Norvin Richards

The Principle of Minimal Compliance

The syntactic literature discusses a number of phenomena in which a constraint which rules out a certain class of syntactic dependencies fails to rule out structures containing both an ill-formed dependency and a well-formed dependency; well-formed dependencies seem to be able to "help" dependences which would be ill-formed in isolation. In this paper I attempt to provide a unified account of these phenomena. I postulate a principle which allows the computational system to "ignore" parts of a syntactic structure which have already been checked with respect to a particular constraint.

Keywords: Reflexivity, Subjacency, weak crossover, Connectedness, parasitic gaps, wh-movement

1. Introduction

The phenomena exemplified in (1-5) seem to share a common quality; in all of these cases, a dependency which would be ill-formed in isolation is somehow "saved" by the presence of a well-formed dependency. In other words, it looks as though in these cases, at least, a given constraint only has to be satisfied once in a certain domain:

Reflexivity (Reinhart and Reuland 1993)

(1) a. * Henk$_i$ wees zich$_i$ aan mij toe.

Henk assigned self to me

b. Henk$_i$ wees mij aan zichzelf$_i$ toe.

Henk assigned me to self-self

c. Henk$_i$ wees zich$_i$ aan zichzelf$_i$ toe.

Henk assigned self to self-self

Weak Crossover (adapted from Hornstein 1995)

(2) a. * Who$_i$ did his$_i$ mother introduce t_i to Mary?

b. Who$_i$ did John introduce t_i to his$_i$ teacher?

c. ? Who$_i$ did his$_i$ mother introduce t_i to his$_i$ teacher?

VP-ellipsis[1] (Danny Fox, personal communication)

(3) a. * I introduced [John and Mary]$_i$ to Bill before they$_i$ could (introduce themselves$_i$ to Bill).

b. John introduced himself to Mary before Bill$_i$ could (introduce himself$_i$ to Mary).

c. I introduced [John and Mary]$_i$ to each other$_i$ before they$_i$ could (introduce themselves$_i$ to each other$_i$).

(4) Subjacency (Brody 1995)

a. * What$_j$ do you wonder who bought t_j?

b. Who$_i$ t_i wonders who bought the book?

c. Who$_i$ t_i wonders who bought what$_j$?

Connectedness (Kayne 1983)

(5) a. * a person who$_i$ people that talk to e_i usually like him

b. a person who$_i$ people that talk to John usually like e_i

c. a person who$_i$ people that talk to e_i usually like e_i

The (a) cases all violate some constraint which the (b) cases obey. Our standard assumptions about the way in which constraints operate lead us to expect that sentences containing both the configuration in (a) and that in (b) should be ill-formed; if the configuration in (a) is what violates the constraint in question, then any sentence containing that configuration should be doomed. In fact, this is not the case; the (c) sentences are better than the (a) sentences. There have been a variety of attempts to explain these facts, of course, but no unified account of all of them has ever been given, to my knowledge. In this paper I will try to isolate the factor that (1-5) have in common, thus simplifying the constraints involved. I will then go on to apply this principle to the notion of Shortest Move, attempting to show that under a certain conception of Shortest Move, we can derive Holmberg's (1986) Generalization, Pesetsky's (1982) Path Containment Condition, the facts of Stylistic Fronting in Icelandic, and the behavior of participial agreement in French from considerations needed to deal with the facts in (1-5).

Traditionally, constraints are thought of as checking the entire structure for violations, and rejecting structures that contain any violations at the point or points in the derivation at which the constraint applies. The data in (1-5) suggest that in some cases, at least, the presence of a dependency that satisfies a constraint can allow the computational system to ignore another dependency which would be ill-formed in isolation. The idea I want to develop here is that the computational system tries to avoid checking the same constraint in the same portion of the structure more than once[2]. As a result, if a constraint is obeyed, the portion of the structure in which it is obeyed can be ignored thenceforth with respect to that constraint.

Let us formalize this idea as the Principle of Minimal Compliance (PMC), given in (6):

(6) <u>Principle of Minimal Compliance</u>

For any dependency D that obeys constraint C, any elements that are relevant for determining whether D obeys C can be ignored for the rest of the derivation for purposes of determining whether any other dependency D' obeys C.

The notion of "relevance" involves at least the following qualifications:

(7) An element X is relevant to determining whether a dependency D with head A and tail B obeys constraint C if

a. X is along the path of D (that is, X=A, X=B, or A c-commands X and X c-commands B).

AND b. X is a member of the class of elements to which C makes reference.

(7a) allows the PMC to apply only to elements along the “path” of a dependency. We could try to build (7a) into the statement of the PMC, but it is interesting to note that there seem to be no constraints making reference to elements that are not in the set defined by (7a) (for example, there are no constraints rendering a dependency ill-formed if some element is present in a position c-commanded by the head which does not c-command the tail[3]). However this fact is to be reflected in our theory, it seems clear that it should not be specific to the PMC; rather, let us assume that the PMC is “aware” of the fact that elements not lying on the path of the dependency are always irrelevant for determining whether the dependency is well-formed. (7b) is more straightforward; it tells us not to bother with things not mentioned by the constraint, and is roughly akin to the basic insight of Rizzi’s (1990) Relativized Minimality.

2. Reflexivity, WCO, VP-ellipsis, Subjacency, Connectedness

Let us go on to see how the data in (1-5) follow from simple constraints in conjunction with the PMC.

2.1 Reflexivity

The facts in (8) now follow straightforwardly from application of the PMC to something like the constraint in (9)[4]:

(8) a. * Henk$_i$ wees zich$_i$ aan mij toe.

Henk assigned self to me

b. Henk$_i$ wees mij aan zichzelf$_i$ toe.

Henk assigned me to self-self

c. Henk$_i$ wees zich$_i$ aan zichzelf$_i$ toe.

Henk assigned self to self-self

(9) In Dutch, all anaphors with a clausemate binder must be *zichzelf*.

Here the dependency between *Henk* and *zichzelf* obeys the constraint on reflexives in (9). The PMC then allows us to ignore all elements which are relevant for determining that the dependency between *Henk* and *zichzelf* obeys (9) for purposes of determining whether other dependencies obey (9). *Henk* and *zichzelf* are themselves among the relevant elements, since the constraint refers to an anaphor and its binder and they are along the path of the dependency. Thus, in particular, we will be allowed to ignore *Henk*[5] for purposes of determining whether other anaphors in the sentence obey (9). Next we look at *zich*, to determine whether it obeys (9). If we ignore *Henk*, the anaphor is not bound within its clause. Thus, (9) is satisfied[6].

Of course, *zich*, being an anaphor, does in fact have to be bound by *Henk* in order to be interpreted. For the account sketched above to be maintained, we must apparently understand *zich* as being subject to two separate constraints, the one in (9) and another which requires it to have an antecedent. The PMC allows us to ignore *Henk* with respect to the constraint in (9), but has no effect on the constraint that requires *zich* to be bound, which is still in force. It seems clear that in principle we do want the PMC to apply separately to different constraints; we do not want an anaphor which is in a well-formed binding relation to be immune to the Case Filter (or its successors), for instance.

At the same time, there is something conceptually awkward about separating the constraints on *zich* in this way. We might prefer, for instance, a theory of anaphoric binding which derives all the properties of anaphoric binding from a single requirement, rather than from several distinct ones. Such a theory, then, would derive both the requirement that *zich* be bound and that it not be locally bound from a single constraint. If we were to construct such a theory, the version of the PMC sketched above would be inadequate; ignoring a binder for purposes of the requirement that *zich* not be locally bound would entail ignoring it for purposes of the requirement that *zich* be bound, and would not improve the structure.

One possible move to make at this point would be to reformulate the PMC as marking parts of the structure as having already obeyed a constraint, rather than allowing us to "ignore" portions of the structure. On such an account, the well-formed relation between *Henk* and *zichzelf* would mark *Henk* as obeying the generalized condition on anaphora, and would allow *Henk* to locally bind *zich*.

Despite the conceptual attractiveness of this move, I will not make it here. The approach as sketched immediately above is too strong. Suppose we were to succeed in developing a theory in which all the properties of anaphoric binding, including the requirement that anaphors be bound, the choice between *zich* and *zichzelf*, and the locality constraints on certain anaphors (such as *zichzelf*), were built into a single constraint. Certain parts of this constraint would appear to be subject to the PMC, including the choice between *zich* and *zichzelf*, as we have seen. Others clearly are not; for instance, a binder which binds an anaphor in an appropriately local fashion is not then licensed to illegitimately bind another anaphor long-distance. The version of the PMC which was just sketched cannot distinguish between these cases; if a binder enters into a well-formed binding relation, it should then be invulnerable to all the constraints on binding, which does not seem to be the case. We must apparently understand anaphors as being subject to a number of separate constraints, which interact with the PMC in different ways[7]. I will discuss this issue further in section 2.6.

2.2 Weak Crossover

The facts in (10) follow in a similar way from the PMC applied to a straightforward constraint on weak crossover:

Weak Crossover (adapted from Hornstein 1995)

(10) a. * Who$_i$ did his$_i$ mother introduce t$_i$ to Mary?

b. Who$_i$ did John introduce t$_i$ to his$_i$ teacher?

c. ? Who$_i$ did his$_i$ mother introduce t$_i$ to his$_i$ teacher?

(11) All pronouns bound by a wh-phrase must also be bound by a trace of that wh-phrase in an A-position.

In (10c), the relation between *who* and the second instance of *his* obeys (11). This will enable us to ignore *who* for purposes of determining whether other pronominal variables obey (11), just as we could ignore *Henk* in the reflexivity case. Thus, the first instance of *his* is effectively not bound by a wh-phrase, since we are allowed to ignore the wh-phrase that does bind it, and the structure is well-formed. The facts are quite similar to those in the Reflexivity case, and fall out in a similar way; a constraint on possible relations between binders and bound elements need only be satisfied once by a given binder. After a binder obeys the constraint once, the PMC allows us to ignore it thenceforth for purposes of evaluating the constraint, and the constraint thus fails to apply to anything else bound by that binder.

2.3 VP-ellipsis

Similar reasoning yields the paradigm of VP-ellipsis described below:

VP-ellipsis (Danny Fox, personal communication)

(12) a. * I introduced [John and Mary]$_i$ to Bill before they$_i$ could (introduce themselves$_i$ to Bill).

b. John introduced himself to Mary before Bill$_i$ could (introduce himself$_i$ to Mary).

c. I introduced [John and Mary]$_i$ to each other$_i$ before they$_i$ could (introduce themselves$_i$ to each other$_i$).

(13) Bound variables in an elided VP must correspond to bound variables in the non-elided VP model.

The facts in (12) seem to show that an R-expression cannot have a correspondent in an elided VP which is a bound variable, unless the elided VP also contains a bound variable which corresponds to a bound variable in the unelided VP. Thus, (12a), where the R-expression *John and Mary* corresponds to a bound variable in the elided VP, is ill-formed. (12c), which contains the same configuration but also contains a bound variable corresponding to the bound variable *each other*, is well-formed[8]. The Principle of Minimal Compliance allows us to understand this pattern as involving the parallelism requirement (13) (probably a specific case of a more general requirement of parallelism between elided VPs and their non-elided counterparts; for discussion of this requirement, cf. Chomsky 1965, Sag 1976, Williams 1977, Fiengo and May 1994, Fox 1995, and much other work). Thus, the relation between *they* and *each other* in the elided VP in (12c) is one of variable-binding, but this relation obeys (13), since *each other* is a bound variable corresponding to a bound variable in the unelided original. This allows us to ignore *they* and *each other* for purposes of computing the well-formedness of the rest of the structure with respect to (13). As a result, we can ignore the binder for *themselves* in the elided VP, and *themselves* is thus effectively not a bound variable, since we can ignore its binder; (13) is therefore satisfied. The pattern is thus entirely parallel to the weak crossover and reflexivity facts discussed above; an element of a certain kind cannot be bound unless another element of a different kind is also bound. The well-formed binding relation allows us to ignore the potentially offending binder, and all other elements bound by that binder are thus free from the constraints in question.

2.4 Subjacency

Subjacency has classically (Huang 1982 and much subsequent work) been claimed to apply only to overt movement, on the basis of contrasts like those in (14):

(14) a. * What$_j$ do you wonder who bought t$_j$?

b. Who$_i$ t$_i$ wonders who bought what$_j$?

(14a) is a standard wh-island violation; *what*, by crossing the [+wh] complementizer whose specifier is occupied by *who*, violates Subjacency. (14b) has an interpretation in which *what* and *who$_i$* both have matrix scope; this reading is classically taken to be derived by LF movement of *what* to the specifier of the matrix CP. This move presumably also violates Subjacency, just like its overt counterpart in (14a), yet the structure is well-formed. The conclusion which has typically been drawn from this is that Subjacency only applies to overt movement.

Brody (1995) suggests an alternative way of looking at the facts in (14), connecting them to Kayne's (1983) theory of Connectedness, to which we will turn in the next section. Translating Brody's insight into our terms, we can see the contrast between (14a) and (14b) as an instance of the PMC; in (14b), the fact that the matrix [+wh] C^o is associated with *who$_i$* in a way which obeys Subjacency frees it from having to obey Subjacency in the case of *what*. Subjacency, then, can be stated as follows:

(15) A [+wh] complementizer cannot be associated with a wh-phrase via movement in a way which crosses barriers of a certain kind (wh-islands, complex noun phrases, etc.).

The PMC, applied to the version of Subjacency in (15), gets the facts in (14) without any need to appeal to a distinction between overt and covert movement. The reasoning is essentially identical to that used in the three cases we have just reviewed. In (14a), the matrix [+wh] complementizer violates Subjacency by being associated with *what*$_j$ across a wh-island. In (14b), the matrix [+wh] complementizer obeys Subjacency when it associates with *who*$_i$. This allows us to ignore the matrix [+wh] complementizer when evaluating the rest of the structure for Subjacency. Next we can evaluate the association between the matrix [+wh] complementizer and *what*$_j$ to determine whether it obeys Subjacency. If we ignore the matrix [+wh] complementizer, it does, since this dependency is no longer subject to Subjacency at all. The well-formedness of (14b) follows.

Making this move will apparently require us to adopt a derivational perspective on the grammar (contra Brody (1995)). In particular, we will apparently need to assume that in cases in which one move triggers the PMC in a way which allows another move to occur, the first move must precede the second move. That is, Subjacency is not simply a constraint on representations; moves are evaluated for obedience to Subjacency as they occur, and if a move violates Subjacency it cannot be saved by a later move's having triggered the PMC.

(16b) involves a case in which, according to our standard ways of looking at derivations, the licit move precedes the illicit move, thus triggering the PMC and allowing the illicit move to take place. If Subjacency were purely representational, we would expect to find cases of the opposite kind, in which an overt Subjacency-violating move is licensed by a covert Subjacency-obeying move. This does not appear to be the case:

(16) a. * Which$_i$ car did John persuade the man who bought t_i to sell the hubcaps?

b. Who$_i$ t_i persuaded the man who bought which car$_j$ to sell the hubcaps?

c. * Which$_i$ car did John persuade the man who bought t_i to sell which hubcaps$_j$?

(16a-b) are parallel to (14a-b). (16a) is a Subjacency violation, in this case a violation of the Complex NP Constraint. In (16b) we can see the now-familiar saving effect of a second wh-word. (16c) is the case which shows that Subjacency must be interpreted derivationally; here the Subjacency-violating move will be followed at LF by the Subjacency-obeying move of *which hubcaps*. This is not enough to save the structure, however; Subjacency is still violated. It looks as though the order in which moves take place is still relevant for Subjacency; licit moves must precede illicit moves in the derivation[9].

Obliterating the distinction between overt and covert movement for purposes of Subjacency will have far-reaching effects on the theory, none of which I can explore adequately here. We expect that in languages such as Japanese, which move all of their wh-words covertly, the first wh-word moved to a specificer of a given CP will have to obey Subjacency[10,11]. This is, of course, the conclusion reached by Watanabe (1992) in his analysis of Japanese wh-movement, although he draws somewhat different conclusions:

(17) a. ?? John-wa [Mary -ga nani -o katta ka dooka] siritagatte-iru no?
John TOP Mary NOM what ACC bought whether know-want Q
'What does John want to know whether Mary bought?'

b. John-wa [Mary-ga nani -o katta ka dooka] dare-ni tazuneta no?
John TOP Mary NOM what ACC bought whether who DAT asked Q
'Who did John ask whether Mary bought what?'

We also expect that in languages which move all of their wh-words overtly, only the first wh-word moved to a specifier of a given CP will have to obey Subjacency. This appears to be the case (Bulgarian, Roumyana Slabakova, Roumyana Izvorski, personal communication)[12]:

(18) a. * Koja kniga$_i$ razprostranjavas#e z#urnalistat

which book spread the-journalist

[malva#ta c#e senatora#t iska da zabrani t_i]?

the-rumor that the-senator wanted to ban

'Which book did the journalist spread the rumor that the senator wanted to ban?'

b. ? Koj z#urnalist$_j$ koja kniga$_i$ t_j razprostranjavas#e

which journalist which book spread

[malvata c#e senatora#t iska da zabrani t_i]?

the-rumor that the-senator wanted to ban

'Which journalist spread the rumor that the senator wanted to ban which book?'

In other words, it appears to quite generally be true that in cases involving multiple wh-movement to a single [+wh] complementizer, only the first moved wh-word will have to obey Subjacency; the other wh-movements are free from Subjacency. This is true regardless of the levels at which the wh-words move, as we expect in a PMC-based explanation of the phenomenon, but not in an explanation which makes crucial reference to the level at which movement takes place.

In this discussion I have assumed that such a thing as covert wh-movement exists. A number of recent works (including Cole and Hermon 1994, Tsai 1994, Reinhart 1995, Ouhalla 1996, among others) have denied this assumption, claiming that some or all cases of wh-in-situ involve interpretation in situ via a mechanism like unselective binding, with no movement relation between the wh-word and its scope position at any point in the derivation. Approaches of this kind, depending on their specific properties, might fail to capture the parallelism between the English facts in (14) and the Bulgarian ones in (18); an account which claims that (14b) is well-formed because the *what* in situ fails to move at any point in the derivation clearly does not generalize to the well-formedness of (18b)[13]. Still, there is no fatal clash between an unselective binding approach to wh-in-situ and the theory developed here. The PMC applies to dependencies other than those created by movement, as we have seen; the Japanese facts in (17) will follow from the PMC whether the relevant constraint is one on movement or on unselective binding.

We may have to redefine our notion of Subjacency to deal with the ill-formedness of (19):

(19) a. * Who$_i$ did John hire t$_i$ [without asking [who had interviewed t$_i$]]?

b. * Who$_i$ did John ask [before hiring t$_i$] [who had interviewed t$_i$]?

(19a-b) both appear to violate Subjacency (cf. Chomsky 1986 for discussion). This is unexpected on the view of Subjacency developed above, according to which Subjacency constrains relations between [+wh] complementizers and wh-words. In each of the sentences in (19), the matrix [+wh] complementizer enters into one relation that obeys Subjacency (in the main clause, in (19a), and in the adjunct clause, in (19b)). This should render the other relation impervious to Subjacency, but this appears not to be the case. One way of dealing with this problem would be to take advantage of the fact that the two wh-chains in (19a-b) both have a single index; we might expand our definition of Subjacency in the following way[14]:

(20) A [+wh] complementizer cannot be associated with a set of wh-phrases which bear a given index in a way which crosses barriers of a certain kind (wh-islands, complex noun phrases, etc.).

2.5 Connectedness

Kayne's (1983) Connectedness also follows from application of the PMC to Huang's (1982) CED, given in (22). Kayne's (1983) insight was that a CED-obeying dependency can eliminate the effect of a single CED island lying along its path. This is shown in (21):

(21) a. * the person Op_i that John described Mary [without examining any pictures of e_i]

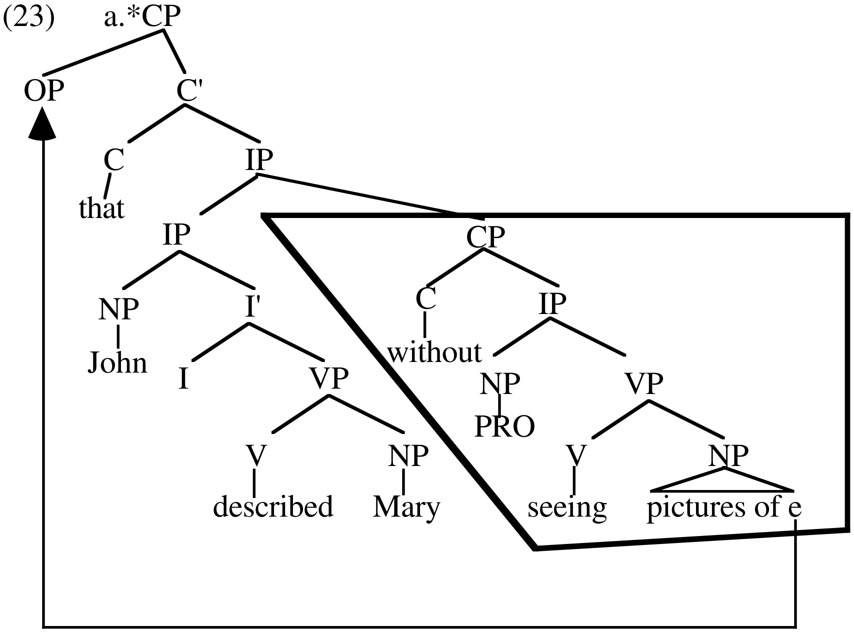
b. ? the person Op_i that John described e_i [without examining any pictures of e'_i]

c. * the person Op_i that John described e_i [without [any pictures of e'_i]being on file]

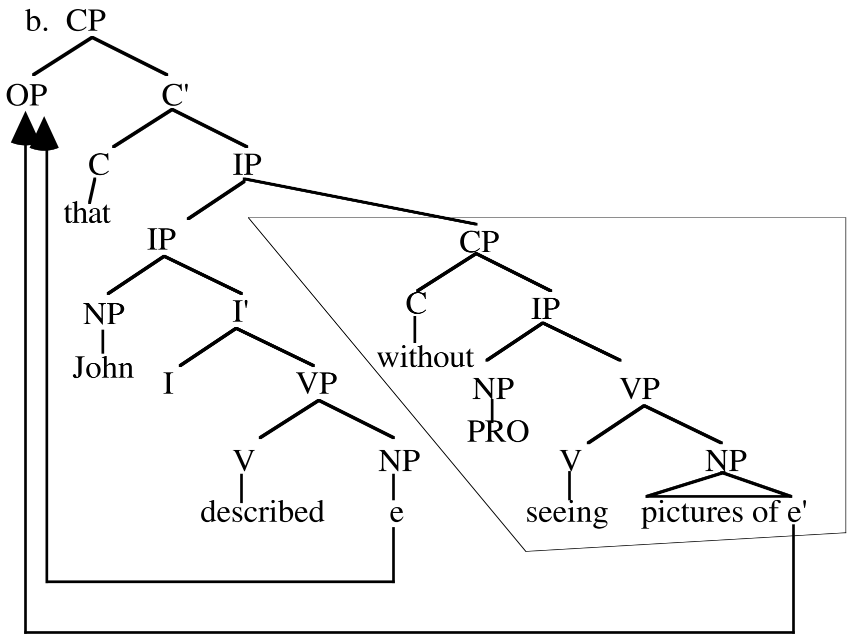
(22) Certain empty categories (including wh-traces) must have a binder which is not separated from them by any ungoverned XP boundaries.

In (21b), the link between OP and e' crosses one CED island (namely, the adjunct CP *without examining...*), while in (21c) the link crosses two CED islands (the adjunct CP and its subject *any pictures of*). The contrasts in (21) follow from the theory developed here, as can be seen from the simplified trees in (23), showing the structure of the relative clauses in (21):

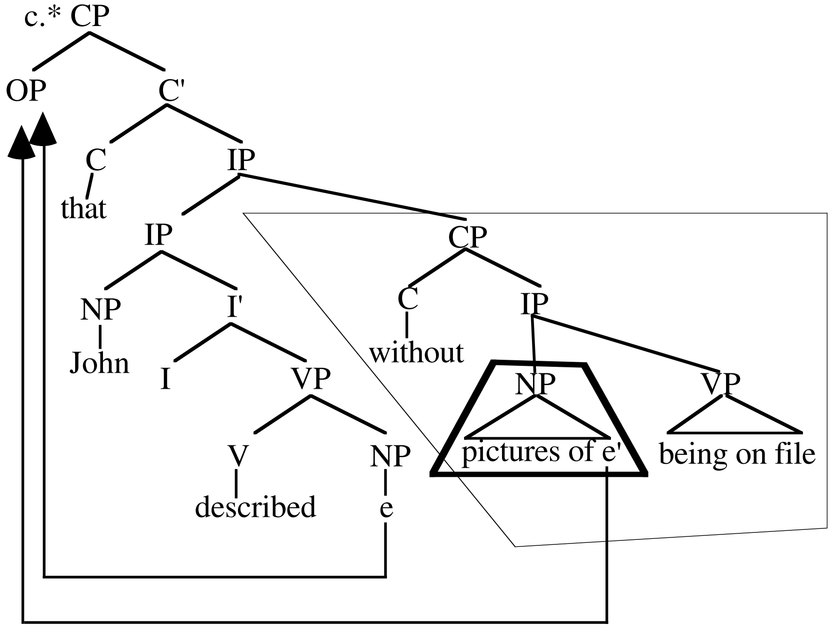
(23) a.*CP



b.



c.*



(23a) shows an ordinary CED violation; the dependency between OP_i and e_i crosses an ungoverned XP boundary (namely, the adjunct CP) and is rendered ill-formed. In (23b), the dependency between OP_i and e_i obeys the CED. The CED makes reference to the head and tail of a chain, and also to ungoverned XPs between them; thus, we can ignore all of these for purposes of determining whether other dependencies obey the CED.

The dependency between OP_i and e'_i has an ungoverned XP boundary (that of the adjunct *without seeing...*) intervening along its path. However, this XP is one of the elements that we are now allowed to ignore for purposes of computing violations of the CED. Thus, the dependency between OP and e'_i is well-formed, since the only ungoverned XP boundary intervening between the head and tail of the chain can be ignored via the PMC.

In (23c), on the other hand, OP_i and e'_i are separated not only by an adjunct but by a subject island; e'_i is contained in the subject of the adjunct clause. We can ignore the adjunct, as before, but the PMC does not allow us to ignore the subject, *[pictures of e]*; not being on the path of the dependency between OP_i and e_i, it is irrelevant for checking the well-formedness of that dependency. The ill-formedness of (23c) follows.

An interesting fact about CED effects is that the set of elements which a well-formed dependency allows us to ignore is the same as the set of elements which are potentially barriers for Shortest Move; namely, the set of elements along the path of the dependency in question. This theory allows us to state this correspondence directly; the elements along the path of the dependency are the ones which are relevant in the technical sense developed here. It is worth stressing that "relevant" is in fact a technical notion, rather than an intuitive one; the islands which are being ignored in the CED case are not ones which can actually interfere with the well-formed dependency (if they could, the dependency would not be well-formed). Rather, they are islands which satisfy the criteria for relevance outlined in (7).

2.6 Interlude: Subjacency vs. Connectedness

Note that in order to account for the contrast between (21b) and (21c), we must prevent the line of reasoning used to deal with the WCO, Reflexivity, VP-ellipsis, and Subjacency facts from being applicable here. Recall that all of these cases involved a single binder that participated in multiple dependencies, one of which would be ill-formed in isolation, while the other would be well-formed. The PMC, triggered by the presence of the well-formed dependency, operated in these cases by allowing us to ignore the binder, thus rendering certain constraints inapplicable to the other elements bound by that binder.

In this case, we cannot allow a CED-violating dependency to be improved by ignoring the binding operator. If this option were left open, (21b) and (21c) would both be predicted to be well-formed, since the operator participates in a CED-obeying dependency in both cases. The particular statement of the CED in (20) avoids this problem, since it requires every trace to have a binder with certain properties, thus marking a trace with no binder as ill-formed. If we were to use the PMC to ignore the binder, then, we would simply doom the structure, rather than saving it[15].

The approach taken here assumes that a dependency may be subject to a number of constraints, each of which may interact independently with the PMC; a dependency which is well-formed with respect to one constraint may, of course, be ill-formed because of another. For instance, wh-movement is subject both to Subjacency, a constraint on [+wh] complementizers which forces them to attract only local wh-words, and to the CED, a constraint on empty categories which forces them to have local binders. The first constraint may be obviated by ignoring the attracting [+wh] complementizer; a Subjacency-obeying dependency renders the [+wh] complementizer involved completely impervious to Subjacency for the rest of the derivation, by making all the dependencies it triggers unrecognizable as dependencies of the type which Subjacency constrains. The prediction, then, is that the PMC can save Subjacency violations which are arbitrarily deeply embedded. This seems to me to be true:

(24) a. * What$_j$ did you wonder [who asked Bill [who bought t$_j$]]?

b. ? Who$_i$ t$_i$ wondered [who asked Bill [who bought what$_j$]]?

The Subjacency-saving effect of the PMC is seen in (24b), even though *what* is embedded in two wh-islands.

This contrasts with the behavior of Connectedness, as we have seen[16]. In Connectedness cases the placement of the islands involved is crucial; only islands along the path of the well-formed dependency may be ignored. In this approach this follows from the nature of the CED, the constraint responsible for Connectedness effects. Because the CED requires empty categories to have a sufficiently local binder, ignoring the binder will not make dependencies immune to the CED; it will simply make them ill-formed. Thus, CED violations can only be remedied by making the offending islands transparent.

Thus, the PMC will have very different effects on constraints like the CED, which force dependent elements to have binders of a certain kind, than it will on constraints like Subjacency, which constrain binders, or on constraints like the one responsible for the Dutch Reflexivity effects discussed in section 2.1, which force dependent elements <u>not</u> to have binders of a certain kind. It is not uncommon for binders to enter into multiple dependencies, and the PMC will allow constraints on binders to be violated by all but a single dependency headed by a given binder. On the other hand, dependent elements typically do not enter into more than one dependency of a given type, and the effects of the PMC on locality constraints on dependent elements will therefore be less frequently seen; the PMC will only have an effect if it can render invisible everything which causes a particular dependent element to violate the constraint. In some cases, of course, this will mean the PMC will have no effect at all. For instance, it is reasonable to assume that local anaphors are subject to a requirement that they be locally bound. If this requirement were like the constraint on Dutch anaphora discussed in section 2.1, we would expect the PMC to operate on it in a similar way, rendering (25c) well-formed:

(25) a. John$_i$ told himself$_i$ [that Mary liked Susan].

b. * John$_i$ told Susan [that Mary liked himself$_i$].

c. * John$_i$ told himself$_i$ [that Mary liked himself$_i$].

In (25c), the second instance of *himself* is bound by an antecedent which has already entered into a well-formed binding relation with the first instance of *himself*. If the requirement that *himself* be locally bound were a constraint on the binder, like Subjacency, we would expect (25c) to be well-formed. On the other hand, if the constraint in question is, like the CED, a constraint on dependent elements of a certain kind which requires them to have a sufficiently local binder, then the ill-formedness of (25c) is expected. Ignoring *John* in (25c) will not improve the status of the dependency between *John* and the second instance of *himself*; the latter will still lack a sufficiently local binder.

Of course, the CED and the requirement that certain anaphors be locally bound presumably make reference to different sets of opacity-inducing elements. One result of this difference is that the PMC apparently never interacts with the locality condition on local anaphors at all; there is no equivalent of Connectedness in this domain. In the CED case, Connectedness effects arise in cases in which one of the elements which can render a dependency ill-formed (an ungoverned XP, in this case) lies along the path of a well-formed dependency. In the theory being developed here, the absence of Connectedness effects for anaphora would follow from the fact that the elements which render an anaphoric dependency insufficiently local can never lie along the path of a well-formed anaphoric dependency. For instance, the relevant opacity-inducing element in (25b-c) might be *Mary* (following Chomsky 1973 and much subsequent work). *Mary* does not lie along the path between *John* and the first instance of *himself*; if it did, that dependency would be ill-formed and would still fail to trigger the PMC[17]. Thus, the locality constraint on local anaphors fails to show PMC effects at all. However, we can derive this result from properties of the relevant constraint, rather than simply stipulating it.

The PMC apparently needs to be able to distinguish between two classes of constraints. In one class, which includes the constraints discussed in sections 2.1-2.4, the constraint is on dependencies of a certain kind, and the PMC can circumvent the effects of the constraint by allowing us to ignore parts of the structure which identify the dependency as a member of the relevant kind. In the second class, the constraint is on dependent elements of a certain kind, and requires them to be associated with appropriate licensers; constraints of this type include the requirement that anaphors be bound, as well as the one responsible for Connectedness effects (section 2.5). Here we have to limit the power of the PMC to remedy violations; triggering the PMC with respect to the licenser cannot make structures legitimate. The version of the PMC in (6) is intended to be capable of making this distinction; in the second class of cases, ignoring the licenser will not improve the structure, because the constraint in question demands a licenser[18]. At the same time, as we have seen, it is in principle possible to remedy violations of such constraints using the PMC, as in the Connectedness case. In other words, it is not simply that the PMC arbitrarily fails to apply to certain constraints.

3. PMC and Shortest Move

Let us now see how we might apply the PMC to Shortest Move. In the following section I will be assuming a version of the theory of Feature-Movement proposed by Chomsky (1995). In particular, I will adopt the notion that overt movements create a number of related but distinct chains, including feature-chains, which drive the process of movement, and category-chains, which are created by pied-piping of categories along with moved features[19].

(26) What$_i$ is he talking about t$_i$?

For example, in a wh-movement case like (26), I will follow Chomsky (1995) in assuming that the element which must move for the derivation to converge is actually the [+wh] feature on the wh-word; this feature moves into CP, forming a wh-feature-chain. However, for morphological reasons, a wh-feature cannot move in English without pied-piping the entire wh-word (just as, in the Italian equivalent of (26), the wh-word could not move by itself, but would have to pied-pipe the PP containing it). Thus, the wh-word also raises into CP, forming a category-chain headed by the word *what*. I will also need to assume that Shortest Move is sensitive both to movable elements and to potential landing sites (contra Chomsky 1995; cf. Murasugi 1992, Ochi to appear, Bo§koviŒ to appear for approaches along these lines); that is, that (27) is a violation of Shortest Move whether X is a movable element that could have moved to A or a possible landing site for the kind of movement involved in movement from B to A.

```
(27)    A     X          B
        :
        z-----m
```

The particular version of Shortest Move I assume is based on Chomsky's (1995) notion of Attract-F, which says that features move as a result of being "attracted" by other features, which force them to raise to a structurally higher position.

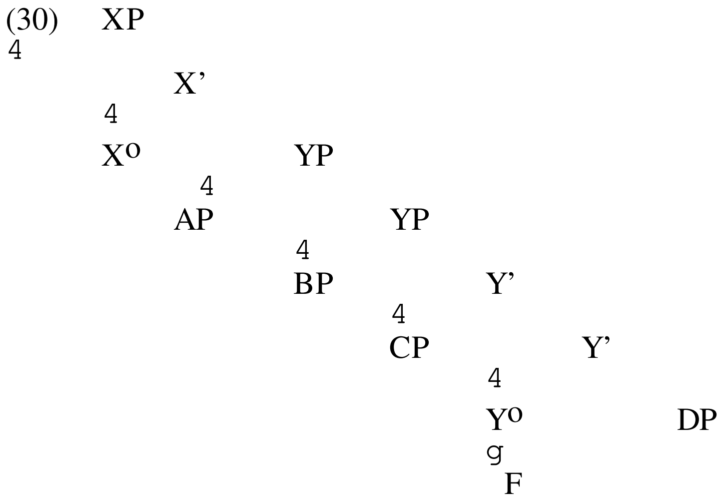
(28) A feature F must attract another feature G, such that G's minimal domain is not separated from F by any feature which could participate in this attraction relation (either as an attractor or as an attractee).

The relevant notion of minimal domain, adapting somewhat the definition given in Chomsky (1995, 299), will be defined as follows, for a feature F:

(29)
a. Max (F) is the smallest maximal projection including F.
b. The *domain* $\delta(F)$ of F is the set of categories included in Max(F).
c. The *minimal domain*Min($\delta(F)$) of F is the smallest subset K of $\delta(F)$ such that for any $\gamma \in \delta(F)$, some $\beta \in K$ (reflexively) dominates γ.

The meanings of the definitions in (29) may be illustrated using the tree in (30), involving a maximal projection YP whose head contains the feature F:

(30)



In (30), Max(F) is YP, and the domain of F is the set of categories contained in YP; namely, AP, BP, CP, DP, Y^o, and all the categories dominated by them. The minimal domain is the smallest subset K of the domain such that for every member of K, some member of the domain dominates it (perhaps reflexively). In other words, the minimal domain is AP, BP, CP, DP, and Y^o.

The definition of minimal domain given in (29) differs from Chomsky's in two regards. One is that it is defined only for features; this is primarily for expository convenience. The other is that the requirement that a category not contain F in order to be in F's domain has been dropped. This means that the head containing F will be a member of F's domain (as will the specifiers of this head, its complement, and any adjuncts to the head's maximal projection).

The definition of Shortest Move given here casts the crucial relation as one between an attracting feature and a minimal domain. The relevant elements, then, will be the attracting feature and the minimal domain of the tail of the chain. As the constraint is stated, there are two possible ways for the PMC to remedy a Shortest Move violation. One would be to ignore the feature which is being skipped by the move (henceforth, the "obstacle"). Another would be to ignore the head of the chain; if the attracting feature is invisible, then Shortest Move does not apply (since it is a constraint on the features that attract other features). Note that the constraint is phrased in such a way that ignoring the tail of the chain will never help; an attracting feature must attract a feature (that is, for every attracting feature there must be some attracted feature), so ignoring the tail of a chain will simply render the structure ill-formed. This guards against, among other things, the prediction that any operation will render a movable element completely immune to Shortest Move. Suppose we imagine a case in which the PMC might have marked an XP ignorable with respect to shortest move, and attempt to take advantage of this by moving the XP too far:

(31)
```
        [ XP
   :          1
   z------m
```

If we try to use the PMC to make the ill-formed structure in (31) legitimate by ignoring XP, we will violate Shortest Move as it is stated in (28); the attractor in question will effectively not be attracting a feature.

There is a further asymmetry between the head and the tail of a chain; the entire minimal domain of the tail is relevant, but only the head of the chain itself, and not its minimal domain, is affected by the PMC. We will see evidence for this asymmetry in the sections on Holmberg's Generalization and on Icelandic Stylistic Fronting.

In principle, then, there are at least two ways that the PMC could save a representation like that in (27); it could allow us to ignore the obstacle X, or it could allow us to ignore A (that is, the head of the chain). Let us consider each of these possibilities in turn.

3.1 Ignoring the obstacle

As we will see, there are two ways of ignoring the obstacle. One would be to perform a Shortest Move past the obstacle, starting from a position in its minimal domain. Since the minimal domain of the tail of a chain is relevant for Shortest Move, this would render the obstacle invisible for Shortest Move and allow movement to take place past it. The other way would be to involve the obstacle itself in a Shortest Move. Given that both the head and the tail of a move are relevant for evaluating the chain for Shortest Move, this will render both the head and the tail invisible for purposes of evaluating the rest of the structure for Shortest Move.

3.1.1 Shortest Movement past the obstacle

3.1.1.1 Holmberg's Generalization

One type of Shortest Move that could enable us to ignore the obstacle for Shortest Move would be a well-formed head-movement past the obstacle. This move is diagrammed in (32) (where the obstacle BP is in boldface):

(32)

```
            XP
     4
     AP              X'
            4
           Xo              YP
     ry              4
     Xo     Yio     BP             Y'
            :               4
            z-----          ti             ZP
```

In (32) head-movement of Y^o obeys Shortest Move, and we can therefore ignore everything in the minimal domain containing the tail of this move. Since BP is in the minimal domain of the tail of the head-movement, we can now ignore BP for purposes of computing Shortest Move violations. We could then perform a move from inside ZP to Spec XP, even if BP were a potential landing site for this move or contained an element which could in principle move to AP. This result has obvious similarities to Chomsky's (1993) Equidistance and Baker's (1988) Government Transparency Corollary. We may use the PMC, then, to derive Holmberg's (1986) Generalization. Here movement of the verb (corresponding to Y^o, above) licenses movement of the object past some obstacle (perhaps the base-position of the subject, as in Chomsky (1993)) (Icelandic, from Jonas and Bobaljik (1993:93)):

(33) a. * Jólasveinarnir hafa búDinginn$_i$ ekki borDuDu t$_i$.

Christmas-trolls have pudding-the not eaten

```
                                :                    1
                                z----------m
```

'The Christmas-trolls haven't eaten the pudding'

b. Jólasveinarnir borDuDu$_i$ búDinginn$_j$ ekki t$_i$ t$_j$.

Christmas-trolls ate pudding-the not

```
                       :          :     1    1
                       1          z--+--m
                       z----------m
```

'The Christmas-trolls didn't eat the pudding'

Note that no other kind of movement past an XP obstacle will help at all. Any movement past an obstacle that could in principle involve the obstacle will violate Shortest Move, and the obstacle will be irrelevant to any movement past it that could not involve it; thus, the PMC will never mark an XP obstacle as a result of XP movement past it.

The PMC-based theory of Shortest Move predicts that head-movement over a specifier position will allow movement past that position to the next available specifier. If AP in (32) is not a possible landing site for the kind of movement in question, the move need not stop there, but can move to the next specifier which is an available landing site. Movement licensed in this way cannot, however, be to an arbitrarily distant specifier; head-movement is simply rendering a potential landing site invisible, allowing movement to skip that site and move on to the next one. This is a looser requirement than Chomsky's (1993) Equidistance, and is closer in flavor to Baker's (1988) GTC.

In Chomsky's (1993) version of Equidistance, head-movement only permits skipping of the specifier of the tail of the movement chain. The specifier of the head of the chain is still visible for Shortest Move. This asymmetry between the head and the tail of the chain is reflected in this theory in the definition of Shortest Move; the relevant relation is one between an attracting feature and a minimal domain, so the minimal domain of the tail of the chain, but not that of the head, is relevant for evaluating the constraint.

3.1.1.2 Icelandic Stylistic Fronting

We have just seen a case in which a head-movement renders the specifier of the tail of its chain invisible for Shortest Move. If minimal domains contain heads as well as XPs[20], then we expect to also find cases in which an XP-movement renders the X^o of the minimal domain containing the tail of its chain invisible for Shortest Move. A phenomenon in Icelandic known as "stylistic fronting" (cf. Platzack 1987, Maling 1990, Rögnvaldsson and Thráinsson 1990, Jónsson 1991, Poole 1996 for discussion) appears to be an example of this kind. Various kinds of heads in Icelandic can front past the tensed auxiliary, just in case the subject has undergone some kind of movement. Fronting can affect participles, predicate adjectives, or certain kinds of adverbs[21], including negation:

(34) a. Hún benti á y msa roskna trésmiDi í bœnum,
she pointed to various old carpenters in town
sem **byrjaD**$_i$ höfDu t_i meD engum efnum.
that begun had with nothing

b. ∏etta er glœpamaDurinn sem **ekki**$_i$ hefur t_i veriD dœmdur.
this is criminal-the that not had been convicted

c. Hann fann mynd sem **vandlega**$_i$ hafDi t_i veriD falin.
he found picture that carefully had been hidden

Stylistic Fronting is only possible if the subject has undergone some kind of movement; it cannot appear in its in-situ position, as in (34a), but can be extracted, as in (34b), or left in a lower position in the tree, as in (34c), a possibility which is only open to indefinites. Stylistic fronting is also possible with impersonal passives; here the subject is phonologically null, so its position is difficult to determine:

(35) a. * Hún benti á bœinn, πar sem **byrjaD**$_i$ höfDu t_i

she pointed to town where that begun had

trésmiDirnir meD engum efnum.

carpenters-the with nothing

b. Hún benti á y msa roskna trésmiDi í bœnum,

she pointed to various old carpenters in town

sem **byrjaD**$_i$ höfDu t_i meD engum efnum.

that begun had with nothing

c. Hún benti á bœinn, πar sem **byrjaD**$_i$ höfDu t_i

she pointed to town where that begun had

nokkrir trésmiDirnir meD engum efnum.

some carpenters with nothing

d. ∏aD var hœtt aD rigna πegar **komiD**$_i$ var t_i πangaD.

it was stopped to rain when arrived was thither

If a sentence contains multiple heads that could in principle undergo Stylistic Fronting, only the highest can be fronted (Jónsson 1991, 5):

(36) a. ∏etta er aDferD sem **ekki**$_i$ hefur t_i veriD reynd á mönnum.

this is method that not has been tried on people

b. * ∏etta er aDferD sem **reynd**$_i$ hefur ekki veriD t_i á mönnum.

this is method that tried has not been on people

In other words, extraction of the subject seems to render the inflected auxiliary transparent for head-movement, but not any of the lower heads.

Icelandic Stylistic Fronting, then, appears to have all the hallmarks of a case of Minimal Compliance. An overly long head-movement is possible just in case the subject (presumably the element in the specifier of the head being skipped) undergoes a Shortest

Move. What we have here, in other words, seems to be the converse of the Holmberg's Generalization case; an XP-movement marks the minimal domain of which it was a member as ignorable for Shortest Move, so the head of that minimal domain can be skipped thenceforth by head-movement.

If we assume some version of the Internal Subject Hypothesis, the Icelandic data again give us evidence for an asymmetry between the head and the tail of a movement chain with respect to the PMC. The claim developed here is that movement of V^o over T^o is only possible in (37) when the subject has been extracted from Spec TP, thus rendering the minimal domain TP (including its head, T^o) invisible for Shortest Move:

(37) $[_{CP}$ $[_{TP}$ SUBJ T^o .... V^o

```
      :            1          1
      z-------+------m
<----------      m
```

If the Internal Subject Hypothesis is correct, however, the subject in (37) is not in its base-position; it has raised out of some internal subject position (perhaps Spec VP). Thus, even in cases in which the subject is not extracted, it has undergone some move (which presumably obeys Shortest Move) to get to Spec TP. This move is not enough by itself, however, to license Stylistic Fronting; the subject must undergo further extraction, as we have seen. In other words, the minimal domain TP is rendered invisible for Shortest Move only when it contains the tail of a movement chain, not when it contains the head of the chain[22]. This was also the case with Holmberg's Generalization, as we saw in the previous section; together, these two phenomena give some support for the particular definition of Shortest Move adopted above, incorporating an important asymmetry between the head and the tail of a movement chain.

There are some cases in which a Stylistic Fronting fronts a participle over several auxiliaries (Jónsson 1991, 5):

(38) ∏etta er versta bók sem **skrifuD**$_i$ hefur veriD t_i.

this is worst book that written has been

If we assume, following Mitchell (1993) and Demuth and Gruber (1994), that the subject moves through the specifiers of all of the auxiliaries, the well-formedness of (38) will be unproblematic; all of the specifiers of all of the auxiliaries will contain tails of Shortest Moves of XPs, and the auxiliaries will therefore be ignorable for Shortest Move. Sentences of this kind raise interesting issues about the obligatoriness of the PMC, which space limitations prevent me from addressing here.

3.1.2 Involving the obstacle in a Shortest Move; Path Containment Condition

The other way of using the PMC to ignore an obstacle would be to involve the obstacle itself in a Shortest Move; this will allow us to ignore both the head and the tail of the chain created by the obstacle, and the move from B to A in (39) can then proceed with impunity:

(39)
```
A    X1      X2     B
:    :
1    z--m           1
z----------m
```

On the assumption that Shortest Move is sensitive both to landing sites and to movable elements, the paths will have to be nested, as shown, for the PMC to save the structure. If X_1 were above A, the movement from X_2 to X_1 would cross a possible landing site (namely A), thereby violating Shortest Move, and the PMC would not be able to license ignoring X_1 and X_2 with respect to movement from B to A. If X_2 were below B, movement from X_2 to X_1 would cross a movable element (namely B), again violating Shortest Move, and the PMC would again be irrelevant[23]. Thus, the PMC interacts with Shortest Move to guarantee that any intersecting moves that are driven by the same feature will have to nest rather than crossing. This is basically Pesetsky's (1982) Path Containment Condition.

Pesetsky (1982) also discusses a class of sentences which contain a total of four variables (from Pesetsky (1982, 430-1)):

(40) a. ?? This Volvo is one car OP_i that I know who_j

[people that talk to e_j about e_i] can persuade e'_j to buy e'_i.

b. * This Volvo is one car OP_i that I know who_j

[people that talk about e_i to e_j] can persuade e'_j to buy e'_i.

Here, as Pesetsky points out, the Path Containment Condition must effectively be obeyed both in the subordinate clause complement of *persuade* and in the relative clause modifying *people*. Depending on how we interpret Shortest Move, the PMC will lead us to expect this effect as well; here, we apparently need to understand Shortest Move as being satisfied in cases where several variables are bound by a single binder only if all the binder-variable links satisfy Shortest Move. That is, Shortest Move might be restated as in (41):

(41) A feature F must attract the maximum possible set of features G such that the minimal domains containing the elements of G are not separated from F by any feature which could participate in this attraction relation (either as an attractor or as an attractee) but fails to do so.

In (40a), then, the relations between *who*$_j$ and its variables both satisfy Shortest Move, thus rendering *who* ignorable for Shortest Move and allowing construction of variable-binding links that cross the position of *who*. In (40b), on the other hand, the links between the binders and their variables within the relative clause both violate Shortest Move, and the structure is ill-formed. This phenomenon bears an obvious similarity to the behavior of Subjacency discussed above; as Chomsky (1986) notes, Subjacency must be obeyed both in the main clause and by the parasitic gap:

(42) a. * Who_i did John hire t_i [without asking who had interviewed t_i]?

b. * Who_i did John ask [before hiring t_i] who had interviewed t_i?

As (42) shows, a sentence involving a parasitic gap is ill-formed if the link between the wh-word and either of the gaps crosses a wh-island; this is true whether the wh-island intervenes between the wh-word and the parasitic gap only, as in (42a), or between the wh-word and the main gap, as in (42b). If the main dependency and the parasitic dependency are separate dependencies (which they crucially are for purposes of the CED, in this theory; see section 2.5), then the ill-formedness of these sentences is surprising; we ought to expect the Subjacency-obeying dependency to rescue the Subjacency-violating dependency.

In section 2.6, I suggested that the ill-formedness of the sentences in (42) might be handled by stating Subjacency as a constraint on a [+wh] complementizer's relations to all wh-words with a given index, rather than simply to a given wh-word. As in the case of Shortest Move, then, the suggestion is that a set of coindexed dependencies ought to undergo the constraints on dependencies as a unit (as in Kayne's (1983) notion of G-projection), rather than applying the constraints to each dependency separately[24].

Another possibility, following Chomsky 1986, would be to claim that parasitic gaps always involve a second operator base-generated in the position of the parasitic gap. The fact that this operator must obey Shortest Move, then, is not surprising. Such a claim obviously involves giving up on the account developed here of the Connectedness effects. The behavior of the Path Containment Condition provides us with an argument against this approach. Consider (40b), repeated as (43):

(43) * This Volvo is one car OP_i that I know who_j

[OP'_i OP'_j people that talk about e'_i to e'_j] can persuade e_j to buy e_i.

In Chomsky's (1986) approach, sentences like (43) contain four operators, rather than two. The operators OP'_i and OP'_j inside the subject NP, which are responsible for the creation of the parasitic gaps, violate the Path Containment Condition, and the structure is therefore ill-formed. If we develop an account of this kind, however, we will need to add a stipulation to ensure that the parasitic-gap-operators do indeed violate the PCC. There is no obvious reason for this to be so; that is, there is no reason for (43) to have the structure in (44a) rather than that in (44b):

(44) a. This Volvo is one car OP_i that I know who_j

[OP'_i OP'_j people that talk about e'_i to e'_j] can persuade e_j to buy e_i.

b. This Volvo is one car OP_i that I know who_j

[OP'_j OP'_i people that talk about e'_i to e'_j] can persuade e_j to buy e_i.

The PCC is violated in (44a), but not in (44b). An account along these lines must apparently stipulate that the parasitic-gap operators must enter into the same hierarchical relation to each other as the operators on which they are parasitic. On the account developed here, on the other hand, (43) has only two operators, rather than four. The fact that the operators binding the e' variables and the operators binding the e variables must be in the same hierarchical relation to each other is now unsurprising, as these operators are identical. The particular redefinition of Shortest Move developed in this section, and the parallel redefinition of Subjacency developed above, thus seem well-motivated.

3.2 Ignoring the head of the chain

Thus far, we have been considering ways of redeeming the Shortest Move violation represented in (45) that involved ignoring the obstacle:

```
(45)  A   X     B
      :
      z------m
```

Now let us consider the possible consequences of ignoring the head of the chain for Shortest Move. We cannot allow this mechanism to work too liberally. We do not want to be able, for example, to save Super-Raising cases by wh-moving the super-raised NP:

(46) * Who_i t'_i seems that it is likely t_i to win?

The Shortest Move connecting *who* and t' cannot allow us to ignore t', thereby saving the Shortest Move-violating dependency between t' and t.

Here it will be useful to make use of Chomsky's (1995) notion of feature-movement; in particular, we will want to make use of the distinction between interpretable and uninterpretable features. In (46), the feature Case drives raising of the nominal *who* from t_i to t'_i, violating Shortest Move[25]. This Case-chain, being headed by an uninterpretable feature, is checked off and vanishes as a result of this movement. When a feature-chain violates Shortest Move and is then eliminated by checking immediately after being formed, it seems reasonable to claim that the chain cannot be saved by any subsequent process. In the Super-raising case above, for example, the feature Case forms a chain and is checked off; later movement cannot save this chain, since the vanished feature cannot participate in movement.

If this is on the right track, then, the PMC will not be able to save the configuration in (45) if the obstacle has not been rendered ignorable and an uninterpretable feature-chain created by the movement violates Shortest Move and is then checked off. On the other hand, the PMC might be active in a case in which a chain is headed by an interpretable feature; such features, by hypothesis, do not vanish under checking, and might be able to participate in further moves that would trigger the PMC and render the structure well-formed.

3.2.1 French Participle Agreement

Participle agreement appears to involve movement triggered by attraction of an (interpretable) ϕ-feature on a noun. This is taken in Chomsky (1995) to explain the well-formedness of sentences like (47) (Italian), which involve agreement on multiple participles:

(47) Gli Australiani sono stati creduti.
the Australians are been-PL believed-PL

If participle agreement involved an uninterpretable feature like Case, it would be odd to find multiple participle agreement, since the Case feature is supposed to vanish after being checked. If, on the other hand, the feature on the noun that enters into a checking relation is interpretable, then the fact that it can enter into several checking relations is not surprising.

French participles apparently may agree with their objects just in case the object undergoes some kind of movement (for discussion of these facts, see Kayne 1989 and Branigan 1992, among many others):

(48) a. * Josèphe a écrite cette lettre.
Joseph has written-FEM this-FEM letter

b. Quelle lettre Josèphe a-t-il écrite?
what-FEM letter Joseph has-he written-FEM

c. la lettre que Josèphe a écrite hier
the-FEM letter that Joseph has written-FEM yesterday

d. Cette lettre, Josèphe l'a écrite hier.
this letter Joseph her-has written-FEM yesterday

e. Les lettres ont tous été écrites.
the-PL letters have all been written-PL

On the assumption that triggering participial agreement involves movement to the specifier of some kind of AgrP projection, these seem to be the kind of cases we need; movement driven by an interpretable feature must be licensed by further movement of the NP triggering agreement. Suppose movement of the object to the position where participle agreement is checked violates Shortest Move; this might be because such movement crosses the base position of the subject, or perhaps the object's own case-checking position (that is, AgrParticipleP might dominate AgrOP)[26]. If this is the case, then the data in (47) are not surprising; the interpretable ϕ-feature driving movement violates Shortest Move, but the head of the chain that violates Shortest Move is then rendered innocuous by a subsequent Shortest Move[27]. We thus arrive at a theory of why agreement on French participles is only licensed by an object which enters the specifier of the AgrP connected with participial agreement but then moves on to some other position[28]. This account can probably be extended to deal with similar facts from Italian, as we will see below:

(49) a. * Ho creduti gli Australiani.
have-1SG believed-PL the Australians

b. Li ho creduti.
them have-1SG believed-PL

c. Gli Australiani sono stati creduti.
the Australians are been-PL believed-PL

3.2.2. postscript to French participles: Greedy features?

The account is not without its difficulties, however. One problem has to do with possible illicit uses of this technique to license Super-raising. We might expect sentences like (50), for instance, to have a well-formed derivation involving Super-raising (Italian, Caterina Donati, personal communication):

(50) *Maria è t sembrata che paia t' intelligente
Maria is seemed-FEM.SG. that appears-SUBJ intelligent

'Maria seemed that it appears intelligent'

Consider the derivation of a sentence like (50). The category *Maria*, along with its ϕ-features and Case-feature, moves from t' to t, the specifier associated with participial agreement on *sembrata* 'seemed'. The ϕ-features and Case-feature thereby violate Shortest Move, but neither is checked off; the ϕ-features are interpretable, and this movement, by hypothesis, is not one in which the Case-feature enters into a checking relation. Next *Maria* will move from t to its surface position, a Shortest Move-obeying move that allows us to ignore the head of the Shortest Move-violating chain just formed. Thus, the sentence should be well-formed, contrary to fact.

The above account depends crucially on the claim that the Case feature of the NP moves along as a "free rider" when the NP moves to the specifier involved in participial agreement. If this is not the case, then the above story collapses. Thus, we might believe in a class of "greedy" features which never move unless they will enter into a checking relation as a result of movement[29].

Postulating a class of greedy features will also help us deal with the Italian participial agreement paradigm in (51):

(51) a. * Ho creduti gli Australiani.
have-1SG believed-PL the Australians

b. Li ho creduti.
them have-1SG believed-PL

c. Gli Australiani sono stati creduti.
the Australians are been-PL believed-PL

d. * Quanti Australiani hai creduti?
how-many Australians have-2SG believed-PL

Italian participial agreement, like French participial agreement, is only licensed by movement; a participle cannot agree with an object left in situ, as in (51a). Italian and French differ, however, in that Italian participles also cannot agree with wh-moved objects, as in (51d). Thus, an Italian participle can agree with a direct object which has undergone a movement that involves ϕ-feature checking, as in (51c), but not one that clearly does not involve ϕ-feature checking, as in (51d) (the motivation for cliticization is unclear, but might, for all we know, have something to do with ϕ-feature checking). We might try to capture this by claiming that Italian ϕ-features differ from French ϕ-features in that the former are greedy, only moving when they enter into a checking relation as a result of movement[30]. Thus, in (51b-c), the ϕ-feature chain associated with the object crosses the internal subject position on the way to the specifier associated with participial agreement and violates Shortest Move, as in French. It then undergoes a Shortest Move that allows us to ignore the head of the Shortest Move-violating chain, and the resulting structure is well-formed by virtue of the Principle of Minimal Compliance. In (51d), on the other hand, the ϕ-feature chain violates Shortest Move in order to trigger participial agreement, as before, and the direct object then undergoes wh-movement; because Italian ϕ-features are greedy, however, and this second movement does not involve checking of ϕ-features, the ϕ-features fail to raise along with the wh-word, and the Principle of Minimal Compliance is irrelevant.

The notion of a class of greedy features thus appears to be a rather useful one. Many questions now arise, of course. Are all uninterpretable features greedy? Does the claim that Italian ϕ-features are greedy while French ϕ-features are not make any predictions outside the data discussed here[31]? Answering these questions is, unfortunately, beyond the scope of this paper.

4. Conclusion

We have seen that we can unify a number of apparently diverse phenomena under the rubric of Minimal Compliance. It may be a general property of human language that constraints need not be satisfied perfectly in all parts of a given structure for that structure to be well-formed. Recognition of the Principle of Minimal Compliance as an independent fact about human language thus promises to simplify considerably our statements about the permissible relations among syntactic elements.

If the theory sketched above is right, we are also driven by acceptance of the PMC to certain conclusions about the structure of the grammar. In particular, we must apparently view the grammar in derivational rather than representational terms; this seems to yield the best analysis of certain facts about Subjacency, as we have seen. Furthermore, as developed above, the PMC entails a view of the grammar in which it contains particular constraints that must be checked. The fact that the PMC interacts in different ways with different constraints (see sections 2.1 and 2.6, in particular, for some discussion of this) suggests that these constraints are in fact formally distinct in some way. I have encoded this distinction here in the statements of the constraints themselves; one can imagine other ways of attempting to encode it (for instance, by having more general constraints which have different effects on different structures, as suggested in footnote 18), but I will have to leave this problem, along with many others, for future research[32].

I am very grateful to Yoo-Kyung Baek, Rajesh Bhatt, Jonathan Bobaljik, David Braun, Andrew Carnie, Noam Chomsky, Kai von Fintel, Danny Fox, Paul Hagstrom, Heidi Harley, Roumyana Izvorski, Martha McGinnis, Shigeru Miyagawa, Rob Pensalfini, David Pesetsky, Colin Phillips, Uli Sauerland, Susi Wurmbrand, and audiences at SCIL VII and CONSOLE 4 for much helpful discussion. For help with particular languages I would like to thank Roumyana Izvorski and Roumyana Slabakova for their Bulgarian judgments, and Caterina Donati for help with Italian. Reponsibility for any remaining errors is mine alone. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship, for which I am also grateful.

[1]Here the material in parentheses corresponds to the elided VP.

[2]We might think of this as an application of Economy to the process of constraint-checking.

[3]Throughout this paper I will use "head" and "tail" for dependencies generally, not only for movement chains. This is simply for expository convenience.

[4]We will obviously want a less language-specific principle to deal with the phenomena in (8), but (9) will suffice for our purposes. For an attempt to derive (9) from general principles of Economy, see Richards 1995.

[5]We can also ignore *zichzelf*, of course, although this is not relevant here.

[6]But see Veraart (1996) for another possible way of understanding these facts.

[7]This need not be as conceptually awkward as it might seem. For instance, a theory of anaphor binding in which the various constraints on anaphors follow from general constraints on dependencies

which are not specific to anaphors (for instance, in which the locality conditions follow from Shortest Move, while the matter of selection among different anaphoric forms is left to Structural Economy) would have this character.

[8]For arguments that reciprocals are or crucially involve bound variables, see Heim, Lasnik, and May 1991, Sauerland 1994.

[9]Note that there may be a difference between Subjacency and Reflexivity in this regard. (i) and (ii) are both well-formed (Reinhart and Reuland 1993, 668):

i. Henk wees zich aan zichzelf toe
Henk assigned self to self-self

ii. Henk wees zichzelf aan zich toe
Henk assigned self-self to self

We must apparently assume either that the constraint driving the Reflexivity effects is purely representational, or that the formation of the dependencies to which the constraint applies may occur in any order (these accounts are identical, as far as I can see, for the data considered here).

[10]We also expect to find the effects in (17) in languages such as Chinese, where they do not seem to appear. See Tsai (1994) and Richards (to appear) for some possible accounts of the absence of island effects in Chinese which do not make reference to the level at which movement takes place.

[11]A reviewer points out that Japanese (like most wh-in-situ languages) does not exhibit Complex NP Constraint effects for movement of argumental wh-words. I have no account of this fact to offer. Note that the absence of CNPC effects is not especially problematic for the theory developed here; we would be more surprised to find a language in which CNPC effects were present and unaffected by the PMC (see (18) below for a case of CNPC effects which obey the PMC).

[12]As one reviewer notes, the order of the wh-words in (18b) is the opposite of that expected on standard assumptions about how multiple adjunction works. The theory here requires *koj senator* ‘which senator’ to move first, triggering the PMC and allowing *koja kniga* ‘which book’ to move in a manner

which violates Subjacency. The resulting derivation appears anti-cyclic. This is a general problem with wh-movement in languages of the Bulgarian type. For instance, multiple fronted wh-words must typically conform to an order which is the opposite of that predicted by Superiority (Bulgarian, from Bo§koviŒ 1995, 3):

i. a. Koj kogo e vidjal

who whom AUX seen

'Who saw whom?'

b. *Kogo koj e vidjal

whom who AUX seen

See Rudin 1988, Bo§koviŒ 1995, Richards to appear for further discussion.

[13]In Richards (to appear) I argue further against approaches of this type, drawing attention to parallels between wh-in-situ languages and languages with multiple overt wh-movement which would be surprising if wh-in-situ received its scope by a mechanism with nothing in common with syntactic movement.

[14]Another possibility, following Chomsky 1986, would be to postulate a distinct operator inside the adjunct clause, which must obey Subjacency separately; this would involve relinquishing the account developed in section 2.5 of the Connectedness effects. See section 3.1.2 for a similar problem raised by Shortest Move, along with an argument against an approach which postulates distinct operators.

[15]This move apparently requires us to understand the PMC as applying optionally.

[16]Brody (1995) claims that Subjacency and Connectedness do not differ in this regard, on the basis of sentences like (i) (from Brody 1995, 56):

i. *Who$_i$ t$_i$ was against [proposals to leave [without waiting for who$_j$]]?

(i) involves a LF-moved wh-word which is embedded in both a complex NP and an adjunct island. As Brody (1995) points out, the sentence is ill-formed, and Brody concludes from this that in both the

Connectedness and the Subjacency cases, only a single island can be rendered transparent by a well-formed dependency.

In order to make this move, however, Brody collapses the categories of Subjacency islands (e.g., wh-islands and complex NPs) and CED islands (any ungoverned XP, including adjuncts and subjects). LF-movement of *who* in (i) crosses one of each type of island, and the CED island is not along the path of the well-formed dependency. The ill-formedness of (i) is therefore unsurprising, if we maintain a distinction between these two types of islands. The well-formedness of (22b), which Brody's story fails to predict, also follows straightforwardly.

[17]If this approach is on the right track, it will place certain constraints on our theories about what the binding domain for a local anaphor is. If, for instance, the embedded CP in (23c) c-commands *himself*, then this CP cannot be the binding domain, or we would expect the PMC to save the structure.

[18]In principle, it might be possible to distinguish between these two classes of cases, not by assuming two different classes of constraints, but by assuming general constraints which apply to different structures in different ways. Distinguishing between an approach of this kind and the one developed in this paper would not be straightforward; we would need to find a case in which the PMC interacts with a constraint differently depending on what kind of structure is being constrained, and try to show that the differences in structure can account for the differences in behavior of the PMC. One such case might be the distinction between head-movement and XP-movement with respect to Pesetsky's (1982) Path Containment Condition (discussed in section 3.1.2); PCC effects are apparently found with XP-movement, but not with head-movement (a fact for which I have no account). I will not try to develop this proposal further here, however.

[19]Chomsky (1995) claims that covert movement only involves movement of features, without pied-piping of a category. Nothing in the following discussion will interact crucially with this assumption, as far as I can see.

[20]contra Chomsky 1993, but in keeping with Chomsky's (1995) reinterpretation of checking domains as the checking domains of features rather than of heads.

[21]It is not clear to me at this point why adverbs should behave this way, on the common assumption that adverbial heads do not participate in head-movement. Perhaps this is (rather weak) evidence that the common assumption is wrong.

[22]On the account developed here, we predict that movement of the subject from the internal subject position to the external position should render the head of which it is base-generated as a specifier transparent to head-movement. The identity of this head is sufficiently disputed to make this prediction difficult to test.

[23]If we were to assume Cyclicity, and if Shortest Move is evaluated at the point at which the move is performed, so that movement of the obstacle must precede movement from B to A, then we would no longer need to assume that Shortest Move is sensitive both to potential movers and to potential landing sites--it could be sensitive to possible movers only. If movement of the obstacle has to precede movement from B to A, then the derivation in which X_1 is above A is ruled out by Cyclicity. Thus, the only derivation that Shortest Move needs to rule out is the one in which X_2 is below B. In other words, movement across possible movers needs to violate Shortest Move, but we no longer need movement across possible landing sites to violate Shortest Move. This would make our definition of Shortest Move more similar to the one adopted in Chomsky 1995; we would only need to postulate Shortest Attract.

[24]This might be implemented either by understanding Subjacency and Shortest Move as representational constraints (but see section 2.4 for an argument against this), or by assuming a form of across-the-board movement in parasitic gap constructions (cf. Haïk 1985, Huybregts and van Riemsdijk 1985, and Williams 1990 for approaches to parasitic gaps along these lines), so that derivational versions of Subjacency and Shortest Move can apply to the movement relations connecting a single binder to its multiple variables.

[25]As a reviewer points out, Chomsky (1995) actually claims that the feature driving this movement is the (interpretable) D-feature, which checks an EPP feature on T; Case merely raises as a "free rider". For the account developed here to work, we will need to assume either that Case can in fact drive movements like the one in (45) or that free riders are themselves subject to Shortest Move (perhaps only when they enter into feature-checking relations).

[26]On the latter story, we would need to assume that this movement leaves behind the Case features of the moving object.

[27]This account seems at first glance to fall afoul of the principle proposed in section 2.4, in the discussion of Subjacency, to the effect that a move which is rendered well-formed only by the PMC must follow the PMC-triggering move which licenses it. This was taken to account for the fact that in languages like English, a Subjacency-violating move cannot be licensed by a Subjacency-obeying move to the same Spec CP at LF; the Subjacency-obeying move must take place first, in the overt syntax. In the French participle case, however, the move which triggers the PMC appears to follow the move which must be licensed by the PMC. We might explain this by appealing to another difference between the two cases, namely that the Subjacency case involves two separate chains while the French participle case involves two links in a single chain. If we assume something like Form Chain (Chomsky 1993, Chomsky and Lasnik 1993), so that all the links of a chain are formed simultaneously, we will make the necessary distinction between the two cases, and can continue to maintain that a move cannot precede any PMC-triggering moves which license it.

[28]We might consider trying to extend this theory to other cases in which agreement cannot cooccur with a lexically overt subject in its specifier (e.g., prepositional agreement in Celtic (McCloskey and Hale 1983) and verbal agreement in Mohawk (Baker 1996)). Similarly, Postal's (1974) Derived Object Condition cases ought to be amenable to an analysis of this kind; these appear to be cases in which ECM must be licensed by further movement of the subject of the lower clause. I will not try to sketch these expansions of the theory here.

[29]A reviewer points out that “greedy” features should be incompatible both with the claim that overt movement always involves pied-piping of all the features of a category, and with the notion of successive-cyclic movement. In neither case, by assumption, would a greedy feature be undergoing checking as a result of movement. It seems to me that both of these problems are surmountable, although a full discussion of them is well beyond the scope of this paper. If the requirement of overt pied-piping follows from a requirement that features be in a single structural location at the point at which the structure is interpreted by the PF interface, then the requirement should only apply to that point in the derivation, and might well be obeyed by greedy features in many cases. As for successive-cyclic movement, we may discover that greedy features never undergo such movement; alternatively, such movement might involve repeated checking of an interpretable greedy feature. I will have to leave these as issues for further investigation.

[30]Another possibility is that ϕ-features are greedy in both Italian and French, and that French differs from Italian in allowing ϕ-feature checking to take place in Spec CP.

[31]For example, one could imagine trying to offer a "feature-geometry"-based account of this, perhaps making reference to the fact that ϕ-features in Italian are morphologically less "fused" with the noun head than their French counterparts (that is, Italian nouns are generally bimorphemic, consisting of a noun stem followed by a vowel indicating the noun's ϕ-features, while French nominal gender is not associated with a particular morpheme).

[32]A reviewer points out that notions like the PMC might be relevant in other areas of the grammar as well. For instance, as the reviewer notes, Chomsky (1964, 1986:39) suggests that processes may only be able to apply once to a given domain, on the basis of the contrast in (i):

(i) a.* Who did you wonder what John [gave t to t]?

b.?? To whom did you wonder what John [gave t t]?

Chomsky notes that (ia) is worse than (ib), and suggests that this might be due to the fact that (ia) involves multiple instances of the same category (NP) being extracted from the same phrase (VP) by the same process (wh-movement). If there is a general principle that processes can only apply once to a given domain, this might well subsume the PMC as it has been developed here. Other phenomena which might fall under this general heading include the "inertness" of intermediate traces (Lasnik and Saito (1984), Chomsky (1995)), which might be seen as an inability on the part of the grammar to accord the same syntactic status to more than a single trace; Cyclicity, which also seems to involve "inertness" of certain portions of the structure which have already undergone certain processes; and the notion of feature checking, whereby a process makes a portion of the structure "invisible" to the computational component. Developing these ideas is well beyond the scope of this paper, however.

Bibliography

Baker, Mark. 1988. *Incorporation: a theory of grammatical function changing*. Chicago: The University of Chicago Press.

Baker, Mark. 1996. *The polysynthesis parameter*. New York: Oxford University Press.

Bo§koviŒ, eljko. 1995. Superiority effects with multiple wh-fronting in Serbo-Croatian. ms., University of Connecticut, Storrs.

Bo§koviŒ, eljko. To appear. Multiple wh-fronting and economy of derivation. To appear in *Proceedings of WCCFL XVI*.

Branigan, Philip. 1992. Subjects and complementizers. Doctoral dissertation, MIT, Cambridge, Mass.

Brody, Michael. 1995. *Lexico-logical form: a radically minimalist theory*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1964. Current issues in linguistic theory. the Hague: Mouton.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, ed. Stephen Anderson and Paul Kiparsky, 232-286. New York: Holt, Rinehart and Winston.

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

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from building 20*, ed. Ken Hale and Samuel Jay Keyser, 1-52. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1995. The minimalist program. Cambridge, Mass.: MIT Press.

Chomsky, Noam, and Howard Lasnik. 1993. Principles and parameters theory. In *Syntax: An international handbook of contemporary research*, eds. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vannemann, 506-569. Berlin: Walter de Gruyter.

Cole, Peter, and Gabriella Hermon. 1994. Is there LF wh-movement? *Linguistic Inquiry* 25.239-262.

Demuth, Katherine, and Jeffrey Gruber. 1994. Constraining XP sequences. Ms., Brown University, Providence, RI, and Université du Québec à Montréal.

Fiengo, Robert, and Robert May. 1994. Indices and identity. Cambridge, Mass.: MIT Press.

Fox, Danny. 1995. Economy and scope. *Natural Language Semantics* 3.283-341.

Haïk, Isabelle. 1985. The syntax of operators. Doctoral dissertation, MIT, Cambridge, Mass.

Heim, Irene, Howard Lasnik, and Robert May. 1991. Reciprocity and plurality. *Linguistic Inquiry* 22.63-101.

Holmberg, Anders. 1986. Word order and syntactic features in the Scandinavian languages and English. Doctoral dissertation, University of Stockholm.

Hornstein, Norbert. 1995. *Logical form: from GB to minimalism*. Oxford: Blackwell.

Huang, James. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation, MIT, Cambridge, Mass.

Huybregts, Reny, and Henk van Riemsdijk. 1985. Parasitic gaps and ATB. In *Proceedings of NELS 15*, 168-187. GLSA, University of Massachusetts, Amherst.

Jonas, Dianne, and Jonathan Bobaljik. 1993. Specs for subjects: the role of TP in Icelandic. In *Papers on case and agreement I: MIT working papers in linguistics 18*, ed. Jonathan Bobaljik and Colin Phillips, 59-98.

Jónsson, Jóhannes. 1991. Stylistic fronting in Icelandic. *Working papers in Scandinavian syntax* 48.1-44.

Kayne, Richard. 1983. Connectedness. *Linguistic Inquiry* 14.223-250.

Kayne, Richard. 1989. Facets of past participle agreement in romance. In *Dialect variation in the theory of grammar*, ed. Paola Beninca, 85-103. Dordrecht: Foris.

Koopman, Hilda, and Dominique Sportiche. 1982. Variables and the bijection principle. *The Linguistic Review* 2.139-160.

Lasnik, Howard, and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15.235-289.

Maling, Joan. 1990. Inversion in embedded clauses in Modern Icelandic. *In Syntax and Semantics 24: Modern Icelandic Syntax*, ed. Joan Malin and Annie Zaenen, 71-91. San Diego: Academic Press, Inc.

McCloskey, James, and Ken Hale. 1983. On the syntax of person-number inflection in Modern Irish. *Natural Language and Linguistic Theory* 1. 487-533.

Mitchell, Erika. 1993. VP-Fronting, do-support, and extended IP in English. Ms., Cornell University, Ithaca, NY.

Murasugi, Kumiko. 1992. Crossing and nested paths: NP movement in accusative and ergative languages. Doctoral dissertation, MIT, Cambridge, Mass.

Ochi, Masao. To appear. Move or attract? To appear in *Proceedings of WCCFL XVI*.

Ouhalla, Jamal. 1996. Remarks on the binding properties of wh-pronouns. *Linguistic Inquiry* 27.676-708.

Pesetsky, David. 1982. Paths and categories. Doctoral dissertation, MIT, Cambridge, Mass.

Pesetsky, David. 1991. Zero syntax II. Ms., MIT, Cambridge, Mass.

Platzack, Christer. 1987. The Scandinavian languages and the null subject parameter. *Natural Language and Linguistic Theory* 5. 377-401.

Poole, Geoffrey. 1996. Optional movement in the minimalist program. *Minimal ideas*, ed. Werner Abraham, Samuel Epstein, Höskuldur Thráinsson, and C. Jan-Wouter Zwart, pp. 199-216. Philadelphia: John Benjamins.

Postal, Paul. 1974. *On raising: one rule of English grammar and its theoretical implications*. Cambridge, Mass.: MIT Press.

Reinhart, Tanya. 1995. Interface strategies. OTS Working Papers, Research Institute for Language and Speech, Utrecht University.

Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24.657-720.

Richards, Norvin. 1995. Towards a theory of head-binding. Ms., MIT, Cambridge, Mass.

Richards, Norvin. To appear. Subjacency forever. To appear in *Proceedings of WECOL '96*.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, Mass.: MIT Press.

Rögnvaldsson, Eiríkur, and Höskuldur Thráinsson. 1990. On Icelandic word order once more. In *Syntax and Semantics 24: Modern Icelandic Syntax*, ed. Joan Maling and Annie Zaenen, 3-40. San Diego: Academic Press, Inc.

Rudin, Catherine. 1988. On multiple questions and multiple wh-fronting. *Natural Language and Linguistic Theory* 6:445-501.

Sag, Ivan. 1976. Deletion and logical form. Doctoral dissertation, MIT, Cambridge, Mass.

Sauerland, Uli. 1994. Codistributivity and reciprocals. ms., MIT, Cambridge, Mass.

Tsai, Wei-tien Dylan. 1994. On economizing the theory of A'-dependencies. Doctoral dissertation, MIT, Cambridge, Mass.

Veraart, Fleur. 1996. On the distribution of Dutch reflexives. Masters Thesis, MIT, Cambridge, Mass.

Watanabe, Akira. 1992. Subjacency and S-structure movement of wh-in-situ. *Journal of East Asian Linguistics* 1:255-291.

Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry* 8.101-139.

Williams, Edwin. 1990. The ATB theory of parasitic gaps. *The Linguistic Review* 6.265-279.

20D-219 MIT

Cambridge, MA 02139

norvin@mit.edu