'Shortest Move' versus 'Fewest Steps'

C. Jan-Wouter Zwart

I will have to put aside a careful development here, but it is intuitively clear how certain basic aspects will enter. Take the phenomena of Superiority (..) and of Relativized Minimality (..). Looking at these phenomena in terms of economy considerations, it is clear that in all the "bad" cases, some element has failed to make "the shortest move". (Chomsky 1993:14)

0. Introduction

In much recent work in generative grammar, derivations are considered to be subject to principles of economy. In Chomsky (1993), economy of derivation is implemented in at least two ways: derivations should involve the shortest possible movements and the fewest possible steps. As Chomsky notes, these two requirements appear to be contradictory. He proposes a *Form Chain* mechanism to resolve the contradiction.

In this paper, I will argue that the Form Chain mechanism should be described as follows. Form Chain consists of two processes, one of which generates empty wh-elements in the intermediate specifier positions required for chain formation over a longer distance. The other process consists in movement *across* these intermediate wh-elements, in compliance with the fewest steps requirement, but in violation of the shortest move requirement. This implies that the shortest move requirement does not apply in long distance movement processes.

In the remainder of this paper, I will argue that the shortest move requirement is equally redundant in the analysis of the other phenomena that seem to call for it, head movement and superraising. My conclusion will be that economy of derivation consists in the single requirement that the number of movement steps should be as small as possible.

1. Economy

Let us start by assuming that subjects are generated inside VP, the VP constituting a kind of kernel sentence consisting of a subject, a verb, and a complement of the verb (an object or a subject-predicate combination (referred to as *Small Clause*)). Let us also assume that the elements of a kernel sentence have to be formally licensed in well-defined syntactic configurations, called *functional projections*. Syntax then consists in moving the elements of a kernel sentence to designated positions in functional projections. The output of the syntactic component is an interface representation serving as the input to other cognitive systems, such as those involved in speech processing and interpretation.

As a minimalist principle, Chomsky (1993) assumes that the interface representations should be pure and simple, stripped of all features that are not relevant to the cognitive systems they provide input for. This he calls *economy of representation*, summarized in (1):[1]

(1) *Economy of representation*:
Use as few symbols as possible in the output of a derivation

[1] Economy of representation ultimately reduces to the principle of *Full Interpretation*, formulated in (36) below.

In addition to (1), Chomsky proposes a second minimalist principle, stating that interface representations should be arrived at in the most economical way. This paper discusses the proper formulation of this second principle, called *economy of derivation*. I will argue for the following formulation:

(2) *Economy of derivation*:
Use as few steps as possible in deriving an output representation

(2) is a standard feature of the minimalist program of Chomsky (1993). Chomsky argues that derivations are governed by principles summarized here under the label *inertness*:

(3) *Procrastinate*:
Move as late as possible

(4) *Greed*:
Move α only if movement contributes to licensing of α.

(3) and (4) can be grouped together as in (5):[2]

(5) *Inertness*
Move as little as possible

(5) and (2) are equivalent.

The formulation of economy of derivation in (2) is more interesting for what it leaves out than for what it contains. In particular, (2) makes no reference to the *length* of the steps involved in a derivation. According to conventional wisdom, short steps are more economical than long steps. Thus, it has been proposed that economy of derivation contains (6) in addition to (2):

(6) *Economy of derivation part 2*:
In deriving a representation, make the shortest possible movements

(6) underlies the concept of *minimality* (Chomsky 1986b, Rizzi 1990), paraphrased in (7):

(7) *Minimality*:
Don't move α across a place where α could have landed

(6) also plays a major part in Chomsky (1993).

2. Equidistance

To illustrate the workings of (6) in Chomsky (1993), consider the derivation of a simple sentence consisting of a subject, a verb, and an object:

(8) John loves Mary

[2] The principle *Procrastinate* reduces to *Inertness*, if the latter is considered to apply at each point in the derivation.

Let us assume that all elements in (8) have been moved out of their kernel sentence positions to their licensing positions.[3] Let us also assume that the architecture of the functional domain is as in (9), where subjects (SUB) are licensed in the spec position in AgrSP, objects (OB) in the spec position in AgrOP, and the finite verb (V) in the head position in TP, and where VP equals the kernel sentence:

(9) $[_{AgrSP}$ spec AgrS $[_{TP}$ spec T $[_{AgrOP}$ spec AgrO $[_{VP}$ SUB [V OB]]]]]

In the syntactic derivation of (8), the object has to move from the position indicated by OB in (9) to the spec position in AgrOP. In doing so, the object crosses the position indicated by SUB in (9). Assuming this to be a position where the object could have landed, movement of OB to the spec position in AgrOP violates the shortest steps requirement (6).

A strict application of (6), then, makes all object movement impossible in the presence of a subject. To remedy this problem, Chomsky proposes that overt movement of the verb to the head position of AgrOP makes the position indicated by SUB in (9) and the spec position of AgrOP 'equidistant' from the position indicated by OB. Spec,AgrOP and SUB being equidistant from OB, movement of the object to Spec,AgrOP is in keeping with (6). Verb movement, then, is the way to make the VP transparent.

As Chomsky notes, this analysis makes the prediction that overt object movement never occurs without overt verb movement to the head of AgrOP. But this is clearly wrong for languages like German and Dutch, as can be concluded from Vikner 1991, section 4.2.5. In German and Dutch, objects always leave the VP, which must be described as movement to Spec,AgrOP in a minimalist approach (Vanden Wyngaerd 1989, Zwart 1993b).[4] Since the object and the verb are not adjacent, the verb cannot have moved to the head of AgrOP in these constructions:

(10)	..daß	Peter	das Buch	gestern	nicht	gekauft hat
	that	Peter	the book	yesterday	not	bought has

The only way to align the Germanic object movement facts with Chomsky's equidistance principle is to assume that the head of AgrOP is located to the right of the VP, an option excluded in minimalist work such as Zwart 1994, 1993a, 1993b, and Kayne 1994.

The generalization that object movement is dependent on verb movement is based on the position of objects in the Scandinavian languages (Holmberg 1986). For example, in Icelandic the object moves to the right of the negative adverb *ekki*, but not if the verb selecting the object is a participle:

(11)	a.		..að	Jón	keypti	bókina	ekki	
			that	John	bought	the book	not	
	b.	*	..að	Jón	hefur	bókina	ekki	keypt
			that	John	has	the book	not	bought-PART

It is assumed that the finite verb *keypti* in (11a) moves to the left, opening up the possibility for the object *bókina* to move as well, whereas the participle *keypt* in (11b) remains in its basic position inside VP, forcing the object to remain inside the VP as well. The equidistance principle

[3] It is irrelevant whether this is the correct analysis of English sentences.

[4] The argument in the texts abstracts away from possible additional A'-movement of the object, as argued by Vanden Wyngaerd (1989) and Mahajan (1990) to account for the observation that object movement in Germanic displays both A-movement properties and A'-movement properties.

of Chomsky (1993) captures this state of affairs in a straightforward way. Without verb movement, the position indicated by SUB in (9) and the Spec,AgrOP position would not be equidistant to the object, blocking object movement by virtue of the shortest move condition.

However, this empirical argument in support of the shortest move condition faces a conceptual problem. According to the Procrastinate principle (3), covert movement (i.e., after the Spell-Out point in the derivation) is preferred over overt movement. Chomsky (1993) describes violations of Procrastination in terms of the strength of the features represented in the functional heads. If these are *strong*, they are visible (but not interpretable) at the representation interfacing with the acoustic-perceptual component of the cognitive system (PF), and must be eliminated before the derivation reaches that stage. Therefore, strong features trigger overt movement, in violation of Procrastinate. The movement is nevertheless inevitable, as without it, economy of representation (1) would be violated.

(11a) suggests that the feature represented in AgrO triggering object movement must be characterized as strong. If so, the absence of object movement in (11b) should lead to a violation of economy of representation, as the relevant features in AgrO are not eliminated in overt syntax.[5]

We must therefore conclude that the equidistance principle is problematic. In what follows, I will argue that the shortest move requirement is not a part of economy of derivation. Since the equidistance principle was prompted by the shortest move requirement, abolishing the latter removes the need for the former, thereby avoiding the problems it poses.

3. Shortest Move Phenomena

There are several phenomena for which the shortest move requirement appears to be relevant. These include head movement, superraising, and wh-movement. I will discuss these phenomena one by one, arguing that the shortest move requirement is irrelevant for the analysis of their properties.[6]

3.1 Head Movement

The history of the *Head Movement Constraint* (Travis 1984:131) is instructive for our purpose:

(12) *Head Movement Constraint*:
An X° may only move into the Y° which properly governs it

This constraint bars the derivation of (13a) from (13b), where *kiss* and *will* are both X°s (heads), and one of these heads must move to C:

[5] See Zwart (1994b) for more extensive discussion of this problem. A way out might be to stipulate that features can be optionally strong. The fact that the object in (11a) may appear both to the right and to the left of the negative adverb *ekki* could be taken to support this move. However, adverbial elements in Germanic in general do not appear to occupy a single fixed position. It is not clear that the variation in the order of object and (negative) adverb in (11a) cannot be described in terms of freedom of placement of the adverb, rather than in terms of optional movement. More generally, it is not clear that allowing features to be optionally strong is necessarily a part of the minimalist approach.

[6] Chomsky (1993:14) also mentions superiority phenomena as illustrating the need for a shortest move requirement. I will refrain from discussing these here, partly because phenomena from ither languages (e.g. Dutch) suggest that the shortest move condition is not an inviolable condition in the domain of superiority phenomena.

(13) a. * Who have John will kissed?
b. Who C [John will [have [kissed]]]
c. Who will John have kissed?

As (13c) shows, only the higher auxiliary *will* may move to C.

From the outset it has been clear that the Head Movement Constraint is part of a principle with a larger scope, the *Empty Category Principle* (see Travis 1984:133, Chomsky 1986b):

(14) *Empty Category Principle*:
Empty categories must be properly governed

A trace is *properly governed* if it is governed by its antecedent.[7] In (13b), movement of *have* to C would leave a trace which is not properly governed by its antecedent (*have*), since the trace would be separated from its antecedent by two maximal projections functioning as *barriers* for government (indicated by the brackets in 13b).[8]

Chomsky (1991) argues that the Head Movement Constraint, if it is reducible to the Empty Category Principle, can be dismissed "as a descriptive artifact, valid only insofar as it does in fact reduce to the ECP." Importantly, Chomsky (1991:429), continuing where Chomsky (1986b:88) left off, considers the ECP a *condition on chains*.

This suggests that in (13) the process of *deriving* (13a) from (13b) is legitimate, whereas the resulting *representation* is illegitimate. The phenomena associated with the Head Movement Constraint thus traditionally fall in the domain of conditions on representations.

In Chomsky (1993), government does not play a role, and hence conditions on head movement cannot be reduced to (12) or (14). It appears that here, the shortest move requirement of economy of derivation becomes crucial. It takes a longer step to derive (13a) from (13b) than it does to derive (13c) from (13b).

However, closer scrutiny indicates that (13a) can be excluded in a more minimalist way. The question that has to be asked first is: What is the trigger for verb movement to C in wh-constructions? If this trigger does not apply to infinitives, (13a) will never be derived because of the inertness principle.

There is ample evidence that verb movement to C in Germanic is closely linked to tense. Consider the following facts from Dutch:

(15) Koopt Jan een huis?
buys John a house
'Is John buying a house?'

(16) a. Jan een huis kopen?
John a house buy-INF
'John buy a house?'
b. * Kopen Jan een huis?
buy-INF John a house

Assuming that the structure of yes/no questions matches that of wh-questions, (15) and (16) are comparable to (13). We may consider the counterpart to the wh-word in (13) to be empty in (15) and (16). This suggests that the verb movement in (15), as in (13c), targets C. As can be seen in (16), such verb movement takes place only when the verb is finite.

[7] I ignore θ-government for expository purposes, cf. Chomsky 1986b:69.

[8] This is the analysis in Chomsky 1986b, Ch. 11.

In terms of Chomsky (1993), we may suppose that C hosts a tense feature, comparable to the V-features of AgrS etc., which must be checked by moving T(ense) to C (cf. Wilder and Ćavar 1993). This triggers movement of the finite auxiliary in (13).

The proper test case for the Head Movement Constraint, then, contains a choice of two finite verbs, only the higher of which may be moved to C:

(17) a. John did not think I could help someone
b. Who did John t not think I could help?
c. * Who could John did not think I t help?

(17c) is correctly excluded by the Head Movement Constraint and the shortest move requirement. However, (17c) can also be excluded on standard minimalist assumptions, once we assume that C hosts a tense feature, to be eliminated by T.

T itself hosts features which must be eliminated by the verb, and conversely, the tense features of the verb must be checked by the corresponding features in T. So, V and T are mutually dependent, as are T and C.

The sentences in (17) contain two finite clauses, each containing a C, a T, and a finite V. In (17a) and (17b), the finite verb *did* links up with the matrix T, and, ultimately, with the matrix C. The finite verb *could* links up with the embedded T and C. In (17c), however, the embedded verb *could* links up with the matrix C. There are various derivations possible to yield this result, but all of these derivations have in common that the proper linking up of V, T, and C is not established. For instance, *could* could move to the embedded T first and then on to the matrix C. This would have the result that the embedded C is not linked up with T, and that the matrix verb *did* is not linked up with the matrix T and C. Or *could* could skip the embedded T and move to the matrix C via the matrix T, but this would again rob the matrix verb *did* of its licensing position T, and it would leave the tense feature in the embedded C unchecked. Finally, movement of *could* via the embedded C would be blocked, as *could* and the embedded T would have all their features checked in C, and there would be no trigger for further movement (by the principle of Greed (4)).

Hence, the derivation yielding (13c) would always leave certain features unchecked, ultimately leading to a violation of economy of representation (1).

More generally, the observation that heads move stepwise rather than long distance is explained by the feature checking requirements of the minimalist approach. If a functional head Y contains a feature to be checked by the verb X, and X moves to a functional head Z, crossing Y, the feature in Y will remain unchecked, and the derivation will not converge. This covers the core cases of the Head Movement Constraint. Hence, for deriving stepwise head movement, no special constraint needs to be formulated.[9]

Taking (13) and (17) to be representative of Head Movement Constraint phenomena, the minimalist analysis of them appears to make the shortest move requirement redundant for the proper understanding of this class of facts.

3.2 Superraising

The shortest move requirement likewise appears to play a role in barring *superraising* phenomena:

[9] The reduction of the Head Movement Constraint to general requirements of feature checking predicts that if no features in a head Y need to be checked by the verb X, X may move to Z crossing Y. In Zwart (1993b) I discuss several of these cases (see also Ouhalla 1989).

(18) a. *John seems is likely to win
b. *John seems it is likely to win

The sentences in (18) are derived from more basic representations in which *John* is the subject of *win*, generated inside the VP as previously assumed. As (19) shows, the subject position (Spec,AgrS) of the embedded clause is a legitimate target for subject movement:

(19) It seems John is likely to win

It seems, then, that the sentences in (18) are derived by moving *John* across a legitimate target for subject movement, in violation of the shortest move requirement of economy of derivation (Chomsky 1993:14).

However, it is immediately obvious that (18a), at least, is excluded on standard minimalist assumptions of movement and feature checking. As discussed in section 2, all movement operations are assumed to be triggered by feature checking requirements. Let us assume that in (19), the subject, *John*, moves to the specifier position of AgrS. This creates the proper configuration for checking the features of AgrS against the corresponding features of the subject. As a result, the features of AgrS and the corresponding features on the subject are eliminated.

Consider the consequences for (18a). Suppose (18a) is derived from an intermediate representation corresponding to (19), (18a'):

(18a') seems John is likely to win

At this point, the subject's features will be eliminated as a result of the checking operation taking place in Spec,AgrSP in the embedded clause, just like in (19). This will make it impossible for (18a) to be derived from (18a'): the subject has lost its features, and cannot serve to check the features of the AgrS in the matrix clause. This derivation of (18a) is excluded.

Another possible derivation of (18a) moves the subject in one swoop from the VP-internal position to the specifier position of the matrix AgrSP. This creates the proper configuration for checking the features of the matrix AgrSP. However, it leaves the features of the embedded AgrSP unchecked. This derivation, therefore, is also excluded.

A final possibility would be to move the subject in one swoop from the VP-internal position to the specifier position of the matrix AgrSP, and to insert an expletive *it* in the specifier position of the embedded AgrSP, in order to check the latter's features. As this would yield (18b), this derivation must be excluded as well. This derivation assumes that *it* is a dummy subject that can be inserted when the derivation requires it. However, it is questionable whether *it* has this dummy status.

Bennis (1986) argues convincingly that expletives of the type of *it* are not dummy subjects, but part of the complement domain of the verb *seem* (see also Moro 1993). If so, *it* can never be simply *inserted* in the specifier position of the embedded AgrSP in (18b). If it ends up in that position, it must have *raised* just like *John* in (19). The ungrammaticality of (18b), then, shows that *it* can only raise to the matrix AgrSP, and *John* can only raise to the embedded AgrSP. How can this be explained?

I would like to propose here that raising verbs like *seem* select a propositional complement (Small Clause, including AgrP), not a sentential complement (CP). Consequently, if a CP appears in the complement domain of a raising verb, it must be either the subject or the predicate of a Small Clause (see Moro 1993). Considering that Small Clauses generally do not have sentential subjects, the starting hypothesis would be that a CP in the complement domain of a raising verb is a Small Clause predicate, needing a subject. My proposal is that *it* is the unique subject that can be associated with the CP predicate.

According to this proposal, (19) must have the structure in (20), with *it* generated as the subject of the complement of the higher raising verb:

(20) $[_{AgrSP1}$ AgrS1 $[_{VP1}$ V1 [*it* $[_{CP}$ $[_{AgrSP2}$ AgrS2 $[_{VP2}$ V2 [*John (to) win*]]]]]]]

It is easy to see that starting from the structure in (20), we will never be able to derive (18b). *John* cannot move to the specifier position of AgrSP1, because this would leave the features of AgrS2 unchecked. *It* is not available for checking the features of AgrS2, because *it* is generated in the matrix clause (and lowering is not allowed). The only correct outcome, then, is (19).

The proposal made here implies that (21) is not a legitimate structure:

(21) $[_{AgrSP1}$ AgrS1 $[_{VP1}$ V1 $[_{CP}$ $[_{AgrSP2}$ AgrS2 $[_{VP2}$ V2 [*it* [*John (to) win*]]]]]]]

In (21), V1 does not have a Small Clause/AgrP complement, but a direct sentential complement CP. This, we have assumed, is not allowed.[10]

What evidence do we have that raising verbs take propositional complements only? Consider first an empirical argument.

Take the Dutch raising verb *schijnen* 'seem, appear'. *Schijnen* appears in standard raising constructions, illustrated in (22):

(22) a. Het schijnt dat Jan ziek is
it seems that John sick is
b. Jan schijnt ziek te zijn
John seems sick to be
c. Jan schijnt ziek
John seems sick

In (22c), *schijnen* takes a propositional complement *[Jan ziek]*, in (22b) an infinitival complement *[e te zijn [Jan ziek]]*. (22a) is the expletive construction. Crucially, *schijnen* cannot take a single noun phrase argument. Thus, the Dutch counterpart of *John appeared* is (23a), not (23b):

(23) a. Jan ver-scheen
John appeared
b. * Jan scheen

The prefix *ver-* has been analyzed as an incorporated predicate by Mulder (1992). If this is correct, *schijnen* in (23a) again takes a propositional complement *[Jan ver]*.[11]

[10] In Zwart (1993b), I explained the illegitimacy of (21) by assuming that *it* cannot be combined with a nonfinite clause. This appears to be a correct generalization, but it would not exclude all possible cases of superraising. For example, we could leave out *it* in (21), and turn the verb *win* into a passive verb. A possible outcome of that structure would be the ungrammatical *The race seems John appears to be won*. This cannot be explained by conditions on the distribution of *it*, but it can on the assumption that CP cannot be a direct complement of a raising verb.

[11] In (23b), the only permissible interpretation is that John is an element that emits light (e.g. a celestial body). In that case, *schijnen* is not a raising verb, as indicated by standard tests. For example, *schijnen* in the sense of 'emitting light' takes a *have*-auxiliary instead of a *be*-auxiliary.

(22b,c) and (23), then, indicate that *schijnen* takes a propositional complement. The only exception would be (22a), unless we adopt the proposal advanced here, according to which *schijnen* in (22a) would take a complement *[het [dat Jan ziek is]]*.[12]

A second observation supporting the analysis proposed here, is that CP complements are *objects*, whereas raising verbs (of the type of *be* and *seem*) select *states* (i.e. subject predicate combinations). Assuming categorial selection to be a function of semantic selection (Grimshaw 1981, Pesetsky 1982 and 1994, Chomsky 1986a), this makes it impossible for raising verbs to directly select a CP complement.

I assume Small Clauses (i.e. subject predicate combinations) to be the canonical structural realization of states. This would make the Small Clause the prototypical raising complement. In the absence of evidence to the contrary, we may conclude that in typical raising constructions like (22a) the prototypical raising complement is realized. This implies that in (22a), too, the complement of *schijnen* is a Small Clause. The analysis of 'expletive' *het* as a Small Clause subject captures this.[13]

In other words, superraising does not result from a violation of the shortest move requirement, but from a violation of the selection requirements of raising verbs (in addition to the violation of Greed discussed in connection with (18a)).

Other superraising constructions can be explained along similar lines. Consider (24):[14]

(24) * $John_i$ seems [it is likely [that we assured them [t_i to win the race]]]

This sentence is excluded because the raising predicates *seems* and *likely* cannot take a CP complement directly. The CP would have to be part of a Small Clause with *it* as its subject. In (24), this condition is met with *likely*, but not with *seems*. The only possible output, therefore, would involve two instances of *it*, but then there would be no position left in which *John* could be licensed:

(25) * It_i seems t_i [it_j is likely t_j [that we assured them [John to win the race]]]

Consider also (26):

(26) * $John_i$ seems [that it was told t_i [that he would win]]

Again, *seems* lacks a Small Clause subject in its complement domain. Inserting *it* there would again rob *John* of its licensing position:

(27) * It_i seems t_i [that it_j was told John t_j [that he would win]]

This explanation for the nonexistence of superraising, introduced to account for the ungrammaticality of (18b), in fact also extends to the cases of superraising illustrated by (18a).

[12] The prefix *ver-* does not appear in the sentences in (22). This might be taken to indicate that in (22), the predicate position is taken by other predicates. This is straightforwardly true in (22b) and (22c), but in (22a) again only if *schijnen* takes a propositional complement in (22a) as well.

[13] I have been assuming that infinitival complements are structurally comparable to Small Clause complements (i.e. Small Clause complements of the category AgrP). It is not clear to me at this point that infinitival complements should be characterized as states also. I would like to postpone discussion of this issue until after further study.

[14] Brought to my attention by Phil Branigan (p.c.).

If *seems* cannot have a CP in its complement without also having *it* in its complement, (18a) will never be generated.

In sum, the superraising constructions in (18) can both be excluded without taking recourse to the shortest move requirement of economy of derivation.

3.3 Wh-Islands

Wh-island facts present a third class of phenomena which have been presumed to motivate a shortest move requirement.

(28) *What did he wonder where John put t?

In (28), *what* moves out of an embedded interrogative clause, crossing the position occupied by *where*. This position (the specifier position of the embedded CP) being a potential landing site for *what*, minimality (7) is violated.

The analysis of wh-movement as being *successive cyclic*, first put forward in Chomsky (1973), had a major impact on grammatical theory. It is probably correct to state that the desirability of having short steps in a derivation goes back to this pioneering work.

However, the shortest steps requirement goes much further than the successive cyclic movement requirement. Successive cyclic movement involves employing intermediate landing sites on the edge of a local domain (in the core cases, CP). But movement to the edge of the local domain (i.e., to Spec,CP) is not the shortest movement imaginable. The possibility of moving via an adjunction position to VP has been explored in Chomsky (1986b), for instance.

The discussion of how many intermediate landing sites are involved in wh-movement does not center around the length of the steps, but around the opacity of a certain local domain. There has to be an intermediate landing site between the boundaries of two opaque local domains.

As Koster (1987) argues at length, locality conditions on movement can be regarded as a subcase of the more general locality conditions on dependency relations. Movement creates a dependency relation between an antecedent and a trace, which therefore have to be in the same local domain. Intermediate elements may serve to link an antecedent and a trace which are not in the same local domain.

From this point of view, intermediate landing sites are not needed to comply with conditions on derivation, but forced by conditions on interpretation.

There are three additional reasons to be skeptical about the relevance of the shortest move requirement in the domain of wh-movement.

First, as Chomsky (1993:15) notes, the shortest move requirement appears to be incompatible with the fewest steps requirement. The number of steps increases with the shortness of the movements. The fewest steps requirement appears to be an essential ingredient of the minimalist program. Procrastinate (3) and Greed (4) are both instantiations of the fewest steps requirement (2). Therefore, if the incompatibility cannot be resolved, it seems that the shortest move requirement must be abandoned.

Chomsky (1993:15), noting this problem, states that successive cyclic movement does not consist in a succession of identical adjunction operations but in a single operation *Form Chain*. This operation turns a representation like (29a) into, for instance, (29b), without the intermediate step that yields (29c):

(29) a. e [you think [e [you love who]]]
b. who [you think [t [you love t]]]
c. e [you think [who [you love t]]]

It would take two steps to get from (29a) to (29b) via (29c), whereas Form Chain derives (29b) from (29a) in one step.

This proposal raises the question whether Form Chain is a new type of structure building process, or a combination of well-known structure building processes. Chomsky (1993) distinguishes two structure building processes, each operating in a bottom-up fashion (both referred to as *generalized transformation*). The first structure building process combines two independent phrase markers. The second structure building process adjoins to a phrase marker K a proper subpart of K. Insertion involves the first process, movement involves the second process.

Form Chain now appears to be a combination of movement and insertion. Therefore, we might want to decompose it into an insertion part, combining the 'intermediate' empty category and the phrase marker *you love who* (30a) and a movement part, adjoining *who* to the phrase marker *you think e you love* (30b):

(30) a. [*e* [you love *who*]]
b. [*who* [you think [*e* [you love *t*]]]]

This has the advantage that Form Chain is not a mysterious, novel structure building process. It has the disadvantage that the movement operation yielding (30b) violates the shortest move requirement. But this disadvantage disappears if the shortest move requirement is not part of economy of derivation, as argued here.

In (30b), *who* and its trace *t* are arguably not in the same local domain. We may hypothesize, however, that the intermediate empty element *e* can serve as a link between the antecedent and its trace, thus replicating the effect of successive cyclic movement.

The second reason to be skeptical about the relevance of the shortest move requirement in the domain of wh-movement is that it is a *vacuous* requirement, given the assumptions on structure building entertained in Chomsky (1993).

Chomsky (1993:22) assumes that the structure building process is subject to a condition of Strict Cyclicity (also known as the *extension condition*). In particular, insertion and movement are allowed only if these processes *extend* the phrase marker affected by them. Thus, the only way to move a subpart α of a phrase marker K is by adjoining α on the outside of K. Adjunction of α at any hierarchically lower point in K is excluded by the Strict Cycle condition.

In terms of conditions on movement, this means that at any point in a derivation, the only movement that is allowed is the *longest* possible movement. Given that a shorter movement is blocked by the Strict Cycle condition, there really appears to be no option at all in terms of the length of a movement.[15]

A third reason to have doubts about the shortest move condition is that in the domain of wh-movement violations of 'shortest move' yield various gradations of ungrammaticality. The status of nonlocal wh-movement appears to be dependent on the nature of the local domain's opacity (strong versus weak islands), and on the status of the moved category (D-linked arguments versus nonD-linked arguments and adjuncts)(see Cinque 1990 for a survey).

It is generally assumed since Chomsky (1991) that violations of economy conditions yield the worst kind of ungrammaticality. In Chomsky (1993), violations of economy conditions lead to a crashing derivation, which cannot be repaired by other components of the cognitive system. To the extent that nonlocal head movement and superraising can be reduced to violations of economy of representation (i.e., certain features remain unchecked), the relevant constructions have the expected quality of ungrammaticality. Nonlocal wh-movement constructions do not in

[15] Chomsky (1993:24) tentatively restricts the Strict Cycle condition to overt movement to specifier positions. This implies that for covert movement, the shortest move requirement is not vacuous.

general, suggesting that no economy violation is involved in the derivation of these constructions.

3.3.1 Form Chain

In the previous section, we have argued that Form Chain can be decomposed into two standard structure building processes (cf. (30)). First, the embedded clause is extended by adjunction of an independent empty wh-element (30a). Second, the matrix clause is extended by movement and adjunction of the wh-word generated in the embedded clause. We have assumed that the intermediate empty wh-element then serves to link the antecedent to its trace through a succession of local dependency links.

The more traditional successive cyclic movement faces a problem that the Form Chain process, made explicit in this way, avoids. Suppose the intermediate empty wh-element were not an independently inserted element, but a trace of the wh-moved category itself. This implies that the first step in the derivation involves movement and adjunction of the wh-word generated in the embedded clause to the embedded clause (which, at that point in the derivation is not yet 'embedded')(31a). In the second step, this wh-word will move on and adjoin to the matrix clause (31b):

(31) a. [*who* [you love *t*]]
b. [*who* [you think [*t* [you love *t*]]]]

The principle of Greed (4) requires that there is a trigger for each of the movements in (31a) and (31b). Arguably, there can be a [+wh]-feature in the embedded C, which needs to be checked against the wh-features of an element in Spec,CP.[16] This [+wh]-feature provides a trigger for the movement in (31a), leading to elimination of both the [+wh]-feature in C and the wh-feature of the moved wh-word. But at this point, no trigger is left for the movement of the wh-word in (31b). Its wh-features have been checked in the intermediate Spec,CP, and no further movement is allowed, by Greed.

In the Form Chain approach pursued here, the [+wh]-features of the embedded clause are checked by an independent empty wh-element, leaving the wh-word generated in the embedded clause free to move to the specifier position of the matrix CP.

Chomsky (1995) avoids the problem for successive cyclic movement posed by the principle of Greed by assuming that checking of features does not automatically result in the elimination of features. In particular, the wh-word in (31a) would check and eliminate the [+wh]-feature of

[16] Long distance wh-movement shows effects of the presence of a wh-element in the specifier position of the embedded CP in a number of languages. For example, in Dutch, a wh-complementizer *ofdat* can appear in long distance wh-movement constructions and embedded interrogatives only (cf. Hoekstra and Zwart 1994):

(i) a. Piet denkt (*of) dat Jan het gedaan heeft
Pete thinks if that John it done has
'Pete thinks that John did it.'
b. Wat denkt Piet (of) dat Jan gedaan heeft?
what thinks Pete if that John done has
'What does Pete think John did?'
c. Piet vraagt zich af wat (of (dat)) Jan gedaan heeft
Pete asks himself off what if that John done has
'Pete wonders what John did.'

the embedded C, but would retain its own wh-features for further checking operations to take place in the matrix CP.

There is reason to believe that a wh-element in the specifier position of an embedded CP still has certain features left which need to be checked in the specifier position of the matrix CP. This becomes clear from partial wh-movement constructions in German (McDaniel 1989, Huybregts 1992, Gamon 1994). In these constructions, the step in (31a) is overt, but the step in (31b) does not take place in overt syntax. Instead, a dummy wh-element *was* 'what' is inserted in the specifier position of the matrix CP:

(32) [*Was* glaubst du [*mit wem* ich *t* gesprochen habe]]
what believe you with whom I spoken have
'Who do you think I talked to?'

Gamon (1994) observes that the construction in (32) is ungrammatical if the wh-element in the specifier position of the embedded CP is located inside a weak island (e.g., a wh-island):

(33) * [*Was* fragst du dich [*weshalb* er glaubt
what ask you yourself why he believes
[*mit wem* ich *t* gesprochen habe]]]
with whom I spoken have
'Who do you wonder why he thinks who I talked to?'

This suggests that the wh-phrase in the embedded clause *mit wem* 'with whom' moves covertly (at LF) to the matrix Spec,CP occupied by the dummy wh-word *was* (cf. also Huybregts 1992).[17]

It is not clear that the trigger for the covert movement would involve wh-features. The dummy element in the matrix Spec,CP must have the morphology of a wh-word, suggesting that checking of the wh-features has been taken care of in overt syntax. Nevertheless, (33) appears to present clear evidence for covert movement from the embedded Spec,CP to the matrix Spec,CP.

Crucially, however, the same evidence is lacking in nonpartial wh-movement constructions in English, as weak islands only yield marginally ungrammatical sentences:[18]

(34) ? *Who* do you wonder whether Bill believes *e* we talked to *t* ?

Apparently, the relation between *who* and the intermediate empty element *e* in (34) is not subject to the same constraints as the relation between the dummy wh-element *was* and *mit wem* in (33). This suggests that the empty element *e* in (34) is not a trace of *who* but an independently inserted intermediate wh-element, as proposed here.

This leads to the following conclusion. Even if it is possible for wh-elements to move successive cyclically, the evidence from partial vs. full wh-movement suggests that a derivation in which the wh-phrase moves long distance in one step must always be available. This would be impossible if the shortest move condition is part of economy of derivation.

[17] In Chomsky (1995), the covert movement would involve movement of the relevant *features* to the matrix C, not movement of the entire wh-phrase to the matrix Spec,CP.

[18] The same contrast between partial and full wh-movement can be observed in Frisian, in which both types of wh-construction are possible (Pytsje van der Veen, p.c.).

3.4 Conclusion

It appears that head movement, superraising, and wh-movement do not provide decisive evidence in support of the validity of the shortest move condition. Nonlocal head movement is excluded by economy of representation: skipping relevant functional heads would leave certain features unchecked. Part of the superraising cases can be explained as economy of representation violations as well. For the remaining cases, I have argued that they could only be derived by starting from illegimitate base structures (i.e., structures that would never be generated). Nonlocal wh-movement yields gradations of ungrammaticality and therefore appears to belong to a different category. It can be argued that successive cyclic movement does not consist in a stepwise movement procedure, but in a combination of inserting intermediate empty wh-elements and moving the wh-phrase long distance to the matrix CP.

4. Consequences

4.1 Equidistance

Let us now return to the Equidistance Principle of Chomsky (1993). This principle allows elements to cross a position where they could have landed, provided the target position is in the same minimal domain as the position which is crossed.

(35) *Equidistance* (Chomsky 1993:17):
If α,β are in the same minimal domain, they are equidistant from γ.

As we have seen, positions end up being in the same minimal domain as a function of head movement. This leads to the prediction that objects can leave the VP, crossing the subject position inside VP, only if the verb has moved to AgrO first.[19]

As argued in section 2, this prediction is not borne out. This suggests that the equidistance principle as stated in (35), under the definitions understood, is not relevant for conditions on movement. But this is not surprising if the shortest move conditions does not exist. Since the equidistance principle is merely an explicitation of the shortest move requirement, we are not surprised to find that it has no empirical substance.

The shortest move requirement, including the equidistance principle, does make one important contribution that is lost in the present approach. Chomsky (1993:19) derives from this requirement the fact that movement of subject and object to their respective licensing positions (*L-related* movement) is always crossing instead of nesting. It is well-known that movement to operator positions (*nonL-related movement*) is nesting rather than crossing (Pesetsky 1982).

This follows in Chomsky's analysis, because (for reasons not discussed here) a minimal domain contains at most two equidistant specifier positions.[20] As a result, if the subject moves to the specifier position of AgrO, so that the subject and its trace occupy the two positions that are equidistant from the object, the object cannot move at all. Consequently, the object has to

[19] See Chomsky (1993:12) for the definition of 'minimal domain'. See also Zwart (1993b:231f).

[20] As pointed out by a reviewer, if the Head Movement Constraint does not hold, it becomes possible for more than two specifier positions to end up in the same minimal domain. The specifier of a head that is skipped in the head movement process would also be included in the minimal domain of the chain resulting from the head movement.

hop over the subject first, and only then is the subject allowed to cross the object, again as a function of head movement creating the proper equidistant landing site.

This important result is lost.[21] We could try to achieve the same result by stipulating the order of agreement projections in the functional domain, but that would clearly be inferior to Chomsky's approach.[22] This, then, is the problem to solve if we accept the point argued for here, namely that the shortest move requirement does not exist.

4.2 Reductions

Accepting the arguments against the existence of the shortest move requirement, economy of derivation can be stated as in (2):

(2) *Economy of derivation*:
Use as few steps as possible in deriving an output representation

The question arises whether economy of derivation can be reduced to economy of representation:

(1) *Economy of representation*:
Use as few symbols as possible in an output representation

Inasmuch as steps reduce the number of symbols, (2) can be reduced to (1).

Movement is generally regarded as a chain creating process. If a chain counts as one symbol, movement will not decrease the number of symbols. However, features must count as symbols, or otherwise they would not be relevant for the principle of Full Interpretation:

(36) *Full Interpretation*:
In an interface representation, do not use useless symbols.

Therefore, only movements that result in checking and elimination of features is allowed by (1).

We can also state that strong features count as symbols at the PF interface, but weak features do not. Movement to check and eliminate strong features then reduces the number of symbols, but movement to check and eliminate weak features does not. *Procrastinate* (3), then, can be reduced to economy of representation holding for the PF interface.

[21] However, as pointed out to me by Liliane Haegeman and Gereon Müller, the Equidistance Principle as stated is not able to accommodate derivations involving two internal arguments and one external argument. Assuming both internal arguments to have designated licensing positions in the functional domain (i.e. there is a sequence of two AgrOPs), the higher internal argument will at some point in the derivation have to cross the basic position of the external argument (i.e. the Spec,VP) and the licensing position of the lower internal argument (i.e. the specifier of the lower AgrOP) on its way to its licensing position (i.e. the specifier of the higher AgrOP). Spec,VP and the two Spec,AgrOPs, however, cannot all three be equidistant from the basic position of the higher internal argument. The Equidistance Principle appears to be unable to derive these structures. If so, the Equidistance Principle captures the crossing requirement on A-movement only partly. However, see Collins and Thráinsson (1993) for an analysis of double object constructions that does not incur this problem.

[22] As pointed out by a reviewer, even stipulating the order of agreement projections does not help much. If we stipulate that AgrSP must be higher than AgrOP, we still cannot exclude that a subject is generated with objective Case morphology and licensed in AgrOP (yielding *She loves him* meaning *He loves her*). It seems, then, that the organization of the functional domain expresses a (partial) link between argument status, morphology, and word order.

Greed (4) follows from (1) if we take (1) to apply in a nonglobal fashion, that is: every element must carry as few symbols as possible in the output representation. Thus, (1) forces a feature bearing element α to move, but prohibits movements that do not result in the elimination of a feature carried by α.

If this conception of *Greed* is correct, economy of derivation (i.e. the 'fewest steps' requirement) and the principles that fall under economy of derivation (the *inertness* principles) can be seen to reduce to economy of representation.

Finally, note that economy of representation can be reformulated as in (37):

(37) *Economy of representation (reformulated)*
In an output representation, do not use superfluous symbols.

It is clear from this reformulation that economy of representation, now incorporating economy of derivation, is equivalent to the principle of *Full Interpretation* (35).

5. Conclusion

I have argued that economy of derivation does not contain a requirement that steps be as short as possible. Restrictions on head movement, superraising, and wh-movement follow from well established principles and mechanisms made explicit in recent minimalist work. This resolves the conflict between the 'fewest steps' requirement and the 'shortest move' requirement in Chomsky (1993). I argued that Chomsky's *Form Chain* mechanism must be decomposed into a 'trace' insertion mechanism and a movement mechanism, where the intermediate 'traces' are inserted before the movement takes place, in accordance with strict cyclicity. Finally, I offered some speculation on the possibilities of reducing economy of derivation to economy of representation, and economy of representation to the principle of Full Interpretation.

It is interesting to note that, if we were correct in the above, generative grammar, even in the purely derivational approach that incorporates the structure building process of generalized transformations (including movement), does not need any constraints beyond conditions on representation.[23]

Acknowledgments

This material was presented at the TABU-dag at the University of Groningen, June 25, 1993, at the Stafseminar of the Department of Linguistics of the University of Tilburg, September 8, 1993, and at the Research Seminar of the Department of Linguistics of the University of Geneva, March 8, 1994. I would like to thank the audiences at these occasions, in particular Jaume Solà, Henk van Riemsdijk, Riny Huybregts, Luigi Rizzi, Ur Shlonsky, and Liliane Haegeman. Also thanks to Eric Hoekstra, Phil Branigan, Gereon Müller, Anthony Kroch, Erich Groat, and K. Scott Ferguson and to two reviewers for this volume.

[23] This final remark presupposes that locality conditions and conditions that ensure the correct order of projections in the functional domain are to be considered as representational rather than derivational conditions. As for the conditions on interpretation requiring a moved element to be related to its trace, it was assumed in the text that these are either representational conditions, or extragrammatical conditions of interpretation.

References

Bennis, Hans. 1986. *Gaps and Dummies*. Dordrecht: Foris.

Chomsky, Noam. 1973. Conditions on Transformations. In S.R. Anderson and P. Kiparsky, eds., *A Festschrift for Morris Halle*. New York: Holt, Rinehart, and Winston.

Chomsky, Noam. 1986a. *Knowledge of Language. Its Nature, Origin, and Use.* New York: Praeger.

Chomsky, Noam. 1986b. *Barriers*. Cambridge: MIT Press.

Chomsky, Noam. 1991. 'Some Notes on Economy of Derivation and Representation'. In R. Freidin, ed., *Principles and Parameters in Comparative Grammar*. Cambridge: MIT Press.

Chomsky, Noam. 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.* Cambridge: MIT Press. Cambridge: MIT.

Chomsky, Noam. 1994. *Bare Phrase Structure.* MIT Occasional Papers in Linguistics 5.

Chomsky, Noam. 1995. 'Categories and Transformations.' Ms., MIT.

Cinque, Guglielmo. 1990. *Types of A'-Dependencies*. Cambridge: MIT Press.

Collins, Chris and Höskuldur Thráinsson. 1993. 'Object Shift in Double Object Constructions and the Theory of Case.' In C. Phillips, ed. *Papers on Case and Agreement II (MIT Working Papers in Linguistics 19).* Cambridge: MIT. 131-173.

Frank, Robert and Anthony S. Kroch. 1992. 'Generalized Transformations in Successive Cyclic and Long Dependencies (or Substitution + Extending = Adjoining)'. Paper presented at the GLOW Colloquium, Lund, April 6.

Gamon, Michael. 1994. 'Partial Wh-Movement in German: A Nonhomogeneous WH-Dependency.' Paper presented at the Western Conference on Linguistics, Los Angeles, October 21.

Grimshaw, Jane. 1989. 'Complement Selection and the Lexicon.' *Linguistic Inquiry* 10. 279-326.

Hoekstra, Eric and C. Jan-Wouter Zwart. 1994. 'De structuur van de CP: Functionele projecties voor topics en vraagwoorden in het Nederlands'. *Spektator* 23.3. 191-212.

Holmberg, Anders. 1986. *Word Order and Syntactic Features in the Scandinavian Languages and English*. University of Stockholm: PhD dissertation.

Huybregts, Riny. 1992. 'Scattering: A Derivational Approach without S-structure'. Paper presented at the 8th Workshop on Comparative Germanic Syntax, Tromsø, November 21.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge: MIT Press.

Koster, Jan. 1987. *Domains and Dynasties. The Radical Autonomy of Syntax.* Dordrecht: Foris.

Mahajan, Anoop. 1990. *The A/A' Distinction and Movement Theory.* MIT: PhD dissertation.

McDaniel, Dana. 1989. 'Partial and Multiple Wh-Movement.' *Natural Language and Linguistic Theory* 7. 565-604.

Moro, Andrea. 1993. *I predicati nominali e la struttura della frase*. Padova: Unipress.

Mulder, René. 1992. *The Aspectual Nature of Syntactic Complementation.* Leiden: PhD dissertation.

Ouhalla, Jamal. 1989. 'Against the Head Movement Constraint.' *UCL Working Papers in Linguistics 1*. 193-205.

Pesetsky, David M. 1982. *Paths and Categories.* MIT: PhD dissertation.

Pesetsky, David M. 1994. *Zero Syntax.* Cambridge: MIT Press.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge: MIT Press.

Travis, Lisa DeMena. 1984. *Parameters and Effects of Word Order Variation*. MIT: PhD dissertation.

Vikner, Sten. 1991. *Verb Movement and the Licensing of NP-Positions in the Germanic Languages*. University of Geneva: PhD dissertation.

Wilder, Christopher and Damir Ćavar. 1993. 'Word Order Variation, Verb Movement, and Economy Principles'. Ms., University of Frankfurt am Main.

Wyngaerd, Guido Vanden. 1989. 'Object Shift as an A-Movement Rule.' *MIT Working Papers in Linguistics 11*. 256-271.

Zwart, C. Jan-Wouter. 1992. 'Matching'. In D. Gilbers and S. Looyenga, eds. *Language and Cognition 2. Yearbook 1992 of the Research Group for Linguistic Theory and Knowledge Representation of the University of Groningen*. Groningen: University of Groningen.

Zwart, C. Jan-Wouter. 1993a. 'Verb Movement and Complementizer Agreement'. In J.D. Bobaljik and C. Phillips, eds. *Papers on Case and Agreement I (MIT Working Papers in Linguistics 18)*. Cambridge: MIT. 297-340.

Zwart, C. Jan-Wouter. 1993b. *Dutch Syntax: A Minimalist Approach.* University of Groningen: PhD dissertation.

Zwart, C. Jan-Wouter. 1994a. 'Dutch is Head Initial'. *The Linguistic Review* 11.3/4. 377-406.

Zwart, C. Jan-Wouter. 1994b. 'On 'Holmberg's Generalization'.' In A. de Boer, H. de Hoop, and H. de Swart, eds. *Language and Cognition 4. Yearbook 1994 of the research group for theoretical and experimental linguistics of the University of Groningen.* Groningen: University of Groningen.