On “Fewest Steps”

E.G. Ruys
Utrecht University

This paper addresses some empirical and theoretical properties of economy constraints as defined in Chomsky 1995, chapter 4. It argues that three such constraints (Procrastinate, Fewest Steps, “No Redundant Features”) have questionable theoretical properties. All are violable, without their interaction being properly defined; the latter two are “global” in such a way as to be incompatible with the overall framework; Procrastinate, furthermore, is underdetermined by the empirical data; and “NRF” appears to defy consistent definition altogether. It is proposed to redefine the Fewest Steps constraint in such a way that the effects of the other two can be derived from this constraint, without the accompanying theoretical problems. This move will also be seen to resolve some serious empirical problems associated with these constraints, which lie in the areas of expletive-placement in English and Subject-Object asymmetries in overt Accusative-checking languages.

1. Introduction

Recent developments in the Minimalist framework have shown a move away from violable, global economy constraints on derivations, in favor of inviolable, local constraints on movement (see Chomsky 1995, Chapter 4; hereafter: Chapter 4). This leads to a considerable reduction in computational complexity within the grammar, as well as providing a principled solution to complex questions arising from the interaction of multiple violable constraints.

Nevertheless, some violable constraints have as yet appeared irreducible, notably Procrastinate, Fewest Steps and Chomsky's "No Redundant Features" constraint; the latter two remaining not only violable, but global as well. In this paper, I will argue that the effects of both Procrastinate and "No Redundant Features" can be derived from Fewest Steps, if properly defined. This leaves us with just one violable, global constraint: a highly desirable result from a conceptual standpoint. Also, I will argue that this reduction has considerable empirical advantages.

I will start with a typology of (economy) constraints (section 2). I will then discuss some empirical and conceptual problems in Chapter 4, and propose a solution. These problems concern Chomsky’s characterization of Procrastinate, his analysis of *there* and *it* insertion, and his “No Redundant Features” condition; the solution comes from a refinement of Fewest Steps.

2. Constraint Types

For the purposes of our discussion, it is necessary to distinguish the following types of constraints: relative constraints vs. absolute constraints, and local vs. global constraints.

I will call a constraint *relative* (or “violable”) when it excludes a derivation (or a step in a derivation) iff the reference set of that derivation (or step) contains an alternative which violates the constraint to a lesser degree. A typical example is Last Resort. Suppose that for any given derivation, the reference set of that derivation with respect to Last Resort is the set of converging derivations with the same underlying numeration. Then Last Resort rules out a derivation A iff A requires a certain operation α, and there exists a convergent derivation B with the same numeration as A which does not require α. If there is no such alternative, however, then A and its operation α are not excluded by Last Resort.

For discussion and comments I am grateful to Peter Ackema, Peter Coopmans, Johan Kerstens, Ellen-Petra Kester, Tanya Reinhart, Maaike Schoorlemmer, and Fred Weerman.

An *absolute* constraint, on the other hand, excludes certain derivations *irrespective as to whether an alternative derivation is available*. Since a derivation that violates an absolute constraint is ruled out without being compared with possible alternative derivations, it is not necessary to provide a definition of Reference Sets for such constraints.

As for the distinction between "local" and "global" constraints: suppose there is a derivation (whose length may be zero) up to point α, and from α there is a choice between the continuations $\beta_1 — \gamma_1 — \delta_1$ and $\beta_2 — \gamma_2 — \delta_2$:

$$\alpha \left\langle \begin{array}{l} \beta_1 — \gamma_1 — \delta_1 \\ \\ \beta_2 — \gamma_2 — \delta_2 \end{array} \right.$$

I will call a constraint "local" if it weighs the costs of β_1 and β_2 and chooses on the basis of these costs, *irrespective of the costs of* $\gamma_1 — \delta_1$ *and* $\gamma_2 — \delta_2$. A "global" constraint, on the other hand, computes the sums (or, conceivably, some other function) of the costs of $\beta_1 — \gamma_1 — \delta_1$ and of $\beta_2 — \gamma_2 — \delta_2$ and chooses between β_1 and β_2 accordingly. When a constraint is local it is possible in principle that it favors a derivation which in the end, from a global perspective, comes out as more costly.[1]

In Chapter 4, Chomsky redefines most existing economy constraints. While most constraints in previous frameworks (see e.g. Chomsky 1991, 1993) were relative (violable) and all were global, most constraints in Chapter 4 are absolute, and some are argued to be local. Chomsky provides an absolute implementation of most economy-constraints by incorporating them into the definition of the movement-transformation. In that way, other options simply cannot be derived and cannot be taken into consideration. In OT-type terminology: most constraints are reinterpreted as properties of GEN. Once a condition is part of the definition of move, it is also local, since move does not create derivations in one fell swoop but creates them "step by step." Hence, most conditions become not only absolute, but local as well.

Chomsky notes several reasons for preferring absolute, local constraints over relative and global ones, most having to do with a reduction of computational complexity. In order to judge a derivation w.r.t. an *absolute* constraint it is no longer necessary to compare a derivation (or a step) with a possibly infinite set of alternative derivations (or steps). Furthermore, in order to judge a derivation w.r.t. a *local* (relative) constraint, it is no longer necessary to compare a set of sequences of steps; one need only compare a set of single steps. Also, Reference Sets for local constraints will presumably be smaller than reference sets for global constraints, since one compares only those derivations that are identical up to the point of the operation being judged (i.e. those derivations that have α in common). Finally, Chomsky points out that the number of available alternatives will decrease as the derivation progresses.

---

[1]Our definition differs from Chomsky's, in that Chomsky also considers a constraint "local" if it compares global costs, but these costs can be predicted locally through some precalculation. Such a constraint can be global in our terms. Conversely, a local constraint on our definition can be "global" in the sense that computationally, lookahead is still necessary (e.g. to determine whether a possible step will lead to convergence). So, we have two disjoint notions of "local/global"; unlike Chomsky we shall not conflate the two and consider only one.

In addition, there is one clearly identifiable problem associated with the use of violable constraints. Consider a situation where derivation α is costly w.r.t. condition A, but inexpensive w.r.t. condition B, whereas derivation β is inexpensive for A but expensive for B. Now in order to decide which derivation is cheaper, one may be forced to add up different, incommensurable "costs."[2] The complex questions that arise in such a situation can be resolved in two ways. One option is to formulate a general theory of constraint interaction, as in Optimality Theory (see e.g. McCarthy & Prince 1994). The other option, pursued here, is to allow only absolute economy constraints (or at most one relative constraint).

Finally, of course, there may be empirical arguments in favor of absolute and local constraints; I shall consider at least one such argument below.

The remainder of this paper is organized as follows. First, in section 3 I will discuss Chomsky's analysis of *there* and *it* insertion. The analysis of *there* insertion is based on a crucially local view of Procrastinate. I will discuss a serious empirical problem for this analysis. I will then show that this problem can be solved if we keep Chomsky's analysis virtually intact, but make use of a (partly local) version of Fewest Steps, instead of Procrastinate. This seems to be a step back, since we then need two global, relative constraints: Fewest Steps, as well as Procrastinate. But in section 4, I will show that we can derive Procrastinate from Fewest Steps; this leaves us with just the one relative, partly global constraint. Finally, in section 5, I will discuss Chomsky's analysis of NP-to-Case movement. I will show, firstly, that this analysis gives independent evidence for Fewest Steps. Furthermore, I will discuss Chomsky's "No Redundant Features" constraint. I will show that this constraint is global and violable (hence should be avoided on general grounds) and, furthermore, does not work as stated, and finally, that its effects (at least those discussed here) can *also* be derived from Fewest Steps. So, rather than being faced with three global and violable constraints (partly ineffective), we are left with just one such constraint, Fewest Steps, which is effective for the data discussed here.

3. There *and* It*: Procrastinate or Fewest Steps*

Chomsky discusses the following pair:

(1) a. There seems t to be someone in the room.
  b.* There seems someone to be t in the room.

Why is (1a) in and (1b) out? Chomsky's analysis runs as follows. Both derivations have (2) in common as a subderivation.

(2) $[_{I'}$ $[_{INFL[EPP]}$ to] $[_{VP}$ be someone in the room]]

In (2), EPP must be checked on INFL. There are two possible ways of achieving this: move *someone* to Spec,IP, or insert *there*. If we move *someone*, we derive (1b) through the steps indicated in (3):

---

[2]An example of this type of problem is the interaction of Fewest Steps and Shortest Link as discussed in Chomsky 1993. If a given movement operation can take few or many intermediate landing sites, then a lower number of operations (sub-movements) implies an increase in the length of links, and vice versa. Chomsky solved this specific problem, not by adopting a general theory of constraint interaction, but by ensuring that the example situation cannot arise (a movement operation which involves various sub-movements counts as one operation Form Chain). See Kitahara 1995, note 26 for another illustration of this type of problem.

(3) 1. $[_{IP}$ someone $[_{INFL}$ to] $[_{VP}$ be $t_{someone}$ in the room]]
2. $[_{IP}$ there INFL seems $[_{IP}$ someone $[_{INFL}$ to] $[_{VP}$ be $t_{someone}$ in the room]]
3. $[_{IP}$ there FF(someone) INFL seems $[_{IP}$ someone $t_{FF(someone)}$ $[_{INFL}$ to] $[_{VP}$ be $t_{someone}$ in the room]]]

After movement of *someone* (step 1) and further derivation, *there* must be inserted to check EPP on matrix-INFL and deplete the numeration (step 2). Finally, after Spell-Out the formal features associated with *someone*, FF(*someone*), move to matrix INFL, checking NOM and φ-features (step 3). This derivation results in the ill-formed (1b). The derivation has one overt movement (violation of Procrastinate) and one covert movement. Now consider the derivation of the well-formed (1a). Again, we start from (2), but we proceed as in (4):

(4) 1. $[_{IP}$ there $[_{INFL}$ to] $[_{VP}$ be someone in the room]]
2. $[_{IP}$ there INFL seems $[_{IP}$ t_{there} $[_{INFL}$ to] $[_{VP}$ be someone in the room]]]
3. $[_{IP}$ there FF(someone) INFL seems $[_{IP}$ t_{there} $[_{INFL}$ to] $[_{VP}$ be someone $t_{FF(someone)}$ in the room]]]

Step 1 inserts *there*. Subsequently, matrix EPP must be checked and the Minimal Link Constraint (MLC) requires that this be done through movement of *there* (step 2). Finally, after Spell-Out, FF(*someone*) move to matrix-INFL, checking NOM and φ (step 3). This derivation has one overt movement (violation of Procrastinate) and one covert movement, as well. The result is well-formed (1a).

Given that both derivations have the same number of movements and violations of Procrastinate, why should (1b) be out and (1a) in? Chomsky's answer is that (1b) violates Procrastinate earlier in the derivation. At the common point in the derivation, (2), there is a choice between moving *someone*, and inserting *there*. Now assume that Procrastinate is a local constraint. It then prefers insertion over movement at this choice point, and the fact that insertion will inevitably be followed by a Procrastinate violation further down the line is “invisible.” Thus, although both derivations are equally costly from a global perspective, (1a) is preferred because the relevant constraint (Procrastinate) operates on a purely local basis. (We have then an empirical argument for a local constraint.)

Now note, first of all, that we can replace “Procrastinate” with “Fewest Steps,” in the analysis presented above, and have the same result fall out in the same way. At the choice point (2), we have a choice between insertion and movement. Given the fact that insertion is “costless,” whereas movement bears a cost, Fewest Steps will prefer insertion, and we derive (1a). (1b) is equally costly w.r.t. Fewest Steps from a global perspective (both have two movement operations, one overt and one covert), but by assuming that Fewest Steps is a purely local constraint, we derive that (1a) is preferred over (1b).

Although “Local Fewest Steps” and “Local Procrastinate” give the same result in this case, I will argue next that “Fewest Steps” is preferable when *it*-expletives are taken into account. Consider first the examples in (5).

(5) a. it_i seems $[_{IP}$ t_i to appear to John $[_{CP}$ that ..]]
b.* it seems $[_{IP}$ $John_j$ to appear to t_j $[_{CP}$ that ..]]

The contrast in (5) shows that *it*-insertion behaves just like *there*-insertion in (1). Both derivations converge, but (5a) is locally cheaper because at the choice point where EPP must be checked on the embedded INFL, a choice has been made for merger instead of movement.

Again, both derivations are equally expensive *globally*; only *local* Procrastinate or Fewest Steps can make the correct distinction. But the situation is reversed in (6):

(6) a.* it seems [$_{CP}$ that t was told John [$_{CP}$ that ...]]
b. it seems [$_{CP}$ that John was told t [$_{CP}$ that ...]]

The contrast between (6a) and (6b) is problematic. Their common substructure is:

(7) [$_{I'}$ INFL[EPP] was told John [$_{CP}$ that ...]]

At this point, there is a choice: move *John*, or insert *it*. If we move *John* (violating local Procrastinate or Fewest Steps), the derivation proceeds as follows:

(8) 1. [$_{IP}$ John INFL was told t_{John} [$_{CP}$ that ...]]
2. [$_{I'}$ INFL[EPP] seems [$_{CP}$ that [$_{IP}$ John INFL was told t_{John} [$_{CP}$ that ...]]]]
3. [$_{IP}$ it INFL seems [$_{CP}$ that [$_{IP}$ John INFL was told t_{John} [$_{CP}$ that ...]]]]

Followed by Spell-out. This derives the well-formed (6b). Alternatively, we can start from (7) and insert *it* (obeying local Procrastinate or Fewest Steps):

(9) 1. [$_{IP}$ it INFL was told John [$_{CP}$ that ...]]
2. [$_{I'}$ INFL[EPP] seems [$_{CP}$ that it INFL was told John [$_{CP}$ that ...]]]
3. [$_{IP}$ it INFL seems [$_{IP}$ that t_{it} INFL was told John [$_{CP}$ that ...]]]

This is followed by Spell-Out. After Spell-Out, we can move FF(*John*) to matrix-INFL, checking NOM on both. The result is:

(10) [$_{IP}$ it FF(John) INFL seems [$_{IP}$ that t_{it} INFL was told John $t_{FF(John)}$ [$_{CP}$ that ...]]]

This derivation converges, and derives the ill-formed (6a). On the basis of the preceding discussion of *there*-insertion, then, we should expect (6a) to be favored over (6b), since the derivation of (6a) is locally more economical.

Chapter 4 attempts to solve this problem by claiming that (6a) does not converge: the Case features on matrix INFL as well as the Case features on *John* remain unchecked — in that case, (6a) is not contained in the Reference Set for (6b) and (6b) is not blocked. But we have seen that this solution is not correct. Note, that in the derivation of the well-formed (1a), Case features on matrix-INFL and on *someone* also remain unchecked in overt syntax. Nevertheless, the derivation converges, because FF(*someone*) can move to matrix-INFL at LF, so that both features are checked. Nothing prevents the same operation from applying in (6a) (see (10)).

So, in sum, local Procrastinate (or Fewest Steps) prefers derivation (7)-(9)-(10) over derivation (7)-(8), hence favors ill-formed (6b) over well-formed (6a). I shall skip discussion of various implausible ways out, and move straight to the solution I want to propose. If we compare ill-formed (7)-(9)-(10) to well-formed (7)-(8), it turns out that the ill-formed derivation has more movement operations: it requires a covert movement of FF(*John*), whereas (7)-(8) requires no covert movement. Thus, although (7)-(9)-(10) is less expensive from a local perspective, it is more expensive from the perspective of a *global* Fewest Steps constraint. I propose, then, a Fewest Steps constraint which is partly local, and partly global:

(11) **Fewest Steps** (FS)
Given two derivations, choose the one with the fewest steps; in the case of a tie, choose the one which is locally cheaper (which takes the cheaper option at the choice point where the derivations diverge).

This gives the same result for *there*-insertion in (1) and *it*-insertion in (5): both derivations are globally equally expensive, hence local economy decides. But in the case of *it*-insertion in (6), global Fewest Steps prefers (6b), as desired.

One may argue that Fewest Steps as we have now defined it is actually two constraints, not one. The point is hardly worth discussing, apart from the fact that, if these are two conditions, their interrelation and interaction is well-defined, and does not cause the type of problems associated with the interaction of violable constraints discussed in section 2.

This reanalysis allows us to choose between Fewest Steps and Procrastinate as the constraint to be held responsible for the data discussed in this section. Whereas Fewest Steps allows for a natural (global) extension in order to capture the data in (6), a similar extension of Procrastinate will not work: (6a) and (6b) each violate Procrastinate once, hence if we were to add a global Procrastinate clause, it would be inoperative and the local clause would incorrectly choose (6a).

This section leaves us with an addition to our rule-set: a global clause in the Fewest Steps constraint. In the next section however, I will argue that adopting local Fewest Steps allows us to eliminate Procrastinate altogether.

4. Deriving Procrastinate

The Procrastinate principle, introduced in Chomsky 1991, is a relative (violable) constraint. A derivation that has an overt operation is excluded by Procrastinate iff the Reference Set of that derivation contains a (converging) alternative derivation that has fewer overt operations. As pointed out in section 2, we want to avoid such constraints. In particular, since we have other relative constraints (such as Fewest Steps) next to Procrastinate, we face the intricate problem of controlling the interaction of several relative constraints. We can avoid this problem by eliminating Procrastinate.

There is a second reason to want to derive Procrastinate from other constraints. Observe that Procrastinate is the exact mirror-image of Pesetsky’s (1989) Earliness principle, which states that movement operations must take place as early as possible. For instance, whereas Procrastinate might explain the difference between Wh-movement in English and Chinese by stating that Wh-movement *can* take place post-Spell-Out in Chinese, hence *must* take place post-Spell-Out, but cannot take place post-Spell-Out in English, hence is bound to occur early, Earliness says that Wh-movement *can*, hence *must* take place early in English, but cannot be early, hence is condemned to occur late in Chinese. We could build a mirror-image of the Chapter 4 model in which Earliness is operative instead of Procrastinate, and in which weak features have the property that they cannot be checked pre-Spell-Out. This implies that Procrastinate as a principle of grammar is underdetermined by the facts. It would be interesting, therefore, if it could be shown that the principle that distinguishes overt from covert movement (Earliness or Procrastinate) *must* be Procrastinate. We achieve this aim if we can derive Procrastinate as an effect of other principles.

In Chapter 4 Chomsky suggests that perhaps, Procrastinate need not be stipulated as an (axiomatic) condition, but may be derivable from other constraints. The idea is roughly this: when a feature moves early, i.e. pre-Spell-Out, the PF-component demands that Pied-Piping takes place: the PF-component cannot interpret “unbound features.” No Pied-Piping is necessary after Spell-Out has occurred. As a consequence, “more material” is moved during overt movement than during covert movement, so that it is natural to suppose that overt movement is more expensive than covert movement; economy then prefers covert movement if possible.

This account is incomplete unless we can determine exactly which economy constraint is involved here. There are two possible views of the matter, depending on the exact properties of the Pied Piping operation.

The first option builds on Chomsky's assumption that overt movement creates (at least) two chains: the chain(s) of the formal features, one of which has been attracted, and the chain of material being Pied-Piped in the process. This option has two disadvantages. First, it implies a complication in the definition of the movement transformation, since it must now be able to create different numbers of chains in a single operation (creating the chain of the Pied Piped material cannot be a separate operation, since this material is not itself attracted). Secondly, in order to derive Procrastinate from this, we must adopt a global, relative economy constraint which counts the number of chains in a derivation. The constraint must be global, as it compares the number of chains resulting from two possible operations that apply at different points in the derivation. This interpretation of Procrastinate takes us further from our goal, in that it provides independent evidence for a global, relative constraint of the type which Chapter 4 seeks to avoid.

A second option is to assume that Attract-F always creates one chain, which optionally consists either of the formal features, or of both the formal features and the pied-piped material. This option allows attract-F to remain relatively simple. But in order to derive Procrastinate from economy we must now assume the existence of an additional global, relative economy constraint on top of those discussed earlier, namely one which measures the "weight" or "amount" of material that is moved by one application of move-F, and compares it to the "amount moved" by a later application of move-F. Chomsky formulates such a condition: "[the attracted feature] F carries along just enough material for convergence." Independent evidence for this condition should come from restrictions on (traditional) Pied Piping, but it rather creates problems in that area, as Pied Piping allows of a fair amount of optionality.

Note, furthermore, that both these options derive Procrastinate from a *global* constraint, so that they are inconsistent with Chomsky's *local* view of Procrastinate as it supposedly applies to *there* insertion (see previous section).

We can avoid all talk of "numbers of chains created by a single application of attract-F," or of "amounts of material moved by Attract-F," and derive Procrastinate from a purely local constraint, along the following lines.

Assume first that Fewest Steps economy (11) needs to consider only those operations that are *potentially* superfluous; i.e., assume the following dichotomy:

Grammatical operations

counted by FS	not counted by FS
Attract	Select
	Merge
	Spell-Out

The intuition underlying this dichotomy is that selecting a certain lexical item from the numeration, merging structure K with some distinct K', and applying Spell-Out to a derivation must each apply once, and cannot apply more than once. An item in the numeration must be selected once, and cannot be selected more than once, etc. Only the movement operation Attract can apply to a given item more than once, and potentially more often than necessary; hence FS considers only Attract.

If we assume that FS applies only to Attract, we immediately derive the correct relative order of operations in a derivation:

Procrastination: Select/Merge | $\text{Attract}_{\text{Strong F}}$ < Spell-out < $\text{Attract}_{\text{Weak F}}$

To illustrate, assume a numeration with a certain number of lexical items with strong and weak features, and consider which operation must be applied at each given point in the derivation. Initially, there is no choice: Select (from the numeration) and Merger are the only options. Only after Select and Merger have applied a number of times, and have created a structure where a head H attracts a feature in its complement, do we have a choice: Merger, or Attract-F.

Because Merger is costless, whereas Movement is costly, local Fewest Steps (as defined in section 3) will always choose Merger at each specific choice point (even if Movement will have to occur sooner or later anyway). Hence, a choice for Movement will be made only when this is necessary for convergence: when H has a strong feature, and merger of HP with a head G will lead to a proscribed structure in which a strong feature is contained in an embedded projection. So, movement for weak feature checking will always be postponed until there simply are no other options left; i.e., until the numeration has been depleted, and Spell-Out has occurred. Any derivation therefore must proceed in the following order: first strong features are checked and lexical items are merged, then Spell-Out is applied, and after that, weak features are checked. In this way, the relative ordering of weak feature checking after strong feature checking and depletion of the numeration falls out as an effect of local economy.[3]

In this section we have resolved the tension between two violable constraints, Fewest Steps and Procrastinate, by deriving one from the other. We have also strengthened the case for local economy, by deriving procrastination effects from the local clause of FS. The next section addresses a third relative constraint that plays a role in Chapter 4.

5. *Case Checking, Fewest Steps and Redundant Features*

Besides Procrastinate and Fewest Steps, Chapter 4 introduces one further global and violable constraint. A numeration underlying a syntactic derivation may not contain features which have no effect on output, i.e. on PF or LF: "α enters the numeration only if it has an effect on output." We will call this the No Redundant Features Constraint (NRF). In the present section we will show that NRF is global and violable (hence to be avoided on general grounds, as indicated above), as well as insufficiently well-defined to be reliably utilized. It will turn out that at least some of the effects attributed to NRF can be deduced, again from Fewest Steps. We will illustrate these issues by means of a single set of examples, involving Case-checking in an overt Accusative-checking language.

A model of grammar which analyses Case-assignment as movement to a Case-checking position must somehow ensure that the correlation between θ-features, Case-features, and overt NP-positions remain intact (the right NP in the right position has the right Case). Thus, well-formed Icelandic (12a) must be allowed but (12b) must be ruled out (for discussion of Case in Icelandic see e.g. Sigurðsson 1989, Jonas & Bobaljik 1993).

---

[3]Deriving Procrastinate in this manner has been considered before (see Zwart 1993), but makes little sense unless Fewest Steps is recognized as a local constraint. Note, that the relative ordering of Spell-Out w.r.t. Select/Merge and Strong feature checking must follow from independent principles: we assume that Spell-Out cannot occur unless the numeration is depleted, and cannot apply to a structure that still contains some strong features. Note further, that we can now do without the global, violable constraint that "F carries along just enough material for convergence."

(12) a. María las bókina. *(Icelandic)*
[$_{Subject}$ Mary-NOM] read [$_{Object}$ the-book-ACC] 'Mary read the book'
b.* Bókina las María
[$_{Subject}$ the-book-ACC] read [$_{Object}$ Mary-NOM] 'The book read Mary'

We start then with an Icelandic numeration with strong DP-features on both T and v, a transitive verb, and two DPs, one (*bókina*) marked with accusative, one (*Mária*) with nominative Case. Initially, we have two options: merge NP_{NOM} in subject position (external θ-role) and NP_{ACC} in object position (internal θ-role), or vice versa. We want the former choice, (12a), to successfully converge, and the latter choice, (12b), to be ruled out.[4]

Let us start with the desired initial derivation:

(13) $[_{vP}\ SU_{NOM}\ [_{v'}\ VB\ [_{VP}\ t_V\ OB_{ACC}\ ]]]$
1 $[_{vP}\ OB\ [_{v'}\ SU_{NOM}\ [_{v'}\ VB\ [_{VP}\ t_V\ t_{OB}\ ]]]]$
2 $[_{T'}\ T\ [_{vP}\ OB\ [_{v'}\ SU_{NOM}\ [_{v'}\ VB\ [_{VP}\ t_V\ t_{OB}\ ]]]]]$
3 $[_{TP}\ SU\ [_{T'}\ T\ [_{vP}\ OB\ [_{v'}\ t_{SU}\ [_{v'}\ VB\ [_{VP}\ t_V\ t_{OB}\ ]]]]]]$

Steps 1 and 2 are forced. Moving SU to outer spec,vP in 1 would have resulted in a Case-mismatch, so OB moves to check the strong features on v. Subsequent structure-building derives step 2. Moving SU (step 3) in order to check strong DP features of Tense correctly derives (12a) (after V2).

Note, incidentally, that the last step could have moved OB to Spec,TP instead, incorrectly deriving a structure with the object in Spec,TP. But this is ruled out by the global clause of Fewest Steps, since it requires that FF(SU) move covertly to check NOM. Thus, this derivation provides independent evidence for global FS. [5]

Consider now the unwanted derivation of (12b) in (14):

(14) $[_{vP}\ SU_{ACC}\ [_{v'}\ VB\ [_{VP}\ t_V\ OB_{NOM}\ ]]]$
1 $[_{vP}\ SU\ [_{v'}\ t_{SU}\ [_{v'}\ VB\ [_{VP}\ t_V\ OB_{NOM}\ ]]]]$
2 $[_{T'}\ T\ [_{vP}\ SU\ [_{v'}\ t_{SU}\ [_{v'}\ VB\ [_{VP}\ t_V\ OB_{NOM}\ ]]]]]$
3 $[_{TP}\ SU\ [_{T'}\ T\ [_{vP}\ t_{SU}\ [_{v'}\ t_{SU}\ [_{v'}\ VB\ [_{VP}\ t_V\ OB_{NOM}\ ]]]]]]$
4 $[_{TP}\ SU\ [_{T'}\ [_T\ FF(OB)\ [_T\ VB\ T\ ]]\ [_{vP}\ t_{SU}\ [_{v'}\ t_{VB}\ [_{VP}\ t_V\ OB{+}t_{FF(OB)}\ ]]]]]$

In step 1 we must move SU to outer Spec,vP, checking ACC and strong DP-features on v (moving OB instead would lead to a mismatch of Case-features, hence terminate the derivation). Subsequent structure-building derives step 2. Now, strong DP-features on Tense must be checked, and MLC only allows SU to move (step 3). Covertly, FF(OB) move to Tense and NOM is checked. This derivation converges; all uninterpretable features have been checked. The result is ill-formed (12b).

---

[4]Note, that we are comparing derivations with different meanings. This is not a problem since we assume, with Chomsky 1995, that Reference Sets are determined solely on the basis of numerations. We ignore the semantic oddity of (12b), which can easily be fixed; what is relevant is that (12b) has an accusative subject.

[5]Chomsky suggests that Fewest Steps works locally in preventing this derivation, but, as we have defined “local constraints,” this is incorrect (cf. footnote 1). In step 3 movement must occur. Locally, moving OB is just as expensive as moving SU. Only a global Fewest Steps will consider it relevant that moving OB will eventually lead to an extra movement operation.

Chapter 4 attributes the ill-formedness of the unwanted derivation resulting in (12b) to the fact that, given that derivation, the numeration that it proceeds from is too expensive, as it contains a superfluous strong DP feature on v. For without that feature we would have arrived at the same PF and LF (with SU moving from Spec,vP to Spec,T in one swoop, without an intermediary landing in outer Spec,vP as in step 1). Let us consider this analysis more closely.

First of all, we see that NRF is not a local condition. Whether a feature will have an effect on output cannot be judged during the composition of the numeration. For whether a feature will have an effect depends on other choices that are made during the derivation. The intended effect of NRF then is to preclude those choices that deprive a feature of the chance of having an effect on output. In the present example, the strong DP-feature on v is supposedly superfluous, but it does play a role in the well-formed derivation (13) which proceeds from the same numeration (without it, OB remains in situ, to the right of VP-boundary adverbials). It is only when the accusative DP is merged as external argument that the feature becomes a superfluous component of the numeration in hindsight. So NRF is not a local constraint. Also, it is a relative constraint; a feature can only be judged to be superfluous in the presence of alternative converging derivations that can do without it.

But the conceptual problems are actually far more serious. We must wonder through what comparison of derivations NRF can produce the effect that Chomsky attributes to it. Consider, in the abstract, a condition C which states that operation O may be applied only if O has an effect on output. How can we ascertain the existence of condition C? We can only ascertain the existence of C by observing that C prevents O from applying in certain cases. And we can only observe whether O has been prevented from applying, in cases where application of O will produce a different output than non-application of O. But in those exact cases, in which application of O yields a different output than non-application of O, C will not prevent O from applying. Hence we can never observe a case in which C has prevented O from applying, for in those cases application or non-application of O is unobservable. C therefore makes no predictions which can be tested from observation: a grammar that contains C yields the exact same PF-LF pairs as a grammar that does not contain C.

Conditions of this type are meaningless, then, unless we adjust them along the following lines. Reformulate condition C as follows: "operation O may be applied only if O has a class A output effect." Now suppose that O can have class A and class B output effects. We can then ascertain the existence of condition C by observing that a class B effect of O occurs only in combination with a class A effect (whereas a class A effect can occur without a class B effect).[6]

Example (12b), e.g. cannot be ruled out by NRF without such a class A-B distinction. NRF might rule out its derivation in (14), but only if output (12b) can be obtained from some sparser numeration — a self-defeating strategy.

Let us see what happens if the numeration for (12b) has no strong DP-feature on v. Given this numeration (no EPP on v, otherwise the same as for (13), (14)), one option is to merge NP_{ACC} as OB and NP_{NOM} as SU. But this will never lead to (12b), as the distribution of θ-roles over NP_{ACC} and NP_{NOM} will then be different.

Another option is to start as in (14); the derivation then proceeds as follows:

(15) 1 $[_{T'}$ T $[_{vP}$ SU_{ACC} $[_{v'}$ VB $[_{VP}$ t_V OB_{NOM}]]]]
2 $[_{TP}$ SU_{ACC} $[_{T'}$ T $[_{vP}$ t_{SU} $[_{v'}$ VB $[_{VP}$ t_V OB_{NOM}]]]]]

---

[6] A complex variety of such a class distinction plays a role in the work of Fox that underlies Chomsky's treatment of the QUANT feature (see Fox 1993).

Step 2 is forced by MLC; the derivation terminates due to mismatch of Case features (ACC on SU against NOM on T). The exact outcome now depends on what is meant by "termination." Either there are no LF and PF at all (in which case we have not managed to derive the same output as in (14)); or (15) step 2 *is* the output, but with a "*" because of termination. In the latter case (15) differs from (14) in two respects: (15) has a "*", and (15) has unchecked Case-features which (14) does not have, as well as having fewer chains. If the first difference were the only difference, this would make for an interesting and acceptable version of a ("class B") "insufficient difference." We could then understand NRF as follows: a converging derivation is ruled out, if from the same numeration minus one feature we can derive an otherwise identical PF-LF pair, that differs at most in that it has a "*". But, as noted, there are other differences between (14) and (15) which would also have to be considered insufficient, and it is not clear how this can be justified. In particular, (15) has an uninterpretable unchecked NOM-feature on OB: we would have to consider this distinction between (14) and (15) insufficient. That is, we must assume that the EPP-feature on v̲ can be "redundant" even if it is necessary for convergence.

We can complicate matters further. We have assumed so far, that NRF rules out a derivation from numeration N_1 with feature F, if (virtually) the same PF-LF pair can be obtained from numeration N_2 which differs from N_1 *only* in that it lacks F. But we might also take into consideration other numerations that differ more radically from N_1: this might lead to a different theory of class A / class B properties of output. In our example, an obvious possibility to consider is that we could compare derivation (14) and its numeration N_1 (with EPP on v̲) with the derivation from a numeration N_2 without EPP on v̲, in which the NP that bears ACC in N_1, bears NOM Case, and vice versa. We can then derive an alternative for (14b), i.e. for (12b), in which the same NPs bear the same θ-roles, but in which the subject bears NOM and the object bears ACC. Since v̲ in N_2 lacks EPP, this would lead to a sentence without overt object movement, hence indeed to (12b) as far as word order is concerned. The PF, however, would be different, because morphological overt Case marking would be distributed differently over the two NPs. After all, the incorrect distribution of overt Case marking is the reason we want to exclude (12b) in the first place. Now, if we ignore this distinction (as a class B effect) then we indeed manage to exclude (12b). Paradoxically, we then manage to explain the Icelandic intuition that Case marking is wrong in (12b) by stipulating that Case marking is the very effect that is ignored when we judge the derivation for (12b) in (14) under NRF. Also, we then call upon a much wider, relative comparison of derivations, since we can no longer use numerations (or simple functions of them) to determine Reference Sets for NRF.

In any case, it should be clear that NRF cannot be said to be well-defined without there being some theory of "class A" (sufficient) and "class B" (insufficient) distinctions among PF-LF pairs, and that the properties of such a theory are not immediately evident. Global Fewest Steps, on the other hand (which by now has considerable independent motivation) straightforwardly explains the distinction in (12): the desired derivation in (13) requires two cases of DP-movement, whereas the undesired derivation resulting in (14) requires three (one covert movement more). Whether all distinctions presumed to follow from NRF can be attributed to Fewest Steps remains to be established, however.[7]

---

[7]Chomsky uses NRF in one other context: to prevent merger of +Wh[+strong] COMP after Spell-Out in English. Here, it appears that NRF *can* apply non-vacuously, because [+strong] is superfluous, but also obligatory. There is no alternative derivation without [+strong], so that pre-Spell-Out merger and overt Wh-movement are the only option. However, here the applicability of NRF is provable only metalinguistically, not within the

In conclusion. While Chapter 4 abolishes most relative constraints and replaces them with absolute properties of Attract, at least three violable constraints remain: Procrastinate, “No Redundant Features,” and Fewest Steps. Procrastinate is held responsible for the distinction between overt and covert movement, and for the *there*-insertion data discussed in section 3. I have proposed that Procrastinate can be done away with, if we assume that Chomsky’s global Fewest Steps is also partly local. Such a constraint can account for *there*-insertion just as well, and for *it*-insertion even better (section 3); it can also be held responsible for the overt/covert distinction (section 4). There is empirical gain here, as well as conceptual gain on three points. Two constraints have been replaced by one. Our analysis of procrastination effects strengthens the empirical basis for a local conception of economy, in that Procrastinate is derived from the local clause of Fewest Steps. And by deriving Procrastinate from other conditions we provide evidence for the assumption that the overt/covert distinction must be viewed in terms of procrastination, rather than earliness. In fact, procrastination is now explained on the basis of the fact that derivations are directional, and proceed “from numeration to LF.” Finally, in the last section, we have shown that there is some promise that Fewest Steps can also replace NRF, a constraint with questionable properties in various respects. While we started out with three violable constraints, whose interaction was undefined, we are now left with just one violable, partly local constraint, and it appears that there may be no need for a theory of constraint interaction in syntax.

6. *References*

Chomsky, N. 1991. Some Notes on Economy of Derivation and Representation. In *Principles and Parameters in Comparative Grammar*, ed. R. Freidin, 417-454. Cambridge, Mass: MIT Press.

Chomsky, N. 1993. A minimalist program for linguistic theory. In *The View from Building 20*, ed. K. Hale & S.J. Keyser, 1-52. Cambridge, Mass: MIT Press.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge, Mass: MIT Press.

Epstein, S.D. 1992. Derivational Constraints on A’-Chain Formation. *Linguistic Inquiry* 23:235-259.

Fox, D. 1993. Quantifier Scope in VP ellipsis. Ms MIT.

Jonas, D. and J. Bobaljik. 1993. Specs for Subjects: The Role of TP in Icelandic. In *MIT Working Papers in Linguistics 18: Papers on Case and Agreement I*, 59-98. Department of Linguistics and Philosophy, MIT, Cambridge, Mass.

Kitahara, H. 1995. Target alpha: Deducing Strict Cyclicity from Derivational Economy. *Linguistic Inquiry* 26:47-77.

McCarthy, J. and A. Prince. 1994. *The Emergence of the Unmarked*. Ms. U. Mass/Rutgers.

Pesetsky, D. 1989. Language Particular Processes and the Earliness Principle. Paper presented at *GLOW*, Utrecht, The Netherlands.

Reinhart, T. 1993. Wh-in-situ in the framework of the Minimalist Program. Ms. UU.

Sigurðsson, H.A. 1989. Verbal Syntax and Case in Icelandic. Diss. University of Lund.

Zwart, C. J.-W. 1993, *Dutch Syntax: a Minimalist Approach*. Diss., University of Groningen.

E.G. Ruys
Utrecht Institute of Linguistics OTS / Utrecht University
Trans 10
NL-3512 JK Utrecht, The Netherlands
eddy.ruys@let.uu.nl

---

system, as there is no derivation without [+strong] that we can use for comparison. See note 3 for our alternative.