example cannot be attributed to Condition (A) or the ECP. Chomsky proposes the Last Resort Principle, which dictates that movement is allowed only when it is necessary to satisfy a morphological requirement of the moved item. Applied to NP-movement, this roughly means that an NP moves only for the purpose of receiving Case. One case that this principle prohibits is NP-movement from a Case position as in (8).

As pointed out explicitly in Lasnik and Saito 1992, the four conditions mentioned above have much overlap in their effects. For instance, the examples in (2) are ruled out by Condition (A), the Last Resort Principle and the ECP. Clearly, it is desirable to eliminate this kind of redundancy. Lasnik and Saito first note that aside from (8), all cases of illicit NP-movement involve movement across a barrier,

[1] The example violates Condition (A) as well. But this does not explain its status since an intervening expletive subject only induces a weak Condition (A) violation. See Chomsky 1981 for discussion on this point.

or more specifically a CP/IP pair.[2] Then, building on a suggestion in Chomsky 1986a that the ECP should be construed as a condition on chains, they propose the following constraint:

(9) Suppose that α bears a θ-role assigned by β. Then, if γ is a barrier for β, γ dominates α.

This constraint in effect prohibits NP-movement across a barrier. It hence rules out (2a-b), (4b), (5b), (6) and (7a-b), and consequently yields Condition (A) and the ECP redundant as conditions on NP-traces.

According to this theory, NP-movement is constrained by (9) and the Last Resort Principle.[3] This is a clear case of an effort to eliminate the redundancy among the conditions on NP-movement, and to have a unified analysis for a larger domain of facts. However, (9) has a stipulative flavor and could not play the unifying role in this endeaver. What played this role instead was the MLC, to which I will turn in the following subsection.

2.2. The Initial Developments in the Minimalist Theory

The MLC developed out of Relativized Minimality, which Rizzi (1990) proposed as a condition on antecedent government. In the initial formulation of the MLC, Chomsky and Lasnik (1993) adapt the idea as a condition on movement, which states roughly that movement must proceed through every possible landing site. The condition, then, prohibits Wh-movement across a CP Spec and NP-movement across an IP Spec.

In addition to its conceptual attractiveness, the condition has far reaching empirical effects. It, like Relativized Minimality, accounts for the "super-raising" example in (5b), and also for the other cases of NP-movement across a subject in (2b), (6) and (7b). Further, it covers the Wh-island effect on Wh-movement, and thus, makes the following simple formulation of the CED/Subjacency possible:

[2] This coincides with Aoun's (1981) observation that "an S' breaks an A-chain."

[3] Note that this raises doubts on the existence of NP traces. Unlike Condition (A) and the ECP, neither (9) nor the Last Resort Principle requires them. However, Lasnik and Saito defended them, arguing that the examples in (i) are to be excluded by the Proper Binding Condition.

(i) a.*[How likely t_i to be a riot]$_j$ is there$_i$ t_j
b.*[How likely t_i to be taken of John]$_j$ is advantage$_i$ t_j

See Kuno 2001 and Saito 2001 for further discussion on this issue.

(10) A maximal projection is a barrier for movement if it is not a complement.

Note that the Wh-island effect had been the major problem in the Bounding theory. Huang's (1982) formulation of the CED (like those of its predecessors in Cattell (1976) and Kayne (1981)) was similar to (10), but he had to retain the classical Subjacency based on bounding nodes as well in order to account for the Wh-island effect. Chomsky (1986a), pointing out the inconsistencies between the CED and the classical Subjacency, made extensive efforts to incorporate the Wh-island effect into the CED. This attempt deepened our understanding in many respects, but overall resulted in much complication in the theory. Now that the MLC covers the Wh-island effect, it becomes possible to return to the simple formulation in (10).[4]

Along with the MLC, the Last Resort Principle played a major role in the development of the analysis of NP-movement. Its interaction with the theory of PRO is particularly important in the present context. Recall first that Chomsky (1986b) discusses two problematic cases for the explanation of the Case Filter effects by Aoun's (1979) Visibility Condition in (11).

(11) An NP is assigned a θ-role only if it has Case.

One has to do with expletives and the other concerns PRO.

(12) a.*It is likely [$_{IP}$ there to be a man in the room]
b. John tried [$_{CP}$ [$_{IP}$ PRO to win the race]]

(12a) appears to indicate that *there* needs to be assigned Case despite the fact that it does not receive a θ-role. Given the LGB account for the distribution of PRO, (12b) is problematic because PRO is ungoverned and hence lacks Case, but is still assigned a θ-role. Chomsky (1986b) proposes a solution for the first problem with his expletive replacement analysis. According to this analysis, the indefinite NP *a man* in (13) raises covertly to the matrix subject position in order to receive Case.[5]

(13) There is a man in the room

---

4 (10) itself is a generalization rather than a principle and requires an explanation. See Takahashi 1994 for relevant discussion.

5 This implies that the Case Filter (or the Visibility Condition) applies only at LF. Since it was the main principle that supported S-structure as a level of representation, this marks the first major step toward the elimination of S-structure. Note also that the notion 'government' was initially proposed based on the S-structure application of Case theory. Thus, the shift of Case theory from S-structure to LF undermines the basic motivation for the notion, and initiates the research project to eliminate it from the theory of syntax. See, for example, Saito 1996 for a discussion on this project.

The expletive is replaced with this operation, and as the result, the example satisfies Full Interpretation. According to this analysis, *there* itself need not be assigned Case. But it must be in a Case position so that it can successfully be replaced with covert movement. If it is not in a Case position, the Last Resort Principle prohibits any NP from moving to its position and hence it remains at LF in violation of Full Interpretation.

Chomsky (1986b), however, left the problem of PRO unsolved, and tentatively assumed the following formulation of the Visibility Condition:

(14) An NP is assigned a θ-role only if it has Case or is PRO.

At this point, the proposal that PRO is assigned Case is almost there. And the proposal is actually made in Chomsky and Lasnik 1993 based on examples such as (15).

(15) John tried [$_{CP}$ [$_{IP}$ PRO$_i$ to be elected $\underline{t}_i$]]

In this example, PRO clearly moves to the embedded subject position. Since the Last Resort Principle allows NP to move only to receive Case, it implies that PRO is assigned Case at the landing site in (15). Chomsky and Lasnik, then, conclude that PRO needs to be assigned null Case by a non-finite INFL. The null Case hypothesis is developed further in Martin 1992, where he argues that only the non-finite INFL with unrealized future tense (in the sense of Stowell 1982) is capable of assigning this Case. This explains the distribution of PRO in place of the PRO theorem, and accounts, for example, for the contrast between (15) and (16).

(16) *John believes [$_{IP}$ PRO to be smart]

One important implication of the null Case hypothesis is noted in Boskovic 1995. What initially motivated the application of the ECP to NP traces were examples such as (4b). The paradigm in (4) is repeated below as (17).

(17) a.*Mary decided [$_{CP}$ [$_{IP}$ John to go to college]]
b.*John$_i$ was decided [$_{CP}$ [$_{IP}$ $\underline{t}_i$ to go to college]]
c. Mary decided [$_{CP}$ [$_{IP}$ PRO to go to college]]

But given the null Case hypothesis, (17b) falls under the Last Resort Principle. The Case must be available in the initial position of *John*, because PRO is allowed in this position as shown in (17c). This means that the movement in (17b) originates in a Case position. The example, then, is a clear case of a Last Resort violation.

At this point, most cases of illicit NP-movement fall under the MLC and the

Last Resort Principle. As noted above, the MLC covers cases of movement across a subject: (2b), (5b), (6) and (7b). The Last Resort accounts for (2a-b), (4b)= (17b) and (8). The only case that is left unexplained is (7a).

2.3. The Elimination of the Last Resort Principle and the MLC

Despite the empirical success of the Last Resort Principle and the MLC, Chomsky (1995) takes up the project to eliminate them from the theory of syntax. The most important motivation for this move is conceptual: although these conditions were entertained as "principles of economy," they actually have no place in the theory of derivational economy.

As Chomsky has repeatedly pointed out, derivational economy must be stated in relative terms in the sense that it should allow necessary operations. Otherwise, no operation and hence no derivation would be possible. Thus, the NP-movement in (1a), repeated as (18), can take place because it makes the necessary feature-checking possible and allows the derivation to converge.

(18) Mary$_i$ was praised t_i by everyone

But then, the NP-movement in (19) should be allowed for the same reason.

(19) a.*John$_i$ seems [(that) t_i is intelligent]
b.*John$_i$ seems [(that) it is likely [t_i to win the race]]

In these cases, the derivation would certainly fail to converge without the movement. This indicates that the Last Resort Principle and the MLC are imposed on syntactic derivations independently of derivational economy. One goal of the Minimalist program is exactly to eliminate such conditions.

Chomsky first proposes an alternative account for the typical Last Resort violations by making the theory of formal features more precise. He divides the instances of the following four types of features into interpretable ones and uninterpretable ones.

(20) a. categorial features
b. ϕ-features
c. Case features
d. strong F, where F is categorial (= the EPP feature)

Categorial features are interpretable as they enter into interpretation, but Case features and the EPP feature are not. ϕ-features are interpretable on NPs but not on Is. Chomsky, then, proposes that uninterpretable features need to be checked and deleted, and further, that Move applies for this purpose.

Given this refinement, the ungrammaticality of (19a) is straightforward. (21) illustrates the feature-checking relation between the NP *John* and the embedded I.

(21)

John	I	be intelligent
D-feature	~~EPP~~	
ϕ-features	~~ϕ-features~~	
~~Case~~	~~Case~~	

Note that the Case feature of the NP is checked and deleted at this point. Hence, when the NP raises to the matrix subject position, it no longer has a Case feature. This means that the Case feature of the matrix I cannot be checked and deleted. The derivation, thus, crashes because an uninterpretable feature remains.

Most examples of Last Resort violations receive a similar analysis, but a notable exception is (22b).

(22) a. There$_i$ is likely [t_i to be a man in the room]
b.*There is likely [a man$_i$ to be t_i in the room]

The example is a Last Resort violation because the NP *a man* has no need to move to the embedded subject position. Yet, the analysis for (19a) is inapplicable in this case: the NP is not checked for Case at the initial position, and further, the embedded I does not have an uninterpretable Case feature to begin with. Chomsky notes that when the derivation reaches the embedded subject position, there is a choice between the insertion of *there* and the movement of *a man*. The former option is legitimate as the grammatical (22a) shows, but the latter is not. Based on this observation, he proposes (23), which is assumed to follow ultimately from derivational economy.

(23) Merge is prefered over Move.

Note that although it is no longer necessary to stipulate that movement applies only when it is required by the moved item, the fact remains that movement is not totally free but applies for the purpose of feature-checking.[6] Thus, Chomsky retains the Last Resort Principle in the weaker form in (24) at this point.

(24) Last Resort: Move F raises F to target K only if F enters into a checking relation with a sublabel of K.

I will return to (24) directly.

---

6 A qualification is necessary here if Japanese/Korean scrambling is not feature-driven. For relevant discussion, see Saito and Fukui 1998 and the references cited there. I will briefly consider examples of scrambling in the following section.

Chomsky's (1995) elimination of the MLC is more involved. As it is clearly a condition on movement, the only way to eliminate it as an independent condition seems to be to incorporate it into the definition of Move. With this reasoning, he first restates it in a form that can be added to the definition of Move as in (25).[7]

(25) α can raise to target K only if there is no legitimate operation Move β targeting K, where β is closer to K.

The basic idea is that in (19b), repeated below as (26), *John* cannot raise to the matrix subject position because the movement of the closer NP *it* is legitimate in the sense that it satisfies (24).

(26) *John$_i$ seems [(that) it is likely [t$_i$ to win the race]]

Next, he notes that the target of movement always contains an uninterpretable feature, e.g., the EPP feature in the case of NP-movement and the strong [+wh] feature in the case of Wh-movement. Finally, based on this observation and (25), he proposes, following a suggestion by John Frampton, that movement is not triggered by a feature of the moved item seeking to enter into a checking relation, but by an uninterpretable feature of the target that looks for and "pulls up" a feature that can check and delete it. This is the reformulation of Move as Attract with the following defining property:

(27) Attract F: K *attracts* F if F is the closest feature that can enter into a checking relation with a sublabel of K.

(27) subsumes both (24) and (25). (24) follows because a target can attract a feature only if the feature can enter into a checking relation with a feature of the target. The MLC effects also follow from (27). For the matrix I in (26), the D-feature of *John* is not the closest feature that can check its EPP feature because of the presence of *it*. The matrix I, then, can attract the D-feature of *it* but not that of *John*.

---

7 The MLC and (25) may make different predictions for examples such as (i).

(i) John$_i$ strikes Mary [t$_i$ as being a fool]

The MLC may allow it (depending on the structure), but (25) seems to exclude it. This may not be a problem for (25) because it is not clear that examples of this kind involve raising as standardly assumed. For example, an expletive cannot appear in the matrix subject position as (ii) shows.

(ii) *There$_i$ strikes Mary [t$_i$ as being a man in the room]

(27) seems to have an empirical advantage as well over the Last Resort Principle and the MLC. Recall that these conditions left (7a) unexplained. The example is repeated below in (28).

(28) *John$_i$ seems [$_{CP}$ (that) [$_{IP}$[$_{NP}$ the proof [$_{IP}$ t$_i$ to be guilty]] was given]]

When the derivation reaches the matrix I, the D-feature of the NP *the proof John to be guilty* is plausibly closer to it than the D-feature of the NP *John*. If so, the matrix I can only attract the D-feature of the larger NP. Then, (27) extends to this example as well.[8]

As illustrated above, the analysis of NP-movement has shown remarkable progress over the past twenty years. First, the overlapping constraints were reduced to two major conditions, the Last Resort Principle and the MLC. Then, these two conditions were eliminated in favor of (27). The motives behind this progress were often conceptual. Attempts were made to eliminiate the redundancy among the conditions, and to eliminate conditions on movement that do not fall under derivational economy.[9] But the purpose of these attempts was to achieve a more abstract, more principled explanation for the core data. In fact, many other phenomena played important roles in this process, e.g., object shift, the so called A-scrambling, and the interaction of head-movement and NP-movement. Although I did not discuss them here, I hope the empirical nature of the project was demonstrated to a sufficient degree by the brief history of the analysis of the classical cases of illicit NP-movement given above.

3. Cyclic Interpretation of Chains and Improper Movement

In this section, I will discuss a case in which the theoretical developments in the Minimalist program may provide a solution to an old, outstanding problem. The phenomenon that I consider is improper movement, which is to my knowledge yet to receive a satisfactory account. Some specific examples are shown in (29)-(30).

(29) a.*John$_i$ seems [$_{CP}$ t$_i$' [$_{IP}$ it is likely [$_{IP}$ t$_i$ to win the race]]]
b.*[John$_i$ to seem [$_{CP}$ t$_i$' [$_{IP}$ t$_i$ is intelligent]]] would be surprising

---

8 As Naoki Fukui points out, a minor adjustment in the definition of 'closeness' may be necessary to accommodate (28) under (27), since Chomsky (1995) defines it in terms of 'c-command'. The required adjustment should be straightforward.

9 Note that the effects of other "independent economy conditions," such as 'the fewest steps', also follow from (27). Another condition of a similar kind, 'procrastinate', is eliminated in Chomsky 1998 with the proposal of 'Agree'.

(30) *John$_i$ is known [how likely [t_i to win the race]]$_j$ it is t_j

(29a) is an example of super-raising, except that the movement takes place through the embedded CP Spec. (29b), which is adapted from Chomsky 1998, is a similar example but in this case the raised NP is checked for Case prior to the movement. (30) from Sakai 1994 involves raising out of a Wh-phrase in the embedded CP Spec. I will suggest an account for these cases based on Chomsky's (1998) theory of phase and cyclic interpretation.

In the following subsection, I will briefly go over the basic motivations for derivational phase as discussed in Chomsky 1998. Then, in Section 3.2, I will suggest a specific mechanism for the cyclic interpretation of chains. Independent evidence for this mechanism is introduced and discussed in Section 3.3. Finally, I will apply the mechanism to cases of improper movement and discuss the consequences of the analysis in Section 3.4.

3.1. Chomsky's (1998) Derivational Phase

Let us first consider again the contrast in (22), repeated below in (31).

(31) a. There$_i$ is likely [t_i to be a man in the room]
b.*There is likely [a man$_i$ to be t_i in the room]

As noted above, Chomsky (1995) proposes to explain (31b) by applying (32) at the point the derivation reaches the embedded subject position.

(32) Merge is prefered over Move.

But (32) faces a problem with examples such as those in (33).

(33) a. There is a possibility [that a proof$_i$ will be discovered t_i]
b. It's fun [PRO$_i$ to [t_i discover a proof]]

There can be merged at the embedded subject position in (33a), and given (32), this should block the movement of *a proof*. (33b), which is attributed to Alec Marantz, is problematic in the same way. Hence, it is necessary to make (32) inapplicable to cases like (33).

Chomsky (1998) first notes that there is one crucial difference between (31) and (33): the embedded clause is an IP in the former but it is a CP in the latter. Then, he proposes that derivation proceeds phase by phase, where CP but not IP constitutes a phase. If each phase has its own numeration, (33a), for example, is correctly allowed. *There* does not belong to the numeration of the embedded CP phase, and hence, is not available when the embedded subject position is filled.

The only way to check the EPP feature of the embedded I, then, is to raise the NP *a proof*. The situation is different in the case of (31b). Since there is no embedded CP phase, *there* is available to fill the embedded subject position. Thus, the raising of *a man* is still blocked by (32).

If derivation indeed proceeds phase by phase, each phase is a complete unit and anything within a lower phase is not accessible to items in a higher phase. But Chomsky notes that an exception must be granted. A verb has a selectional relation with the head of its complement CP. Hence, the head of the CP phase must be "visible" to the verb. Chomsky, then, weakens the condition, and proposes that it is the domain of the head of a phase that is inaccessible to items outside the phase. He states this as in (34).

(34) In phase α with head H, the domain of H is not accessible to operations outside α, but only H and its edge.

This opens up a way to recapture successive-cyclic Wh-movement, which had been left unaccounted for. Given the formulation of the CED=Subjacency in (10) and the reformulation of Move as Attract, it is unclear why Wh-movement must proceed through intervening CP Specs as in (35).

(35) [$_{CP}$ who$_i$ [$_{C'}$ do [$_{TP}$ you think [$_{CP}$ t_i' [$_{TP}$ John saw t_i]]]]]

But (34) makes it clear why it is necessary. If the Wh-phrase is in the initial position, its features are invisible to and hence cannot be attracted by the matrix C. It must first move to the edge of the embedded CP phase to undergo movement to the matrix CP Spec. Given the conclusion in Chomsky 1986a that successive-cyclic Wh-movement proceeds through VP-adjunction and substitution into CP Spec, he suggests that not only C but also *v* define derivational phases.[10]

---

[10] Examples such as (i) constitute clear evidence for successive-cyclic movement through CP Spec.

(i) Which picture of himself$_i$ does John$_i$ think Mary liked

Barss (1986) argues that VP-adjunction must be assumed for examples like (ii).

(ii) ?Which picture of himself$_i$ does John$_i$ wonder why Mary liked

Here, the Wh-phrase must move to an intermediate position to make the local binding of *himself* by *John* possible. Since the embedded CP Spec is unavailable, the position adjoined to the matrix VP seems to be the most plausible option. In the present context, this argument is still suggestive but it does not seem decisive. For example, the unexpected weakness of the Wh-island effect on argument Wh-phrases may be

Here, a technical question still remains with respect to the trigger of successive-cyclic Wh-movement. In particular, there must be an uninterpretable feature on the embedded C in (35) that attracts a feature of the Wh-phrase. Chomsky then proposes (36).

(36) The head H of Phase PH may be assigned an EPP- or P-feature.

These features "pull up" Wh-phrases to intermediate positions, and make successive-cyclic movement possible. The EPP feature, when assigned to *v*, is responsible also for object shift.[11]

An important idea that Chomsky entertains with the proposal of derivational phase is cyclic spell-out and cyclic interpretation. As noted above, phases are more or less complete units in the derivation, and the domain of a phase head cannot be affected by operations in higher phases. Hence, it is possible to send information to the interpretive components upon the completion of each phase. For the C-I interface, this results in derivational interpretation, which is similar in many respects to what is proposed in Epstein, et al. 1998. I will outline a specific instantiation of this idea for the interpretatioin of chains in the following subsection.

3.2. A Mechanism for the Cyclic Interpretation of Chains

In Saito 2001, I suggested a mechanism based on Chomsky's (1993) copy and deletion analysis of movement for the cyclic interpretation of chains. The purpose there was to provide a uniform analysis for the main properties of Japanese/Korean scrambling, e.g., its radical reconstruction property, its A/A' properties, and the fact that its trace is strictly constrained by the Proper Binding Condition. In this subsection and the next, I will briefly summarize the proposal.

---

because multiple CP Specs are marginally allowed.

11 It seems to be assumed here that the P-feature and the EPP feature are assigned to C and *v* respectively. But I am not sure if the EPP feature should play any role in the case of Wh-movement. For example, consider the movement of *what* to the matrix *v*P Spec in the following cases:

(i) a. What$_i$ did you persuade John [$_{CP}$PRO to buy t$_i$]
b.?*What$_i$ did you persuade who [$_{CP}$PRO to buy t$_i$]

In both examples, a P-feature on the embedded C can raise *what* to the embedded CP Spec. The fact that (ib) is a superiority violation suggests that the matrix object position is closer to the matrix *v* than the embedded CP Spec. Then, if it is the EPP feature that "pulls up" *what* to the matrix *v*P Spec in (ia), the intervening NP *John* should block the movement. See Section 3.4 below for a related discussion.

Chomsky's copy and deletion analysis of movement provides an elegant way to interpret Wh-chains. According to this analysis, (37) is interpreted as illustrated in (38).

(37) Whose$_i$ did John see t$_i$

(38) a. [$_{CP}$ Who [$_{C'}$ did [$_{IP}$ John see who]]]
b. [$_{CP}$ Who [$_{C'}$ did [$_{IP}$ John see x]]]

The movement initially copies the Wh-phrase at the landing site as in (38a). Then, deletion applies at all positions of the chain. The operator part of the Wh-phrase is retained at the CP Spec, and the other parts are deleted at this position. This deletion is vacuous in the case of (38a). The Wh-phrase *who* in CP Spec is eventually interpreted as '[for which x: x a person]'. At the initial site, the part that corresponds to the operator is turned into a variable.

This interpretive procedure can be readily stated in terms of features. I will adopt a somewhat simplified picture for the sake of exposition, and assume that the Wh-phrase *who* consists of three relevant sets of features, the phonetic features (π-features), the operator feature (O-feature), and the categorial D-feature. The D-feature of an NP, I will assume, enters into binding/coreference relations and makes it possible for the NP to have a reference. Then, the copying in (38a) is more precisely as in (39).

(39) [$_{CP}$ Who [$_{C'}$ did [$_{IP}$ John see who]]]
{π,O,D} {π,O,D}

Now, suppose that deletion applies to the features so that they are retained in only one position. The π-features must be deleted at the initial site and retained at the landing site. I will simply assume this as given because it is almost the definition of overt movement. For the other features, it is clear that the O-feature must be retained at the landing site. The D-feature, on the other hand, should be at the initial site if it is responsible for the interpretation of the copy in this position as a variable. The desired result, then, is as in (40).

(40) [$_{CP}$ Who [$_{C'}$ did [$_{IP}$ John see who]]]
{π,O} {D}

And this result is obtained if the O-feature and the D-feature can only be retained at positions where they are selected (in a broad sense to be specified below). The D-feature is selected by the verb at the initial site, but not by the C at CP Spec. Hence, it must be retained at the object position. The O-feature is clearly not selected at the initial site. But it is at the landing site if we interpret 'selectional

relations' in a broad sense so as to include checking relations. Thus, the selection requirement guarantees that deletion applies properly to the features.[12]

Construed this way, Chomsky's copy and deletion analysis can be applied derivationally. I will illustrate this with the example in (41), paying attention to CP phases.

(41) Who do you think John saw

Since the embedded CP is a phase, the Wh-phrase *who* must move to its edge as in (42a) so that it can eventually move out of the CP.

(42) a. [$_{CP}$ who [$_{IP}$ John saw who]]
{π,O,D} {π,O,D}

b. [$_{CP}$ who [$_{IP}$ John saw who]]
{π,O} {D}

c. [$_{CP}$ who [$_{IP}$ John saw x]]
{π,O}

At this point, deletion applies exactly as in the case of (40), as shown in (42b). The π-features are retained at CP Spec. The O-feature is also retained there because it enters into a checking relation with the P-feature on C that triggers the movement.[13] The D-feature, on the other hand, is deleted at this position since it is selected only at the object position. Note that the Wh-phrase in CP Spec is not ready for interpretation because it is subject to further operations. But the TP complement of C is. The copy of *who* with the sole feature D is interpreted as a variable, and the TP is construed as an open sentence 'John saw x'.

The Wh-phrase in (42c) undergoes further movement in the matrix CP as in (43a).

(43) a. [$_{CP}$ who [$_{C'}$ do [$_{IP}$ you think [$_{CP}$ who [$_{IP}$ John saw x]]]]]
{π,O} {π,O}

12 This requirement is a variant of Lee's (1994) idea that only feature-checking positions are retained in chains at LF.

13 I assume, following Chomsky 1998, that the P-feature is uninterpretable and hence must be deleted after it is checked. The deletion must apply prior to the point when information on the C is sent to the interpretive component.

b. [$_{CP}$ who [$_{C'}$ do [$_{IP}$ you think [$_{CP}$ who [$_{IP}$ John saw x]]]]]
{π,O} {}

c. [$_{CP}$ who [$_{C'}$ do [$_{IP}$ you think [$_{CP}$ [$_{IP}$ John saw x]]]]]
{π,O}

The π-features are retained at the landing site as before, and the O-feature can be as well because it enters into a checking relation with a feature of the matrix C. The result is (43b), where the copy of *who* in the embedded CP Spec disappears as there are no features left in this position. The information on the matrix CP phase can be sent to the interpretive components at this point. In particular, *who* is spelled-out at the matrix CP Spec because its π-features are there, and its O-feature yields its interpretation at this position as '[for which x: x a person]'.

3.3. A Piece of Evidence for the Mechanism

Among the desirable consequences of the chain interpretation mechanism outlined above is that it allows us to provide a uniform analysis for the so called A and A' scramblings.

As noted by Mahajan (1990), a phrase preposed by clause-internal scrambling can serve as the antecedent of a lexical anaphor, but long-distance scrambling lacks this property. The following examples illustrate this generalization:

(44) a.?*[$_{IP}$ [Otagai -no sensei]-ga karera-o hihansita] (koto)
each other-GEN teacher-NOM they -ACC criticized fact

'[Each other's teachers] criticized them'

b. ?[$_{IP}$ Karera-o$_i$ [[otagai -no sensei]-ga t_i hihansita]] (koto)
they -ACC each other-GEN teacher-NOM criticized fact

'Them$_i$, [each other's teachers] criticized t_i'

(45) a. *[$_{IP}$ [Otagai -no sensei]-ga [$_{CP}$ [$_{IP}$ Tanaka-ga karera-o
each other-GEN teacher-NOM -NOM they -ACC

hihansita] to] itta] (koto)
criticized that said fact

'[Each other's teachers] said that Tanaka criticized them'

b. *[$_{IP}$ Karera-o$_i$ [[otagai -no sensei] -ga [$_{CP}$ [$_{IP}$ Tanaka-ga t_i
they -ACC each other-gen teacher-NOM -NOM

hihansita] to] itta]] (koto)
criticized that said fact

'Them$_i$, [each other's teachers] said that Tanaka criticized t_i'

The contrast between (44b) and (45b) suggests that there are two distinct types of scrambling. Mahajan (1990) indeed proposes that there are A- and A'-scramblings with different landing sites, and that the former is basically clause-bound.

However, if chains are interpreted derivationally as suggested above, this bifurcation of scrambling seems unnecessary at least for Japanese and Korean. Let us first consider the movement in (44b) illustrated in (46).

(46) a. [$_{IP}$ karera-o [... otagai-no ... karera-o ...]]
{π,D} {π,D}

b. [$_{IP}$ karera-o [... otagai-no ... karera-o ...]]
{π} {D}

The scrambled NP is copied at the sentence-initial position in (46a).[14] With deletion, the π-features are retained only at the landing site as in the cases of other overt movements. Now, suppose that Japanese/Korean scrambling is not feature-driven as I argued elsewhere.[15] Then, the D-feature is not selected at the landing site and must be deleted there. This yields (46b).

The structure in (46b) appears to be inconsistent with the grammaticality of (44b) given that it is the D-feature of an NP that enters into binding relations. However, the problem is only apparent if Condition (A) is an anywhere condition (i.e., it applies derivationally) as argued on independent grounds by Belletti and Rizzi (1988), Lebeaux (1988) and Epstein, et al. (1998), among others. The D-feature of the scrambled NP is in a position c-commanding the lexical anaphor at

[14] If *v*P is also a phase, the object NP must first move to its edge. I assume that this is a case of "VP-internal scrambling," which is known to have properties similar to object shift unlike the other cases of scrambling. See Tada 1990 and Nemoto 1993 for detailed discussion on this type of local scrambling, and Saito 2001 for a possible analysis that is consistent with the discussion in the text.

[15] See Saito and Fukui 1998, Saito 2000, and the references cited there. This implies that there is no head that attracts a feature of the scrambled phrase, and that scrambling results from an operation similar to pure Merge.

one point of the derivation, as shown in (46a). Thus, the lexical anaphor is successfully licensed and interpreted.

(45b) is derived in a similar way, but in this case, the scrambled NP must first move to the edge of the embedded CP because it is a phase. This initial movement is illustrated in (47a).

(47) a. [$_{CP}$ karera-o [... karera-o ...]]
{π,D} {π,D}

b. [$_{CP}$ karera-o [... karera-o ...]]
{π} {D}

At this point, deletion yields (47b) in the same way that it created (46b). Then, the head of the chain in (47b) undergoes further movement into the matrix clause as in (48a).

(48) a. [$_{IP}$ karera-o [... otagai-no ... [$_{CP}$ karera-o [... karera-o ...]]]]
{π} {π} {D}

b. [$_{IP}$ karera-o [... otagai-no ... [$_{CP}$ karera-o [... karera-o ...]]]]
{π} {} {D}

c. [$_{IP}$ karera-o [... otagai-no ... [$_{CP}$ [... karera-o ...]]]]
{π} {D}

As before, the π-features are retained only at the landing site. The copy in the embedded CP Spec, then, disappears since all if its features are deleted, yielding (48c).

Although the resulting structure in (48c) is virtually identical to the case of clause-internal scrambling shown in (46b), there is one crucial difference between the derivations in (46) and (48). In the case of (48), the D-feature of the scrambled NP c-commands the lexical anaphor at no point of the derivation. Thus, the ungrammaticality of (45b) follows. According to this analysis, scrambling is strictly uniform whether it is clause-internal or long-distance. It involves copying and only the π-features of the moved phrase are retained at the landing site. The hypothesis that all features except the π-features are deleted at the landing site is in accord with the proposal in Saito 1989 that scrambling can be "undone" in LF. It also captures the insight of the hypothesis often entertained since Chomsky and Lasnik 1977 that scrambling is PF movement. It is not, but it is indistinguishable from PF movement in terms of the structures it creates, as can be observed in (46b) and (48c).

3.4. An Account for the Improper Movement Phenomenon

Let us now apply the chain interpretation mechanism to the cases of improper movement mentioned at the outset of this section. The most straightforward example is Sakai's (1994) (30), repeated below as (49).

(49) *John$_i$ is known [how likely [$\underline{t}_i$ to win the race]]$_j$ it is $\underline{t}_j$

Here, the Wh-phrase *how likely John to win the race* is first copied at the embedded CP Spec as in (50a).

(50) a. [$_{CP}$ [how likely [John to ...]] [$_{IP}$ it is [how likely John to ...]]]
{π,O,A} {π,O,A}

b. [$_{CP}$ [how likely [John to ...]] [$_{IP}$ it is [how likely John to ...]]]
{π,O} {A}

c. [$_{CP}$ [how likely [John to ...]] [$_{IP}$ it is [x much likely <u>John</u> to ...]]]
{π,O} {D-Case}

The O-feature and the π-features of the Wh-phrase are retained at the CP Spec. On the other hand, the categorial feature A of the phrase is "reconstructed" to the initial site, along with those of the terms it dominates.[16] We thus obtain (50b).

At this point, information on the IP complement is sent to the interpretive component as in (50c). But there is an obvious problem. Suppose, as it seems reasonable, that the Case feature and the ϕ-features of an NP are associated with and pattern along with its D-feature. Then, the Case feature of *John* is on the copy within the IP complement, and is sent for interpretation without being checked and deleted. The interpretation should fail receiving an uninterpretable feature. Further, an additional problem arises even if the derivation is allowed to proceed. Since the categorial features are deleted at the landing site, *John* within the CP Spec does not have a D-feature that can be attracted by the matrix I. Thus, (49) fails to be generated in any event.

The example in (29a), repeated in (51), can be analyzed in basically the same way.

(51) *John$_i$ seems [$_{CP}$ $\underline{t}_i$' [$_{IP}$ it is likely [$_{IP}$ $\underline{t}_i$ to win the race]]]

16 Since the moved phrase is complex, I assume that the internal structure of the phrase is preserved at both positions of the chain. At the landing site, each lexical item contains its π-features. At the initial site, we have a phrase of categorial features without the O-feature and the π-features.

Suppose that a P-feature triggers the initial movement to CP Spec. Then, the movement is not possible for *John* because the NP lacks an O-feature. But this problem can be easily circumvented by substituting a Wh-phrase for the NP as in (52), which is still ungrammatical.

(52) *Who$_i$ seems [$_{CP}$ $\underline{t}_i$' [$_{IP}$ it is likely [$_{IP}$ $\underline{t}_i$ to win the race]]]

In this case, the initial movement creates the structure in (53a).

(53) a. [$_{CP}$ who [$_{IP}$ it is likely [$_{IP}$ who to win the race]]]
{π,O,D} {π,O,D}

b. [$_{CP}$ who [$_{IP}$ it is likely [$_{IP}$ who to win the race]]]
{π,O} {D}

c. [$_{CP}$ who [$_{IP}$ it is likely [$_{IP}$ <u>x</u> to win the race]]]
{π,O} {D-Case}

Then, the D-feature is deleted at CP Spec, and the π-features and the O-feature are deleted at the tail of the chain, as in the case of regular successive-cyclic Wh-movement illustrated in (42). Thus, we obtain (53b). The IP complement of C is sent for interpretation at this point, but the copy of *who* at the initial site cannot be interpreted as a variable because of the unchecked Case feature. Further, the Wh-phrase in CP Spec cannot move on to the matrix IP Spec as in (54), since it does not contain a D-feature that can be attracted by the matrix I.

(54) [$_{IP}$ who seems [$_{CP}$ who [$_{IP}$ it is likely [$_{IP}$ <u>x</u> to win the race]]]
{π,O} {π,O} {D-Case}

Hence, the uninterpretable features of the matrix I remain unchecked.

The final case (29b), adapted from Chomsky 1998, is repeated below in (55).

(55) *[John$_i$ to seem [$_{CP}$ $\underline{t}_i$' [$_{IP}$ $\underline{t}_i$ is intelligent]]] would be surprising

This example has interesting additional consequences. Suppose again that it is a P-feature assigned on the C-head that triggers the movement of the NP to the CP Spec. Then, in this case also, the example has to be changed slightly as in (56) to make this movement possible.

(56) *Who said that [who$_i$ to seem [$_{CP}$ $\underline{t}_i$' [$_{IP}$ $\underline{t}_i$ is intelligent]]] would be surprising

The initial movement in this example is illustrated in (57).

(57) a. [$_{CP}$ who [$_{IP}$ who is intelligent]]
{π,O,D} {π,O,D}

b. [$_{CP}$ who [$_{IP}$ who is intelligent]]
{π,O} {D}

c. [$_{CP}$ who [$_{IP}$ x is intelligent]]
{π,O} {D}

The structure is derived and interpreted exactly as in the case of (53), but there is one crucial difference. Since *who* at the tail of the chain in (57b) is checked for Case, it can successfully be interpreted as a variable. Hence, unlike in the cases of (49) and (52), the only place where the derivation fails in (56) is the second step of the movement. Here, the EPP feature of the non-finite I cannot trigger the movement of the Wh-phrase to the IP Spec as in (58) because the D-feature of the Wh-phrase is already deleted.

(58) [$_{IP}$ who to seem [$_{CP}$ who [$_{IP}$ x is intelligent]]
{π,O} {π,O} {D}

This suggests that although the failure of the second step of the movement is redundant in the accounts for (49) and (52), it is nevertheless a real phenomenon.

Another implication of the proposed account for (55)/(56) is that a C-head can be assigned a P-feature but not an EPP feature, as I have been assuming. Whether an EPP feature can be assigned to a C-head to trigger successive-cyclic movement is immaterial in cases like (51)/(52). In (51), for example, even if an EPP feature is assigned to the embedded C, it fails to attract the D-feature of the NP *John* because of the intervening expletive *it*. But it matters in (55)/(56). Suppose, for example, that an EPP feature is assigned to C and makes it attract the D-feature of *John* in (55). Then, by hypothesis, the D-feature can be retained at the CP Spec because it enters into a feature-checking relation there. In the next phase up, the non-finite I should be able to attract this feature, and there should be nothing wrong with the example. Hence, the assignment of an EPP feature to C should be banned altogether.[17]

---

17 As noted in Footnote 11, it is not clear that the EPP feature plays any role in successive-cyclic Wh-movement. If the speculation there is correct, and phases are defined by *v* and C as Chomsky proposes, then it is the P-feature assigned to *v* and C that makes successive-cyclic Wh-movement possible. Further, there seem to be no cases where an EPP feature assigned to *v* (or *v** in the sense of Chomsky 1999) plays a role in successive-cyclic NP-movement. If *v* (or *v**) implies the presence of an external argument, NP-movement out of a *v*P (or *v**P) phase is ruled out on independent grounds. Then, object shift seems to be the only phenomenon where

The final implication has to do with the 'activation condition' proposed in Chomsky 1998. Considering examples such as (55), he suggests that "an uninterpretable feature activates the goal." This means in the present context that only NPs with uninterpretable features are subject to movement. Thus, *John* in (55) cannot move since its uninterpretable Case feature is checked and deleted at the initial site.[18] But given the account for (55)/(56) proposed above, the motivation for the activation condition is weakened. The theory of phase implies that the movement in these examples must proceed through the edge of the CP phase. And once this happens, we have the case of improper movement discussed above. Hence, if the analysis above is on the right track, the activation condition is redundant for examples of this kind.

The other place where Chomsky (1998) appeals to the activation condition is in his account for the illicit partial Wh-movement as in (59).

(59) *Who thinks [$_{CP}$ what$_i$ [$_{IP}$ John bought t_i]]

Given the activation condition, a Wh-phrase must have an uninterpretable feature to undergo movement to CP Spec in general. This feature is checked by a [+wh] C and deleted at the CP Spec where the Wh-phrase takes scope. But a P-feature assigned to an intermediate C should not check and delete this feature: if it does, the activation condition prevents the second step of the successive-cyclic movement in (60), for example.

(60) What$_i$ does Mary think [$_{CP}$ t_i' [$_{IP}$ John bought t_i]]

Then, in (59), *what* has the uninterpretable feature because it moves to the embedded CP Spec. But this feature cannot be checked there, and hence, remains unchecked. The ungrammaticality of (59) thus follows.

Although this account for (59) is elegant, the application of the activation condition to Wh-movement seems to have an unwanted consequence: it implies that there are two kinds of Wh-phrases, one with and the other without the uninterpretable feature. This is so because in examples of multiple-Wh questions like (61), the moved Wh-phrase must have the feature but the Wh-phrase in situ cannot have it.

---

the EPP feature on *v* is needed.

18 The activation condition can readily be stated as part of the definition of Attract/Move (or Agree in the theory of Chomsky 1998, 1999), and hence, does not have the conceptual problem that the classical Last Resort Principle did. But it surely looks like a residue of this principle.

(61) What$_i$ did John give t_i to whom

This is quite different from the case of NP-movement, where every NP (DP) is assumed to enter the derivation with an uninterpretable Case feature.

Here, Tsai's (1994) analysis of Wh-in-situ suggests an alternative account for (59). He proposes that only Wh-phrases with D-feature can be interpreted through unselective binding (or absorption in the sense of Chomsky 1993). This proposal takes care of Huang's (1982) classical example in (62).

(62) *Who$_i$ t_i left early why

Now, if unselective binding applies to the head of a chain, the ungrammaticality of (59) follows straightforwardly from Tsai's proposal. After deletion, the Wh-chain in this example will be as in (63).

(63) $[_{CP}$ what $[_{IP}$ John bought what]]
{π,O} {D}

Unselective binding fails here because the head of the chain lacks a D-feature. This forces the interpretation of the chain as an operator-variable chain, but this should also fail because the O-feature cannot be interpreted at the CP Spec position of a proposition.

Another, probably more principled, alternative possibility is to rule (63) out by appealing to Full Interpretation of positions, which can be formulated as in (64).

(64) Every position (term) must contribute to interpretation.

Suppose, as seems reasonable, that information on every position that contains formal or semantic features is eventually sent to the C-I interface. This applies to the CP Spec in (63), because it contains the O-feature. But none of its formal or semantic features is interpreted in that position. Hence, the position is redundant at the C-I interface in violation of (64).[19] If these alternative accounts for (59) are tenable, then it may be possible to eliminate the activation condition altogether.

4. Summary and Further Issues

In this paper, I first reviewed the developments in the analysis of illicit NP-movement. The purpose of this discussion was to show that the Minimalist program is a direct descendant of the "earlier theories," and is pursued in much the same way. In the second part, I suggested a mechanism of cyclic chain interpretation based on the theory of phase, and argued that it solves an old outstanding problem concerning improper movement. If the proposed analysis can be maintained, it shows that the theoretical proposals in the Minimalist program has a further empirical consequence. If not, the discussion illustrates an empirical domain in which further research in the Minimalist program is called for. There are in fact many more empirical problems to be resolved as we can see in the vast Minimalist literature. I will briefly discuss one of them before I conclude this paper.

It is probably fair to say that the Barriers theory (Chomsky 1986a) was developed on the basis of detailed examination of Wh-movement. It achieved much success in this domain, but left many problems unsolved for NP-movement. On the other hand, as noted at the outset of this paper, the analysis of NP-movement played a central role in the development of the Minimalist theory. This means that much further research is required on Wh-movement.

Let us consider the superiority phenomenon, which has received some attention in the Minimalist program. (See, for example, Boskovic 1997 and the references cited there.) It may seem that the following contrast follows directly from the definition of Attract/Move:

(65) a. Who$_i$ t_i bought what
b.*What$_i$ did who buy t_i

The elimination of the activation condition suggested above may in fact be helpful here. If only Wh-phrases with an uninterpretable feature can move, and if a Wh-phrase can but need not have this feature, (65b) cannot be straightforwardly ruled out. More concretely, if *what* has this ferature and *who* does not, it is not clear that the movement of *what* should be blocked by the closer Wh-phrase *who*. But once the activation condition is eliminated, there is no need to assume that those Wh-phrases that move to CP Spec enter the derivation with uninterpretable features. It can be assumed instead that every Wh-phrase has an O-feature, and the C head attracts this feature. Then, the C in (65) must attract the O-feature of *who* since it is clearly closer than that of *what*.

However, there are problematic examples, including (66), discussed in Lasnik and Saito 1992.

(66) ?Who$_i$ t_i knows what$_j$ who$_k$ bought t_j

This example is a superiority violation under the interpretation where *who*$_k$ takes

---

[19] Note that (64) may subsume Chomsky's (1999) proposal in (i).

(i) *v** is assigned an EPP-feature only if that has an effect on the outcome.

embedded scope. But it is not under the reading where the Wh-phrase takes matrix scope. This indicates that superiority is relative to the scope position. If who_k takes embedded scope, then it competes with *what* for the embedded CP Spec. Since who_k is closer, the movement of *what* is excluded. On the other hand, if who_k takes matrix scope, this Wh-phrase and the matrix who_i compete for the matrix CP Spec. Since the closer who_i is actually moved, the example is fine under this interpretation.

The relevance of interpretation to movement observed in (66) is not easy to capture under the definition of Attract/Move in (27). If every Wh-phrase carries the O-feature, and the embedded C in (66) attracts this feature, then the example should be a superiority violation under any interpretation. What seems to be needed here is some sort of binding relation between C and Wh-phrases that is established prior to movement. For example, let us say that the embedded C binds the O-features of all and only those Wh-phrases that take embedded scope, and only those O-features bound by the C can be attracted by it. This will give us the desired result in (66) because the O-feature of *what* will be the only possible attractee for the embedded C when who_k takes matrix scope. But examples like (67) pose further complication.

(67) ??Who$_i$ t_i knows what$_j$ John thinks that who$_k$ bought t_j

As Roger Martin (personal communication) reports, this example is worse than (66) but has a similar interpretive property. The matrix interpretation is marginal and the embedded interpretation is out for who_k. The contrast in this case implies that the most deeply embedded C with *that* is allowed to attract the O-feature of *what* when who$_k$ takes matrix scope. Then, if the binding mechanism suggested above is extended to this case, a [-wh] C with a P-feature must be allowed to bind an O-feature of a Wh-phrase even though it has nothing to do with its scope. At this point, the proper analysis of the superiority phenomenon is far from clear.

This short discussion on superiority indicates that there are many empirical issues on Wh-movement that need to be addressed and resolved in the Minimalist program. The development of generative grammar has allowed us to gain insights into a wide range of facts in various languages. On the other hand, as Chomsky repeatedly points out, the task to explain these facts becomes more difficult and challenging with a restrictive theory. It is not surprising that there are more empirical issues now than ever before.

References

Aoun, J. (1979) "On government, Case-marking, and clitic placement," ms., MIT.
Aoun, J. (1981) *The formal nature of anaphoric relations*, Ph.D. dissertation, MIT.
Barss, A. (1986) *Chains and anaphoric dependence*, Ph.D. dissertation, MIT.

Belletti, A. and L. Rizzi (1988) "Psych-verbs and theta theory," *Natural Language & Linguistic Theory* 6, 291-352.
Boskovic, Z. (1995) *Principles of economy in nonfinite complementation*, Ph.D. dissertation, University of Connecticut.
Boskovic, Z. (1997) "On certain violations of the superiority condition, AgrO, and economy of derivation," *Journal of Linguistics* 33, 227-254.
Cattell, R. (1976) "Constraints on movement rules," *Language* 52, 18-50.
Chomsky, N. (1981) *Lectures on government and binding*, Foris Publications, Dordrecht.
Chomsky, N. (1986b) *Barriers*, MIT Press, Cambridge, Mass.
Chomsky, N. (1986a) *Knowledge of language*, Praeger, New York.
Chomsky, N. (1993) "A minimalist program for linguistic theory," in K. Hale and S.J. Keyser, eds., *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, MIT Press, Cambridge, Mass.
Chomsky, N. (1995) "Categories and transformations," in N. Chomsky, *The minimalist program*, MIT Press, Cambridge, Mass.
Chomsky, N. (1998) "Minimalist inquiries: The framework," MIT Occasional Papers in Linguistics #15.
Chomsky, N. (1999) "Derivation by phase," MIT Occasional Papers in Linguistics #18.
Chomsky, N. and H. Lasnik (1977) "Filters and control," *Linguistic Inquiry* 8, 425-504.
Chomsky, N. and H. Lasnik (1993) "The theory of principles and parameters," in J. Jacobs, et al., eds., *Syntax: An international handbook of contemporary research*, Walter de Gruyter, Berlin.
Epstein, S., E. Groat, R. Kawashima and H. Kitahara (1998) *A derivational approach to syntactic relations*, Oxford University Press, Oxford.
Huang, C.-T. J. (1982) *Logical relations in Chinese and the theory of grammar*, Ph.D. dissertation, MIT.
Kayne, R. (1981) "ECP extensions," *Linguistic Inquiry* 12, 93-133.
Kuno, M. (2001) "Why does movement sometimes leave a trace and sometimes not?", ms., University of Tokyo.
Lasnik, H. (1985) "Illicit NP movement: Locality condition on chains?", *Linguistic Inquiry* 16, 481-490.
Lasnik, H. and M. Saito (1984) "On the nature of proper government," *Linguistic Inquiry* 15, 235-289.
Lasnik, H. and M. Saito (1992) *Move α*, MIT Press, Cambridge, Mass.
Lebeaux, D. (1988) *Language acquisition and the form of grammar*, Ph.D. dissertation, University of Massachusetts, Amherst.
Lee, R. K. (1994) *Economy of representation*, Ph.D. dissertation, University of Connecticut.
Mahajan, A. (1990) *The A/A-bar distinction and movement theory*," Ph.D. dissertation, MIT.

Martin, R. (1992) "On the distribution and Case features of PRO," ms., University of Connecticut.
Nemoto, N. (1993) *Chain and Case positions: A study from scrambling in Japanese*, Ph.D. dissertation, University of Connecticut.
Rizzi, L. (1990) *Relativized minimality*, MIT Press, Cambridge, Mass.
Saito, M. (1989) "Scrambling as semantically vacuous A'-movement," in M. Baltin and A. Kroch, eds., *Alternative Conceptions of Phrase Structure*, University of Chicago Press, Chicago.
Saito, M. (1996) "Barriers for A-adjunction," *Japanese/Korean Linguistics* 5, 1-36.
Saito, M. (2000) "Japanese scrambling in a comparative perspective," ms., Nanzan University.
Saito, M. (2001) "A derivational approach to the interpretation of scrambling chains," ms., Nanzan University.
Saito, M. and N. Fukui (1998) "Order in phrase structure and movement," *Linguistic Inquiry* 29, 439-474.
Sakai, H. (1994) "Derivational economy in long distance scrambling," *MIT Working Papers in Linguistics* 24, 295-314.
Stowell, T. (1982) "The tense of infinitives," *Linguistic Inquiry* 13, 561-570.
Tada, H. (1990) "Scrambling(s)," ms., MIT.
Takahashi, D. (1994) *Minimality of movement*, Ph.D. dissertation, University of Connecticut.
Tsai, W.-T. D. (1994) *On economizing the theory of A-bar dependencies*, Ph.D. dissertation, MIT.