Move 1st: A Dynamic Economy Plan*

Takashi Toyoshima

Kyushu Institute of Technology/Cornell University

0. Introduction

In Chomsky (1995), the operation Merge is considered "costless," whereas Move/Attract is not, so that Merge always has priority over Move/Attract, whenever there is a choice — global static economy. On the other hand, Collins (1997) proposes a framework of local economy, in which Merge and Move/Attract are not comparable, so that neither has priority over the other; in effect, they have equal "cost." Thus, Merge and Move/Attract are always options whenever there is a choice, leading to different outputs — non-deterministic local economy.

In this paper, I propose a deterministic economy principle that makes a dynamically local choice of operations, generally favoring Move/Attract over Merge, yet sometimes choosing Merge over Move/Attract. I call this the Principle of Minimum Feature Retention (MFR), which chooses the operation that leaves the fewest features in the structure produced by that operation. This principle is empirically motivated by an over-generation problem that arises in the account of super-raising offered in Chomsky (1995) — a classic case of a Tensed-S Condition violation (Chomsky 1973), which has not yet received any satisfactory account in the minimalist literature. Conceptually, the MFR is motivated from the viewpoint of computational complexity theory, in which it is known that the space requirement is more crucial than time measurement (Johnson 1990, Papadimitriou 1994, among others). For the computational system of human language, the memory load required to keep track of unchecked features is more important than the number of steps to yield a convergent derivation. The MFR offers a local solution to minimize the number of steps, by reducing the memory load, thus having a propensity of synergistic effect for the reduction of the overall computational complexity.

This paper is organized as follows: in section 1, we first see the over-generation problem of the super-raising account in Chomsky's (1995) global static economy as well as in Collins' (1997) non-deterministic local economy, reviewing their accounts of the *there*-expletive constructions, which is the principal empirical motivation for holding the view that Merge is equally or more economical than Move/Attract. In section 2, we see

* The material of this article is excerpted from Chapter I and Chapter III of Toyoshima (forthcoming). The usual disclaimer applies, and any misunderstandings or misinterpretations are of my sole responsibility.

Pius Tamanji, Masako Hirotani, and Nancy Hall (eds.), NELS 29: 409-425

that Move/Attract needs to take priority over Merge sometimes, but not always, demonstrating that the over-generation problem of the super-raising account is a violation of the Tensed-S Condition (Chomsky 1973). I will formulate the MFR in section 3, explicating the over-generation problem of the super-raising account. In section 4, I will offer an account of the *there*-expletive constructions, which appear to run counter to the MFR at first sight, in terms of selection by an unaccusative verb. We review in section 5 the erudition on computational complexity, outlining the theory of computational complexity, studied in combinatorial mathematics and theoretical computer science. Finally, section 6 discusses some implications of the MFR, and concludes the paper.

1. The Problem of Over-Generation
1.1 Expenses of Costless Merge

Chomsky (1995) holds the view that Merge is "costless," in that its insufficient application fails to yield a convergent derivation, and hence no question arises about the motivation of its application or about economy. On the other hand, Move/Attract is a last resort operation triggered by the need of feature-checking, so that it is not "free." Furthermore, an application of Move/Attract in the covert component is considered "cheaper" than in the overt component, on the assumption that covert operations are "wired-in" reflexes of the computational system (Procrastinate). In consequence, whenever there is a choice between Merge and Move/Attract, Merge always has priority over Move/Attract. That is, the economy evaluation is static, and requires global comparison among convergent derivations.

Apart from the problem of computational complexity it entails, Chomsky's global static economy has a serious empirical problem within the mechanisms provided in Chomsky (1995): the problem lies in the account of super-raising (*op. cit.*:295ff.).

(1) *John$_i$ seems (that) it was told t_i that IP

Incorporating the Minimal Link Condition (or the "Shortest Movement" Condition) as a defining property of Move/Attract, Chomsky claims that (1) cannot be generated with available operations. (1) was presumably derived from the following stage in the derivation:

(2) I^0 seems (that) [$_{IP}$ it was told John [$_{CP}$ that IP]]

In this structure, *it* is closer than *John* is to the matrix I^0, so that the matrix I^0 attracts *it*, not *John*, yielding the following:

(3) *It$_i$ I^0 seems (that) [$_{IP}$ t_i was told John [$_{CP}$ that IP]]

Chomsky claims that the matrix I^0 cannot check its Case-feature in (3), since the Case-feature of *it* has been checked and erased when *it* is merged to the embedded Spec(IP) — hence the derivation crashes. Thus, the derivation leading to (3) is discarded, and the computational system must seek an alternative derivation that converges. The decisive point in the derivation is the following stage:

(4) [$_{I'}$ was told John [$_{CP}$ that IP]]

As considerations of economy arise only if there is a convergent continuation (Chomsky 1995:297), merging *it* is not a choice at this stage, and the only option is to raise *John*, and the derivation can continue as follows:

(5) a. [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]
  b. [$_{CP}$ (that) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]
  c. [$_{VP}$ seems [$_{CP}$ (that) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]]
  d. [$_{I'}$ I^0 [$_{VP}$ seems [$_{CP}$ (that) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]]]
  e. [$_{IP}$ it [$_{I'}$ I^0 [$_{VP}$ seems [$_{CP}$ (that) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]]]]

However, there is a serious oversight in this explanation; covert raising of *John*'s Case-feature in (3) should be able to check the Case-feature of the matrix I^0, as *John*'s Case-feature has not been checked, so that (3) should converge, contrary to Chomsky's claim. As (3) converges, raising of *John* and merger of *it* do compete for the purpose of economy at stage (4), and insofar as Merge is more economical than Move/Attract, merger of *it* must be chosen, leading to (2, 3). Thus, even though super-raising (1) is correctly ruled out, another ungrammatical derivation (3) is over-generated as convergent and more economical in Chomsky's static preference of Merge over Move/Attract: merger of *it* is more economical than raising of *John* at stage (4). This state of affairs seriously undermines the correctness of Chomsky's (1995) account of super-raising, threatening the whole approach to Relativized Minimality effects in terms of the shortest movement.

The primary empirical motivation for taking Merge as "costless" or at least more economical than Move/Attract comes from the account of the *there*-expletive, in the context of the following kind of paradigm (Chomsky 1995:344ff.):

(6) a. There seems to be a man in the room.
  b. *There seems a man to be in the room.
(7) a. *Someone believes to be a man in the room.
  b. Someone believes a man to be in the room.

The paradox here is that *a man* cannot raise in (6), but has to raise in (7). These presumably had the common intermediate stage as in the following:

(8) [$_{I'}$ to be [a man in the room]]

Given a numeration containing *there* for (6), the option at stage (8) is to either merge *there* or raise *a man*, in order to check the strong EPP-feature of I^0. Raising of *a man* is assumed to be more costly than merging of *there*, so that the latter option is chosen.

(9) [$_{IP}$ there [$_{I'}$ to be [a man in the room]]]

After this stage, this structure is embedded under *seems*, and the matrix I^0 will be further merged as below:

(10) a. [$_{VP}$ seems [$_{IP}$ there [$_{I'}$ to be [a man in the room]]]]
  b. [$_{I'}$ I^0 [$_{VP}$ seems [$_{IP}$ there [$_{I'}$ to be [a man in the room]]]]]

At this point, the strong EPP-feature of the matrix I^0 needs to be checked. On the assumption that the numeration contains no more lexical items bearing a D-feature, the choice is to raise *there*, since *there* is closer than *a man* is to the matrix I^0, yielding (11).

(11) $[_{IP}$ there$_i$ $[_{I'}$ I^0 $[_{VP}$ seems $[t_i$ [to be [a man in the room]]]]]] (= 6a)

After this stage, the Case-feature and the ϕ-features of *a man* will raise covertly to check the Case-feature and the ϕ-features of the matrix I^0, and the derivation converges.

For (7), however, whose numeration is assumed not to contain *there,* but another D category *someone*, the same consideration at stage (8) will choose the merging of *someone* over the raising of *a man*, yielding the following:

(12) $[_{IP}$ someone $[_{I'}$ to be [a man in the room]]]

This structure will be further embedded under *believes,* merging the matrix I^0, and *someone* will be raised to the matrix Spec(IP), as below:

(13) a. $[_{VP}$ believes $[_{IP}$ someone $[_{I'}$ to be [a man in the room]]]]
b. $[_{I'}$ I^0 $[_{VP}$ believes $[_{IP}$ someone $[_{I'}$ to be [a man in the room]]]]]
c. $[_{IP}$ someone$_i$ $[_{I'}$ I^0 $[_{VP}$ believes $[_{IP}$ t_i $[_{I'}$ to be [a man in the room]]]]]] (= 7a)

After this stage, the Case-feature of *a man* raises covertly to be checked against the Case-feature of *believes,* and the derivation should converge.

Consider now the following continuation for (7b) from the stage (8).

(14) a. $[_{IP}$ a man$_i$ $[_{I'}$ to be $[t_i$ in the room]]]
b. $[_{V'}$ believes [a man$_i$ [to be $[t_i$ in the room]]]]
c. $[_{VP}$ someone $[_{V'}$ believes [a man$_i$ [to be $[t_i$ in the room]]]]]
d. $[_{I'}$ I^0 $[_{VP}$ someone $[_{V'}$ believes [a man$_i$ [to be $[t_i$ in the room]]]]]]
e. $[_{IP}$ someone$_j$ $[_{I'}$ I^0 $[_{VP}$ t_j $[_{V'}$ believes [a man$_i$ [to be $[t_i$ in the room]]]]]]] (= 7b)

This derivation should also converge, but is more costly than (7a = 13c), as it involves Move/Attract of *a man*, instead of Merge of *someone* at the same stage (8). Therefore, (7a = 13c) should be chosen as the optimal derivation, clearly a wrong result.

Chomsky (1995:347) claims that what is wrong about (7a = 13c) is that *someone* fails to receive any θ-role, which causes the derivation to crash: the failure of θ-assignment violates the principle of Full Interpretation. As the crashing derivations do not compete for the purpose of economy, (7b = 14e) will be the only convergent derivation, and no issue of economy arises against (7a =13c).

For this account to go through, unassigned θ-roles need to be treated on a par with unchecked [– interpretable] formal features, both violating the principle of Full Interpretation, which is yet to be clearly formulated.

1.2 Non-Deterministic Local Economy
1.2.1 Collins' (1997): Overview

Collins (1997) makes a proposal which in effect renders Merge and Move/Attract equal in cost. He argues that economy conditions should be formulated in a local fashion, in the sense that the decision about whether an operation may apply to Σ (the set of syntactic objects) is made only on the basis of information available in Σ. Collins takes Move/Attract to be essentially a combination of Copy and Merge, and Select to be the Copy operation out of the lexicon (*op. cit.*:90ff.). Thus, Select + Merge and Move/Attract both as Copy + Merge need to be triggered as last resort operations. Collins (*op. cit.*:9) proposes Last Resort as an independent local economy condition, rather than a defining property of an operation, as in the following:

(15) *Last Resort*
An operation OP involving α may apply only if some property of α is satisfied.

For Move/Attract, Collins adopts the standard Minimalist assumption that the relevant property to be satisfied is feature-checking. For Merge, Collins (*op. cit.*:66) proposes the following principle called Integration:

(16) *Integration*
Every category (except the root) must be contained in another category.

Collins (*op. cit.*:9) also proposes Minimality, as another independent local condition.

(17) *Minimality*
An operation OP (satisfying Last Resort) may apply only if there is no smaller operation OP′ (satisfying Last Resort).

For Move/Attract, Collins adopts the Minimal Link Condition of the standard minimalist conception as the measure of "smallness" of the operation. For Merge, the number of merged objects is counted, introducing a new formulation of Merge, which he calls Unrestricted Merge (*op. cit.*:75ff.).

Unrestricted Merge is a generalized grouping operation that applies to any number of constituents. But the vacuous application of Unrestricted Merge to a null element does not satisfy Last Resort, so that it is not possible. The binary application to two elements is smaller than ternary, quadripartite, etc. applications involving more than two elements. Given Last Resort and Minimality, the unary application to a single element is the "smallest" application of Unrestricted Merge, but if we allow such a unary application, it inevitably yields infinite recursion, so that it is stipulated as impossible. Therefore, only the binary application involving two elements is chosen by Minimality.

Given these preliminaries, Collins claims that Move/Attract and (Unrestricted) Merge are not comparable, and they are equally economical insofar as they obey Minimality, so that neither one is more costly than the other. In effect, they are equal in cost, and the options are not just potential, but real, whenever there is a choice.

1.2.2 Problems with Collins' Non-Deterministic Economy

Collins' local economy faces an additional over-generation problem that Chomsky's global static economy does not have, as Collins' local economy is non-deterministic, yielding a wider range of options. Consider, first, the super-raising case in (3), repeated as (18) below:

(18) *It$_i$ I^0 seems that [$_{IP}$ *t*$_i$ was told John [$_{CP}$ that IP]]
(19) [$_{IP}$ it was told John [$_{CP}$ that IP]]

Insofar as the stage (19) can be generated, the same over-generation problem arises as in Chomsky's global static economy. Collins' local economy cannot block (19), as merging of *it* and raising of *John* are presumably not comparable, and hence equally economical in effect. Thus, as in Chomsky's global static economy, (18 = 3) should be a legitimate derivation, contrary to fact.

An additional over-generation problem arises with regard to the *there*-raising contrast (6), repeated below as (20).

(20) a. There seems to be a man in the room.
b. *There seems a man to be in the room.

As with Chomsky, the common intermediate stage was (8), repeated below:

(21) [$_{I'}$ to be [a man in the room]]

At this point, Chomsky was forced to merge *there.* Collins, on the other hand, has an option to raise *a man,* or Merge any element, as Collins (1977:89ff.) rejects the postulation of numeration as an unnecessary theoretical construct.

If we choose to raise *a man*, the result will be (22).

(22) [$_{IP}$ a man$_i$ [$_{I'}$ to be [*t*$_i$ in the room]]]

A possible continuation from this structure is the following:

(23) a. [$_{VP}$ seems [$_{IP}$ a man$_i$ [$_{I'}$ to be [*t*$_i$ in the room]]]]
b. [$_{I'}$ I^0 [$_{VP}$ seems [$_{IP}$ a man$_i$ [$_{I'}$ to be [*t*$_i$ in the room]]]]]
c. [$_{IP}$ a man$_i$ [$_{I'}$ I^0 [$_{VP}$ seems [$_{IP}$ *t'*$_i$ [$_{I'}$ to be [*t*$_i$ in the room]]]]]]

In this way, the grammatical sentence can be successfully generated, but at the same time, the ungrammatical one can also be generated as a continuation from (23b), since there is no numeration in Collins' theory.

(24) [$_{IP}$ there [$_{I'}$ I^0 [$_{VP}$ seems [$_{IP}$ a man$_i$ [$_{I'}$ to be [*t*$_i$ in the room]]]]]] (= 20b = 6b)

This is a perfectly legitimate continuation, with the conditions given so far: Last Resort and Minimality are both respected. Thus, to rule out the cases like (24), Collins (1997:123) proposes the following principle.

(25) *Chain-Formation Principle 4*
If there are two operations OP_1 and OP_2 applicable to a set of representations Σ (both satisfying Last Resort and Minimality), then choose the operation that extends an incomplete chain.

An incomplete chain here roughly means a non-trivial chain that contains an unchecked [– interpretable] feature. The net effect is that once an element is moved, and if it still contains an unchecked [– interpretable] feature, its movement has priority over merging another element.

For the ECM contrast in (7), repeated below as (26), there was the same intermediate stage (21 = 8), repeated again as (27) below:

(26) a. *Someone believes to be a man in the room.
b. Someone believes a man to be in the room.
(27) [$_{I'}$ to be [a man in the room]]

A possible continuation from (27 = 21 = 8) is the following:

(28) a. [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]
b. [$_{V'}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]
c. [$_{VP}$ someone [$_{V'}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]]
d. [$_{I'}$ I^0 [$_{VP}$ someone [$_{V'}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]]]

Despite the Chain-Formation Principle 4, *a man* cannot raise over *someone* to the matrix Spec(IP) by Minimality. Thus, the grammatical sentence will be successfully derived as in (29 = 26b = 7b).

(29) [$_{IP}$ someone$_j$ [$_{I'}$ I^0 [$_{VP}$ t_j [$_{V'}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]]]]

At the same time, however, the ungrammatical (26a = 7a) will be over-generated from (27 = 21 = 8) as in the following:

(30) a. [$_{IP}$ someone [$_{I'}$ to be [a man in the room]]]
b. [$_{VP}$ believes [$_{IP}$ someone [$_{I'}$ to be [a man in the room]]]]
c. [$_{I'}$ I^0 [$_{VP}$ believes [$_{IP}$ someone [$_{I'}$ to be [a man in the room]]]]]
d. [$_{IP}$ someone$_i$ [$_{I'}$ I^0 [$_{VP}$ believes [$_{IP}$ t_i [$_{I'}$ to be [a man in the room]]]]]] (= 26a = 7a)

For Chomsky, this is a crashing derivation (12–13), since *someone* fails to receive any θ-role, violating the principle of Full Interpretation. For Collins (1997:71ff.), however, a violation of the θ-Criterion does not lead to a crashing derivation, but simply yields an "uninterpretable" derivation. That is, the "uninterpretable" derivation (30d = 26a = 7a) is over-generated as convergent.

Furthermore, as Collins rejects postulating a numeration, (29) is not the only continuation from (28d). Instead of raising *someone*, *there* could have been merged to the matrix Spec(IP) as below:

(31) *[$_{IP}$ there [$_{I'}$ I^0 [$_{VP}$ someone [$_{V'}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]]]]

After this stage, *a man* can covertly check the Case-feature and the ϕ-features of *believes*, and *someone* with the matrix I^0. There should be no interpretive problem: θ-assignment is complete, and the existential reading for *there* should also be obtained — even if *a man* cannot be associated with *there* in this configuration, *someone* should be able to — thus, (31) should converge as perfectly grammatical.[1]

To summarize, Collins' local economy, forfeiting the comparison between Merge and Move/Attract, is non-deterministic, which exacerbates the over-generation problem. It requires independent conditions to rule out unwanted over-generation, such as the Chain-Formation Principle 4, which does not seem to have any application other than to rule out the cases like (24 = 20b = 6b).

2. Move First Plan: Local Determinacy

The problem of the super-raising case (3 = 18), repeated below as (32), is that *it* is extracted from the tensed clause.

(32) *It_i I^0 seems (that) $[_{IP}$ t_i was told John $[_{CP}$ that IP]]

This is a classic case of a Tensed-S Condition violation (Chomsky 1973), which has not received any satisfactory account yet in the minimalist literature.[2]

The problem of (32 = 3 =18) arises at the point when *it* is merged at the embedded Spec(IP), in spite of the fact that *John* is present that can check all the feature of the embedded I^0 as well as *John*'s own. We have seen that insofar as the stage (19), repeated below as (33), can be generated, it is doomed to over-generate (32 = 3 = 18).

(33) $[_{IP}$ it was told John $[_{CP}$ that IP]]

Before *it* was merged to Spec(IP) in (33 = 19), there was the stage (4), repeated below as (34).

(34) $[_{I'}$ was told John $[_{CP}$ that IP]]

For Collins' local economy, available options are raising *John* and merging whatever element. Since unmoved elements are not subject to the Chain-Formation Principle 4, Collins could not block the merger of *it*.

Suppose, however, that Move/Attract has priority over Merge. Then, (35) will result, and the derivation can further continue as (36).

(35) $[_{IP}$ John_i $[_{I'}$ was told t_i $[_{CP}$ that IP]]]

---

[1] Superficially, the equivalent of (31) appears grammatical in languages that allow the so-called Transitive Expletive Construction (TEC), but it is ungrammatical in English, and hence it must be ruled out somehow. TEC typically involves an expletive that surfaces as the clause-initial element in the V2 context. Thus, it is not obvious that the expletive in TEC is base-generated as Spec(IP) as in (31).

[2] Nunes (1995) claims that what is wrong with (32 = 3 = 18) is that the trace of *it* is crossed by the covert raising of *John*, inducing a violation of the Minimal Link Condition; this cannot be correct, however, as we will see below in (43–45). The discussion here is concerned only with the Tensed-S Condition effect for movement, not its effect for non-movement binding dependency, originally included in Chomsky (1973).

(36) a. [$_{CP}$ (that) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]
b. [$_{VP}$ seems [$_{CP}$ (that) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]]
c. [$_{I'}$ I^0 [$_{VP}$ seems [$_{CP}$ (that) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]]]
d. [$_{IP}$ it [$_{I'}$ I^0 [$_{VP}$ seems [$_{CP}$ (that) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]]]]

This derivation satisfies all the morphological requirements of the embedded I^0 by raising *John*, and those of the matrix I^0 by merging *it*. The unwanted derivation (32 = 3 = 18), on the other hand, satisfies the morphological requirements of the embedded I^0 by merging *it*, and those of the matrix I^0 by overt raising of *it* and covert raising of *John*. That is, the desired derivation (35–36) involves one step fewer than the unwanted derivation (32 = 3 = 18): one movement for the former, and two movements for the latter, relevant merger being once for both. Thus, ironically, Chomsky's global static economy, imposing the fewest derivational steps, may have correctly chosen the desired derivation — but only if Move/Attract could have taken precedence over Merge at stage (34 = 4).

Nevertheless, the preference for Move/Attract over Merge cannot simply be fixed globally as a static principle, since it will violate the Tensed-S Condition in another context. Consider, again, the desired derivation from (34 = 4), repeated below as (37).

(37) [$_{I'}$ was told John [$_{CP}$ that IP]]

We want Move/Attract to take precedence over Merge at this point, yielding (35), repeated below as (38).

(38) [$_{IP}$ John$_i$ [$_{I'}$ was told t_i [$_{CP}$ that IP]]]

The derivation should continue as (36a–c), but at stage (36c), if Move/Attract is always favored over Merge, raising of *John* will be chosen again over merging *it*, yielding the following:

(39) *[$_{IP}$ John$_i$ [$_{I'}$ I^0 [$_{VP}$ seems [$_{CP}$ (that) [$_{IP}$ t'_i [$_{I'}$ was told t_i [$_{CP}$ that IP]]]]]]]

This derivation crashes, of course, since *John* cannot check the Case-feature of the matrix I^0, violating the Tensed-S Condition again. That is, although the ungrammatical (39) is correctly ruled out, the static preference of Move/Attract over Merge makes the merger of the expletive *it* never possible, and the grammatical (36d) cannot be generated. We cannot simply appeal to a numeration, in order to force the merger of *it* here; there is no guarantee that a numeration contains *it*.

3. Principle of Minimum Feature Retention: The Dynamic Principle
3.1 Super-Raising Revisited

What we need is a dynamic economy measure that deterministically chooses an appropriate operation depending on the contexts. To seek such a measure, let us return to the first decisive stage (37 = 4 = 34), repeated as (40) below, of the derivation for the super-raising case.

(40) [$_{I'}$ was told John [$_{CP}$ that IP]]

Here, the choice arises: whether to merge some element, or raise *John*. We want *John* to raise, instead of merging the expletive *it* in this case.

The next decisive point in the derivation is (36c), repeated below as (41), upon which we want *it* to be merged, instead of raising *John* again.

(41) $[_{I'}\ I^0\ [_{VP}$ seems $[_{CP}$ (that) $[_{IP}$ John$_i$ $[_{I'}$ was told t_i $[_{CP}$ that IP]]]]]]

What could motivate this difference? From the stage (40 = 4 = 34 = 37), the raising of *John*, will check the EPP-feature, the ϕ-features, and the Case-feature of I^0, as well as the Case-feature of *John*, while if *it* were merged, even though all the relevant features of I^0 can be checked, the Case-feature of *John* will be left unchecked. On the other hand, merging *it* to (41 = 36c) will check the EPP-feature, the ϕ-features, and the Case-feature of the matrix I^0 as well as the Case-feature of *it*, whereas raising *John* will leave the Case-feature of the matrix I^0 unchecked. That is, the unwanted operations, merging *it* in the former and raising *John* in the latter, both leave a feature unchecked that is already in the structure. Exploiting this fact, I propose the following principle:

(42) ***Principle of Minimum Feature Retention***
Choose the operation that leaves the fewest unchecked [– interpretable] features in the structure produced by that operation.

The intuitive idea behind this principle is that an operation that eliminates more features is more economical than another operation that eliminates fewer features. Informally speaking, keeping track of unchecked features imposes heavier memory load than just checking them off, so that the former is more costly than the latter. This is not to make any direct psychological claim: the memory requirement is a more important factor than the time measurement of computational complexity on purely formal grounds (section 5).

If *John* was not in the derivation at the stage (40 = 4 = 34 = 37), i.e., the verb tell is used as a mere transitive, not as a ditransitive, *it* could have been merged as in (43).

(43) $[_{IP}$ it was told $[_{CP}$ that IP]]

Suppose that (43) is further embedded under *seems* and I^0 is merged as in the following:

(44) $[_{I'}\ I^0\ [_{VP}$ seems $[_{CP}$ (that) $[_{IP}$ it was told $[_{CP}$ that IP]]]]]

Raising of *it* to the matrix Spec(IP) will check the EPP-feature and the ϕ-features of the matrix I^0, but not the Case-feature of the matrix I^0, as *it* has lost its Case-feature when merged to the embedded Spec(IP). Thus, the derivation crashes.

(45) *$[_{IP}$ it$_i$ $[_{I'}\ I^0\ [_{VP}$ seems $[_{CP}$ (that) $[_{IP}$ t_i was told $[_{CP}$ that IP]]]]]]]

If the numeration had another *it*, that would have been merged to the matrix Spec(IP), yielding the following convergent derivation:

(46) $[_{IP}$ it $[_{I'}\ I^0\ [_{VP}$ seems $[_{CP}$ (that) $[_{IP}$ it was told $[_{CP}$ that IP]]]]]]

Merging another *it* will check the EPP-feature, the Case-feature, and the ϕ-features of the matrix I^0 as well as the Case-feature of that *it*, whereas the raising of the embedded *it* checks only the EPP-feature and the ϕ-features of the matrix I^0 — the Case-feature of the embedded *it* is already checked when merged to the embedded Spec(IP). Thus, the merger of another *it* is more economical than raising of the embedded *it*, according to the Principle of Minimum Feature Retention (MFR). In this way, the MFR correctly allows the preference of Merge over Move/Attract in cases like this.

In the same vein, if (40 = 4 = 34 = 37), after raising *John*, is further embedded under *seems* and higher structure as in (47), merging *it* is more economical than raising *John* again.

(47) $[_{I'}$ I^0 $[_{VP}$ seems $[_{CP}$ (that) $[_{IP}$ John$_i$ was told t_i $[_{CP}$ that IP]]]]]

This correctly captures the Tensed-S Condition effect, with the MFR dynamically making a "locally" deterministic choice whether to apply Move/Attract or Merge case by case, depending on how many features will be left unchecked by each operation.[3]

4. Selected *There*

An immediate question will arise with respect to the *there* constructions, as it was the primary reason to maintain Merge as equally or less costly than Move/Attract. In a simple finite context, the MFR appears to be too strong. Consider the following pair:

(48) a. $[_{IP}$ there $[_{I'}$ is [someone in the room]]]
b. $[_{IP}$ someone$_i$ $[_{I'}$ is $[t_i$ in the room]]]

These presumably had the following common intermediate stage:

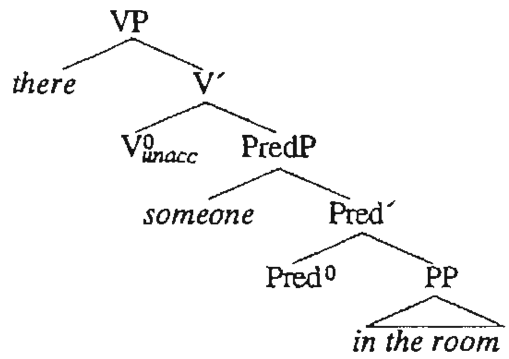
(49) $[_{I'}$ is [someone in the room]]

Raising of *someone* can check the EPP-feature, the Case-feature, and the ϕ-features of I^0, as well as the Case-feature of *someone*, whereas the merger of *there* can check only the EPP-feature of I^0. Thus, by the MFR, (48a) should be blocked in favor of (48b) — an undesired result — a problem of under-generation.

Nevertheless, this under-generation problem arises, only if *there* is merged at Spec(IP), as standardly assumed. Contra Chomsky (1995:363ff.), I claim that *there* is not merged at Spec(IP), but to the external domain of an unaccusative verb as below, assuming that the existential *be* is an unaccusative verb:

[3] Here, the notion of "local" is different from the one defined in Collins (1997:4). It is "local," in the sense that it applies to a local transitional relation in the derivational sequence, $<\Sigma_n, \Sigma_{n+1}>$.

(50)



This kind of structure is motivated by the fact the expletive *there* is virtually limited to an unaccusative construction, modulo the definiteness restriction.

(51) a. There [$_{VP}$ arrived [three men (from Chicago)]].
b. *There [$_{vP}$ many girls danced (in the disco)].
c. *There [$_{vP}$ some boys played baseball (in the park)].

This conforms to the fact that the expletive *there* cannot stand in for the base object, unlike the "expletive" *it* (Postal and Pullum 1988; *cf.* Rothstein 1995). Thus, I take it justified that *there* is optionally selected as a kind of "indirect object" by an unaccusative verb, when its complement subject is indefinite.[4]

Thus, the relevant intermediate stage for (48a) was not (49), but the following:

(52) [$_{I'}$ is [$_{VP}$ there t_{is} [someone in the room]]]

Even though the raising of *someone* would leave fewer features than the raising of *there*, *someone* cannot raise over *there* by the Minimal Link Condition. If there is another DP category available, the MFR will choose the merger of that DP, since raising of *there* will leave the ϕ-features and the Case-feature of I^0, as well as the Case-feature of *someone* unchecked, whereas the merger of that DP will leave only the Case-feature of *someone* unchecked.

(53) [$_{IP}$ DP [$_{I'}$ is [$_{VP}$ there t_{is} [someone in the room]]]]

However, such a derivation crashes, as *someone* cannot check its Case-feature. Yet, it means that if such a DP is available, (48a) will never be generated, given the MFR as a local condition. In turn, this means that a numeration is still necessary.[5]

In (48b), *there* is not selected, and the relevant intermediate stage was (54).

[4] Lasnik (1995:624ff.) considers a possibility of a selectional restriction for *there* by an unaccusative verb, but dismisses it as hard to state in terms of semantic coherence, and due to the semi-productive nature of the constructions, yet pointing out that it deserves further investigation if it could be stipulated in some way that *there* can be introduced only as the subject of unaccusative verbs. Here, I am appealing to selection as subcategorization of an "indirect object," not (semantic) selectional restriction.

[5] This is different from the claim I made in oral presentation that the MFR is neutral to the necessity of a numeration, which was an oversight.

(54) $[_{I'}$ is $[_{VP}$ t_{is} [someone in the room]]] (= 49)

Merger of another DP is not an option in this case, as such merger will leave the Case-feature of *someone*, whereas raising it will leave no features that need to be checked.[6]

(55) $[_{IP}$ someone$_i$ $[_{I'}$ is $[_{VP}$ t_{is} [t_i in the room]]]]

Let us return now to the contrast in the raising context (6 = 20), repeated as (56) below.

(56) a. There seems to be a man in the room.
b. *There seems a man to be in the room.

Given that *there* is selected by *be*, the common intermediate stage is (57):

(57) $[_{I'}$ to $[_{VP}$ there be [a man in the room]]]

As *there* is closer than *a man* is to I^0, *there* will be raised. In this case, either raising *there* or *a man* would check only the EPP-feature of I^0, and the number of unchecked features in the derived structure is the same, viz. one, the Case-feature of *a man*. On the other hand, merger of another DP will add an extra unchecked feature in the structure that is not present in (57): the Case-feature of that DP. Therefore, the MFR will chooses raising of *there*.[7]

(58) $[_{IP}$ there$_i$ $[_{I'}$ to $[_{VP}$ t_i be [a man in the room]]]]

The derivation can further continue as in the following:

(59) a. $[_{VP}$ seems $[_{IP}$ there$_i$ $[_{I'}$ to $[_{VP}$ t_i be [a man in the room]]]]]
b. $[_{I'}$ I^0 $[_{VP}$ seems $[_{IP}$ there$_i$ $[_{I'}$ to $[_{VP}$ t_i be [a man in the room]]]]]]
c. $[_{IP}$ there$_i$ $[_{I'}$ I^0 $[_{VP}$ seems $[_{IP}$ t'_i $[_{I'}$ to $[_{VP}$ t_i be [a man in the room]]]]]]]] (= 56a)

If *there* is not selected, *a man* will raise as below:

(60) a. $[_{I'}$ to be [a man in the room]]
b. $[_{IP}$ a man$_i$ $[_{I'}$ to be [t_i in the room]]]

Again, merger of another DP at stage (60a) will add an extra unchecked feature in the structure, so that it is not an option.

The derivation can further continue as in the following:

(61) a. $[_{I'}$ I^0 $[_{VP}$ seems $[_{IP}$ a man$_i$ $[_{I'}$ to be [t_i in the room]]]]]
b. $[_{IP}$ a man$_i$ $[_{I'}$ I^0 $[_{VP}$ seems $[_{IP}$ t_i' $[_{I'}$ to be [t_i in the room]]]]]]

---

[6] The copulative *be* in the so-called predicative or equative constructions is irrelevant here, as it is presumably a transitive verb, not an unaccusative verb.
[7] This is where the derivationally-transitional notion of "locality" (fn.3) comes in.

Merger of another DP at stage (61a) will also add an extra unchecked feature in the structure, besides leaving the Case-feature of *a man* unchecked. Again, such merger of another DP is therefore not an option. This captures the effect of Collins' Chain-Formation Principle 4, without independently stipulating it. In this way, (56b = 6b = 20b) is correctly blocked, and the contrast in (56 = 6 = 20) follows.

One may wonder why the raising verb *seem* cannot optionally select *there*, as unaccusative verbs and the raising verbs share common properties in relevant respects — the lack of an internal Case and an "external" θ-role — and in particular, the structure depicted in (50) without *there* is essentially the raising configuration. It would have yielded the following representation, grammatical for (56b = 6b = 20b).

(62) There$_i$ [$_{I'}$ I^0 [$_{vP}$ seems$_v$ [$_{VP}$ t_i [$_{V'}$ t_v [$_{IP}$ someone$_j$ [$_{I'}$ to be [t_j in the room]]]]]]]

The reason is that the simple unaccusative verbs and the raising verbs have different selectional properties; instead of small clause complement, the former can select a DP complement, whereas the latter cannot, and the position where *there* could have been merged in the raising VP is an optional θ-position for the Experiencer argument, such as *to me* in the following:[8]

(63) *There [$_{I'}$ I^0 [$_{vP}$ seems$_v$ [$_{VP}$ **to me** [$_{V'}$ t_v [$_{IP}$ someone$_i$ [$_{I'}$ to be [t_i in the room]]]]]]]

Thus, *there*, as an expletive without a referential or anaphoric potential, cannot be merged within the raising VP.

If (60b) is embedded under an ECM verb *believe*, for example, (64) will result.

(64) [$_{VP}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]

Then, the derivation can continue as follows, yielding (7b = 26b):

(65) a. [$_{vP}$ someone v^0 [$_{VP}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]]
b. [$_{I'}$ I^0 [$_{vP}$ someone v^0 [$_{VP}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]]]
c. [$_{IP}$ someone$_j$ [$_{I'}$ I^0 [$_{vP}$ t_j v^0 [$_{VP}$ believes [$_{IP}$ a man$_i$ [$_{I'}$ to be [t_i in the room]]]]]]]]

As the merger of another DP at stage (60a) is not an option, (66) cannot be generated, and hence no such derivation as (13a–c = 30b–d), repeated below as (67), is possible.

(66) [$_{IP}$ someone [$_{I'}$ to be [a man in the room]]
(67) a. [$_{VP}$ believes [$_{IP}$ someone [$_{I'}$ to be [a man in the room]]]]
b. [$_{I'}$ I^0 [$_{VP}$ believes [$_{IP}$ someone [$_{I'}$ to be [a man in the room]]]]]
c. [$_{IP}$ someone$_i$ [$_{I'}$ I^0 [$_{VP}$ believes [$_{IP}$ t_i [$_{I'}$ to be [a man in the room]]]]]]

Thus, the contrast for the ECM paradigm in (7 = 26) follows, without an unclear notion of "uninterpretability" as in Collins (1997:71ff.), or taking the failure of θ-assignment to be a Full Interpretation violation as Chomsky does.

---

[8] For the shell-structure of the raising verb, see Chomsky (1995:304ff.)

5. Computational Complexity

In the theory of computational complexity studied in combinatorial mathematics and theoretical computer science, problems formalized as mathematical objects are classified into complexity classes, in terms of computational power and resources that an algorithm requires to solve those problems (Johnson 1990, Papadimitriou 1994, among others). The computational power is modeled on a formal automaton, and the standard model of automaton for the measurement of computational complexity is the basic off-line Turing machine with a single read-write head and a single one-dimensional semi-finite tape, for which the input problem is encoded in a bit sequence.

The resource requirements are usually characterized by the growth rate of time and/or space, with respect to the "size" of the problem. Time is measured by the number of (discrete) operational steps of the algorithm to solve a problem, and space by the memory capacity expended by the algorithm. Any problem solving procedure that can be reasonably called an "algorithm" can be formalized as a Turing machine (Church's Thesis). The complexity class of decision problems that can be solved by *deterministic* Turing machines in *polynomial-bound time* is called the class **P**. The complexity class of decision problems that can be solved by *non*-deterministic Turing machines in polynomial-bound time is called the class **NP**. Informally speaking, the class **P** problems are computationally tractable, that is, "efficiently solvable," whereas the class **NP** problems are not.[9]

As for the relation between the time measurement and the space requirement, it is known that the polynomial-*space* bound classes **(N)PSPACE** properly include the polynomial-*time* bound classes **(N)P**, and the exponential-*space* bound classes **(N)EXPSPACE** properly include the exponential-*time* bound classes **(N)EXP**. That is, the space requirement is more crucial than the time measurement for the overall computational complexity.

A standard measure for economy of derivation posited in the literature is the step count for a derivation to converge. The number of operational steps is the time measurement, and what we need to be more concerned with is the memory load that the computational system of human language has to bear.

The proposal of the Minimum Feature Retention (MFR) is best construed in these contexts, aiming at the reduction of the memory load for the computational system of human language. Leaving [– interpretable] features unchecked imposes more memory load than just checking them off. The MFR does a "computational trick" (Chomsky 1989) that overcome the computational complexity, inherent to a derivational system of the generative procedure.

[9] Although it has not yet been formally proven, it is generally believed that the class **NP** properly includes the class **P**, as non-finite state non-deterministic automata are generally more powerful than deterministic class. It does not mean that the class **NP** problems can never be solved, but that there is no general algorithm known to solve arbitrary instances of their types. If a given instance of a problem is "small" enough, it can be solved without difficulty by an exhaustive, viz. global search.

6. Concluding Remarks

An implication of the MFR is that movement motivated by feature-checking preempts merger that does not involve feature-checking. This means that object shift precedes, if it takes place overtly, the merger of the subject into the *v*P projection. That is, the base position of the subject is higher than the derived object position, a hypothesis which has been proposed in various forms (Johnson 1991, Koizumi 1993, Bobaljik 1995, among others). In fact, his is also a suggestion that Chomsky (1995:358ff.) makes in the Agr-less structure with the multiple specifier potential, as an alternative to the view that the object shifts over the merged subject.

Although the discussion thus far has been tacitly limited to the operations in the overt component, there is no reason that the MFR is inapplicable in the covert component, given the Uniformity Thesis that the computation from Numeration to the LF representation λ is uniform. In fact, I have already included the weak features in counting the number of features checked by an overt operation, so that the strength of features should not matter for the purpose of the MFR. Thus, a natural conclusion is that the MFR holds throughout a derivation, allowing a "covert" operation in the "overt" component.

This entails that no covert component is necessary, and all the syntactic operations should be executed before the operation Spell-Out applies. That is, Spell-Out applies to the LF representation λ that leads to the PF representation π, a hypothesis already proposed in various forms (Bobaljik 1995, Brody 1995, Groat and O'Neil 1996, Pesetsky 1996, among others). The leading idea is that the difference between "overt" and "covert" movement does not have to be expressed as the timing of movement with respect to the Spell-Out point, but can be captured as the determination of which copy of the moved category is "pronounced."

Obviously, the MFR is reminiscent of Pesetsky's (1989) Earliness Principle that requires any syntactic conditions to be satisfied as early as possible in the derivation. Rather than syntactic conditions, the MFR locally requires [– interpretable] features to be checked as many as possible at a given stage of derivations, and thus consequently as early as possible (*cf.* Ishii 1997).

References

Bobaljik, Jonathan D. 1995. Morphosyntax: The syntax of verbal inflection. Doctoral dissertation, MIT.

Brody, Michael. 1995. *Lexico-Logical Form: A radically minimalist theory.* Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. S. R. Anderson and P. Kiparsky, 232–286. New York: Holt, Rinehart and Winston.

Chomsky, Noam. 1989. Some notes on economy of derivation and representation. In *MIT Working Papers in Linguistics* 10:43–74.

Chomsky, Noam. 1992. A minimalist program for linguistic theory. MIT Occasional Papers in Linguistics 1. Cambridge, Mass.: MITWPL.

Chomsky, Noam. 1995. Categories and transformations. In *The Minimalist Program*, 219–394. Cambridge, Mass.: MIT Press.

Collins, Chris. 1997. *Local Economy*. Cambridge, Mass.: MIT Press.

Groat, Erich and John O'Neil. 1996. Spell-out at the LF interface. In *Minimal Ideas: Syntactic studies in the minimalist framework*, ed. W. Abraham, S. D. Epstein, H. Thráinsson and C. J.-W. Zwart, 113–139. Amsterdam: John Benjamins.

Ishii, Toru. 1997. An asymmetry in the composition of phrase structure and its consequences. Doctoral dissertation, University of California, Irvine.

Johnson, David S. 1990. A catalog of complexity classes. In *Handbook of Theoretical Computer Science A: Algorithms and Complexity*, ed. J. van Leeuwen, 67–161. Amsterdam: Elsevier Science/Cambridge, Mass.: MIT Press.

Johnson, Kyle. 1991. Object positions. *Natural Language & Linguistic Theory* 9:577–636.

Koizumi, Masatoshi. 1993. Object agreement phrase and the split VP hypothesis. In *MIT Working Papers in Linguistics* 18:99–148.

Lasnik, Howard. 1995. Case and expletives revisited: On greed and other human failings. *Linguistic Inquiry* 26:615–633.

Nunes, Jairo M. 1995. The copy theory of movement and linearization of chains in the minimalist program. Doctoral dissertation, University of Maryland, College Park.

Papadimitriou, Christos H. 1994. *Computational Complexity*. Reading, Mass.: Addison-Wesley.

Pesetsky, David. 1989. Language-particular processes and the earliness principle. Ms. MIT.

Pesetsky, David. 1996. Some optimality principle of sentence pronunciation. In *Is the Best Good Enough?: Optimality and competition in syntax*, ed. P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky, 337–383. Cambridge, Mass.: MIT Press/MITWPL.

Postal, Paul M. and Geoffrey K. Pullum. 1988. Expletive noun phrases in subcategorized positions. *Linguistic Inquiry* 19:635–670.

Rothstein, Susan D. 1995. Pleonastics and the interpretation of pronouns. *Linguistic Inquiry* 26:499–529.

Toyoshima, Takashi. forthcoming. Optimization of phrase structures and their derivational processes. Doctoral dissertation, Cornell University.

Department of Human Sciences
Faculty of Computer Science and Systems Engineering
Kyushu Institute of Technology
Kawazu 680-4, Iizuka
Fukuoka 820-8502 JAPAN

toyoshima@lai.kyutech.ac.jp