From Merge and Move to Form Dependency[*]

M. RITA MANZINI

0 Introduction

In this article, I shall first briefly present the minimalist framework that emerges from the recent work of Chomsky (1995). As I shall show, a number of classical problems in movement theory, concerning in particular locality in the sense of Manzini (1992), remain open under minimalism. What is more, while the operation Merge can be argued to be necessary on conceptual grounds alone, there is no analogous motivation for the operation Move, but only considerations of language use. I shall then outline a theory which substitutes Chomsky's (1995) movement rule with a rule of Form Dependency. As I shall show, this rule is sufficient to subsume Merge, thus effectively eliminating the Merge-Move dualism. Finally, I shall indicate how the classical problems in locality theory can also be solved within the version of minimalism that I propose.

1 Merge and Move

In the minimalist model of Chomsky (1995), each linguistic expression is characterized by two representations, a P(honological) F(orm) representation, interfacing with the articulatory-perceptual system(s) and a L(ogical) F(orm) representation, interfacing with the conceptual-intentional system(s). This represents a clear departure from previous models, even within the principles and parameters framework, which are characterized by at least two additional syntactic representations, namely D(eep)-S(structure) and S(urface)-S(tructure). The minimalist model is obviously preferable on conceptual grounds, since in it linguistic representations are seen to simply match the interfaces between language and other cognitive systems. Furthermore, it is compatible with known empirical evidence and arguably superior in dealing with at least some of it.

A central property of the minimalist model is what Chomsky (1995) terms inclusiveness. In other words, not only inputs but also outputs of linguistic

[*]During the first half of 1995, this article formed the basis for talks at the Dipartimento di Scienze Cognitive, Hospital San Raffaele, Milan, at the Università di Firenze and Siena, and at the GLOW Colloquium in Tromsoe. I am grateful to the various audiences involved for valuable feedback.

computations are entirely expressed in terms of properties of lexical items. This implies the impossibility of adding information in the course of the derivation and/or in the output representation in the form of such standard devices of previous models as bar levels, indices, and the like.

The basis of each linguistic computation is represented by a set of lexical items. One of the operations of grammar, Select, applies to this set, selecting one of its members and introducing it into the set of syntactic objects of which a derivation consists at each of its stages. New syntactic objects are formed by the operation Merge, which takes a pair of given syntactic objects (S_i, S_j) and creates a single syntactic object out of them, namely a labelled set (S (S_i, S_j)), where the label S is either S_i or S_j. The operation Move in effect is Merge of A and K, where A and K are (contained in) a syntactic object already formed, and A raises to target K. The operation Move then creates two copies of the same element, A. A chain is the ordered pair consisting of A in derived position, and of its copy, technically its trace t_A, in the original position.

At some stage of the derivation an operation of Spell-Out applies stripping away the phonological properties of lexical items, and leaving only semantic and formal features. The computation then continues on these, and on phonological features separately.

As Chomsky (1995) points out, no real question arises about the motivation behind the operations Select and Merge. Select simply must apply till the set of lexical items that forms the basis of the computation is exhausted. Similarly, Merge must apply till a single syntactic object is formed. As for the reason why Move exists, however, only considerations of language use can be invoked: 'facilitation of parsing ..., the separation of theme-rheme structures from base-determined semantic (theta) relations, etc...'.

In relation to Merge, Chomsky (1995) considers the question whether and how the theory of Kayne (1994) can be incorporated into the minimalist model. Kayne's (1994) theory centers on the L(inear) C(orrespondence) A(xiom), which translates asymmetric c-command into linear precedence. In this way, the LCA derives principles of phrase structure as well as principles of word order. In particular, since linear order must be complete, terminals in phrase structure configurations must be completely ordered by asymmetric c-command as well.

Only two phrase structure configurations are allowed by the theory of Kayne (1994). The first one is what we can refer to as the complementation configuration, where a head is sister of a maximal projection, its complement. In this configuration the head asymmetrically c-commands the head of the complement; thus by the LCA, the head precedes it. The second configuration allowed by the theory is the external argument configuration, where a maximal projection is sister of a segment of another. From the point of view of linear order, the adjunct always precedes the head of the

projection it is adjoined to, since it asymmetrically c-commands it, i.e. it is left-adjoined.

Chomsky's (1995) model is less retrictive. In particular, it allows for a head complement of another head, if the complement is both minimal and maximal. Indeed Merge, being a binary operation, could not create a complement consisting of a head exhaustively dominated by its projection. Furthermore, Chomsky (1995) seeks to maintain the traditional distinction between specifiers and adjuncts, as well as the possibility of multiple specifiers and/ or adjunctions.

Let us now turn to Move in some detail. Move is crucially restricted by a property of Last Resort. According to it, a feature F raises to a target K only if F enters into a checking relation with a feature of the head of K, technically a sublabel of K. Strictly speaking indeed, Move applies to features. A problem of empirical adequacy arises in connection with this latter assumption, to the extent that overt movement takes the shape of movement of entire phrases, rather than of (bundles of) features. According to Chomsky (1995) this problem is straightforwardly solved by taking into account the fact that (bundles of) features in isolation cannot be pronounced. Thus it is reasons of PF interpretation that force entire phrases to pied-pipe in overt syntax. In covert syntax, where PF features have been stripped away, we can assume that movement takes the pure form of Move F(eature).

A second crucial property of the operation Move is the M(inimal) L(ink) C(ondition). According to the MLC, A can raise to target K only if there is no B such that raising of B to K is legitimate under Last Resort and B is closer to K than A is. Consider for instance a typical context for a wh-island violation, as schematically represented in (1):

(1) Q' ... [wh Q ... wh

Under the formulation of the MLC just given, in (1) movement of the lower *wh* to check its features against Q' is blocked, in that the *wh* in the Spec of Q can also move to Q' to do the same, and crucially is closer to it.

Superraising effects can be accounted for in an analogous way. Consider (2):

(2) seems [that it was told John [that ...

Under the MLC, movement of *John* to the matrix subject position to check Case is blocked by the presence of the intermediate *it*, which can also move to it to check its features. The feature theory that allows these results to be achieved is essentially irrelevant for present purposes.

Chomsky (1995) also suggests that the formulation of the MLC is more natural if Move is thought of in terms of attraction. Thus incorporating Last Resort into this new conception, we can define Attract/ Move by saying that K attracts F if F is the closest feature that can enter into a checking relation with a sublabel of K.

In essence, the MLC exploits the generalization first drawn by Rizzi's (1990) Relativized Minimality. Indeed, the MLC derives two of the main consequences of Relativized Minimality, namely wh-islands and superraising. A systematic comparison of the two principles, however, reveals that they are not necessarily equivalent, when the wider class of weak islands is taken into account.

In particular, in order to predict inner islands, created by a negation, under the MLC, it is necessary to prove that the same target K can attract both wh-phrases and negative quantifiers, so that the blocking effect of the latter on the former can be derived. In principle, this is perfectly possible. For instance, wh-phrases and negative quantifiers could both belong to a larger class of Focus elements, along the lines suggested for Hungarian by Brody (1990) among others. If so, Focus could be the property that attracts both wh and negation, thereby explaining their interaction under the MLC. Needless to say, the blocking effect that Foci proper do have on wh-movement would also be predicted.

A major difference between Relativized Minimality and the MLC explicitly noted by Chomsky (1995) concerns Travis's (1984) H(ead) M(ovement) C(onstraint). In essence, the MLC excludes those cases of the HMC where a head moves across another head attracted by the same target. However, unlike the HMC, the MLC does not necessarily exclude all cases of movement of a head across another head. Interestingly, the empirical adequacy of the HMC has been extensively discussed quite independently of the proposals of Chomsky (1995).

A wellknown prima facie problem for the HMC is represented by so-called long head movement, as studied notably by Rivero (1994). In essence, it appears that in a number of languages a lexical V moves to C under appropriate conditions, crossing an auxiliary in I. These cases are used by Roberts (1994) to argue in favor of a revision of Relativized Minimality that distinguishes between A and A' heads. As is evident, these same cases potentially favor the MLC, depending on whether some property of C can be shown to be involved, that attracts V but not I. Similar considerations apply to the revision of Relativized Minimality proposed by Baker and Hale (1990). It is then a purely empirical question whether the inability of the MLC to derive the HMC makes it inadequate in some respect, or viceversa provides a solution to long-standing problems with the HMC itself.

A more serious problem for the MLC is acknowledged by Chomsky (1995) when he states that the superraising violation, which falls under the MLC, is 'far more severe' than the wh-island violation involving arguments, which should fall under the

same principle. In fact, while Chomsky (1995) seems to take wh-extraction of arguments from a wh-island to be degraded, the classical study of Rizzi (1982) treats many instances of them in Italian as fully wellformed. Suppose they are. As already noted, Chomsky (1995) is fully explicit in construing the MLC as part of the definition of movement. If so, we are faced with the problem of a subset of movement cases that do not appear to conform to a part of the movement rule. As Chomsky (1995) is careful to point out, the minimalist framework leaves the problem completely open.

A second problem in locality theory that the minimalist approach of Chomsky (1995) leaves completely open is that of strong islands. An independent principle appears to be envisaged, presumably along the lines of Huang's (1982) CED. Still, it remains unclear whether the CED is to become a part of the definition of Move, like the MLC, or not. In either case, it also remains unclear why all types of movement are equally sensitive to strong islands and hence the CED, as opposed to weak islands and the MLC.

Finally, the theory of Chomsky (1995) fails to make any predictions concerning a third classical problem in locality theory, that concerns a class of phenomena apparently irreducible to either the MLC or CED, and accounted for by standard models within the principles and parameters framework in terms of some notion of head government. These are phenomena of the *that*-t class, typically involving asymmetries between object and subject.

The grammar of Chomsky (1995) excludes the notion of head government on what are essentially simplicity grounds. In particular, the (head, complement) relation and the (Specifier, head) relation are independently recognized by the grammar, in that they correspond to configurations created by Merge. Thus the only head government relation that does not correspond to a configuration created by Merge is the relation between a head and the Spec of its complement; the notion of head government needs then to be defined essentially for this case. Chomsky (1995) however argues that this configuration systematically reduces to one of the other two. For instance, there is no Exceptional Case Marking of the subject of *believe*-type complements by the matrix V; rather Exceptional Case Marking effects are due to the subject raising to the matrix (Spec, AgrO).

Suppose we accept this argument against head government. This means of course that we can no longer have recourse to accounts of *that*-t phenomena based on this notion, such as Rizzi's (1990). But other branches of the theory make no predictions about them either.

Interestingly, the last three problems illustrated here for Chomsky's (1995) minimalist grammar are individuated as the fundamental problems in locality theory by Manzini (1992). In pre-minimalist terms, the first corresponds to the disjunction

between argument and adjunct wh-movement with respect to antecedent government; the second to the disjunction between antecedent government, as relevant for weak islands, and the CED, as relevant for strong islands; and the third to the disjunction between antecedent government and head government. It seems fair to conclude that the minimalist approach makes no substantial progress not only with respect to the first question, as explicitly recognized by Chomsky (1995), but also with respect to the other two.

2 Form Dependency

As I have indicated above, according to Chomsky (1995) the operation Merge applies to a pair of syntactic objects S_1, S_2, creating a labelled set (S, (S_1, S_2)) out of them, where S is either S_1 or S_2. The operation Move applies to A and K in a phrase marker S, merging a copy of A with K; the two copies of A then form a chain.

The central theoretical proposal of this article consists in a revision of this standard conception of movement and chains. Important work in this direction is carried out by Brody (1994; forthcoming), arguing in particular that chains are primitives of the theory. In what follows, I shall in fact suggest that there are no chains, understood as ordered sets of copies of a given element, as well as no rule of movement to create them. I shall refer to the successors of chains within the present theory as dependencies, and I shall call the operation that creates them Form Dependency.

Specifically, I propose that Form Dependency takes as its input lexical items from the set N that forms the basis of the derivation according to Chomsky (1995) and creates ordered pairs out of them, roughly as in (3):

(3) *Form Dependency*
A, B → (A, B)

Unlike Move, Form Dependency is an elementary operation and does not presuppose Merge.

Last Resort is fully compatible with the conception of dependencies in (3). In particular we can assume that Form Dependency will take place only if the interpretive needs of the lexical items involved, i.e. effectively those of their features, force it to. Interpretive needs obviously subsume LF interpretation needs; but they can subsume also PF interpretation needs. I shall express this by saying that all dependencies must have an interpretation, a principle which is reminiscent of Full Interpretation. This in turn can naturally be combined with the idea that only dependencies can be interpreted, which is reminiscent of Chomsky's (1995)

characterization of chains as (perhaps) the only interpreted objects at LF. The combination of these two principles then yields what I shall call the Principle of Interpretation, as in (4):

(4) *Principle of Interpretation*
All and only dependencies are interpreted

Concrete examples of elementary dependencies formed by (3) because of interpretive needs are easy to provide. Thus PF interpretation needs can be argued to drive the formation of a dependency of the form (I, V), on the assumption that PF, or the Morphology subpart of it, will not be able to provide an interpretation of V+I as a word unless a dependency is established between its component parts. To provide another example, there are a number of reasons to assume that I is systematically bound by C; thus it is well-known that different types of C match different types of I. Under (4), this (C, I) dependency can correspond to the interpretive need for the T variable in I to be bound by an operator in C, and/or viceversa for a temporal operator in C to bind a variable in T.

Similarly, though Chomsky (1995) explicitly states that the establishment of theta-relations does not depend on movement and hence chain formation, within the present model we can assume that the interpretive needs of an argument head, such as D, and of a predicate head, such as V, drive them to form a dependency (V, D). Sentential complementation dependencies (V, C) can be driven by analogous reasons; and so on. Note that here, as throughout, categories such as C, I and V are employed just as a convenient shorthand for lexical items, which are the only primitives that syntactic operations manipulate.

In short, with respect to Last Resort, no major conceptual change is required by the shift from Move to Form Dependency. Under the model proposed here, on the other hand, Merge and Form Dependency create entirely independent syntactic objects, raising the problem of the mapping between them. This problem however is far from novel, since it can be raised given a standard conception of chains as well, as the problem of the mapping between phrase structure configurations and chains.

To be more precise, the question does not arise within derivational models, such as Chomsky's (1995), where an appropriately restrictive formulation of Move can dispense with the need for any further restrictions on the mapping between chains and phrase structure. It does however arise within representational models, where c-command and locality restrictions, however the latter may be formulated, apply to chains rather than to Move. By analogy with these models, we can then tentatively surmise that the mapping between phrase structure configurations and dependencies in the present model will include hierarchical order and locality conditions.

In particular, I assume that the hierarchical order constraint can be expressed by c-command, and in fact by asymmetric c-command, given Kayne's (1994) conclusion that this is the notion of c-command relevant to the grammar. Thus given a dependency (A_i, A_j), as part of its mapping to phrase structure we can require A_i to asymmetrically c-command A_j.

The issue of locality is needless to say more controversial. Remember that according to Chomsky (1995), the MLC, as included in the definition of Move, insures that movement of A to a target K is possible only if there is no shorter movement from B to K that also satisfies Last Resort. This in turn is shown to derive wh-islands and superraising, i.e. two of the three main consequences of Rizzi's (1990) Relativized Minimality to which it is conceptually related. However, as Chomsky (1995) explicitly points out, the MLC does not appear to derive the third major consequence of Relativized Minimality, namely the HMC. Indeed under the MLC nothing prevents a head from crossing another head with no features, or different features, to be checked, while this is not possible under the HMC.

Here, following in essence Manzini (1994a), I shall pursue the opposite conclusion that an appropriate version of the HMC is somehow basic with respect to (other) shortest movement effects. Thus I shall assume without discussion that the locality constraint on the mapping between dependencies and phrase structure is represented by some version of the HMC. Given a dependency (A_i, A_j), this can require in particular that no potential dependency member A_k intervenes between A_i and A_j, where intervention is defined in terms of c-command, hence of asymmetric c-command within the range of theories considered here.

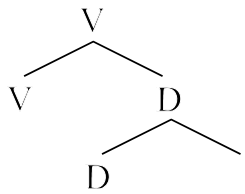
Interestingly, both the hierarchical order constraint and the locality constraint are now formulated in terms of the same primitive notion of asymmetric c-command. As already pointed out by Manzini (1994b) this means that the two constraints can be easily unified. Here I shall refer to the conjunction of the fundamental principles of c-command and locality as the Mapping Principle, to reflect its apparent role in the theory. The Mapping Principle reads as in (5):

(5) *Mapping Principle*
If (A_i, A_j) is a dependency,
(a) A_i asymmetrically c-commands A_j and
(b) there is no A_k such that A_k asymmetrically c-commands A_j and A_i asymmetrically c-commands A_k

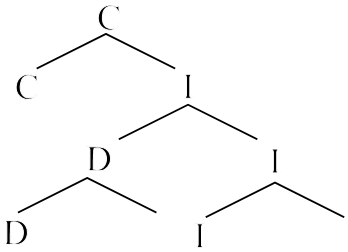
Consider for instance the two phrase structure configurations allowed by Kayne's (1994) LCA, namely what I have called the complementation configuration and the external argument configuration. The first one is exemplified by the canonical case

of a direct object of a verb, as in (6), the second one by the equally canonical case of a nominative subject, as in (7):

(6)



(7)



Obviously enough, (5) defines a wellformed mapping between (6) and the complementation dependency (V, D). Indeed the V head asymmetrically c-commands the D head in (6), nor is there any element that intervenes between them.

The case of external arguments is somewhat more complex. Let us assume that D bears a Case relation to I, encoded in a (D, I) dependency. (5) defines a wellformed mapping between the dependency (D, I) and the structure in (7), provided that D is taken as the D projection and I as the I head. In that case, D asymmetrically c-commands I and there is no other element that has the property of intervening between them. In general, in order for the Mapping Principle to work, it must be possible for a member of a dependency to be construed either as a head or as its projection in a phrase structure configuration. I take it that this possibility is allowed for by Chomsky's (1995) theory of phrase structure, under which head and projection are characterized by exactly the same properties, ultimately those of a lexical item; it is of course lexical items that enter into dependencies.

Notice that in (6), (V, D) cannot be construed as involving the V projection and the D head or projection because the former dominates the latter. Furthermore, (V, D) cannot be construed as involving the V head and the D projection because c-command holds beween them, but not asymmetrically. In (7), (D, I) cannot be construed as involving the D head and the I head or projection, because the former simply does not c-command the latter. Finally, (D, I) cannot be construed as involving the D projection and the I projection, since under the appropriate definition, the I projection contains the D projection, preventing c-command again.

Let me then at least tentatively assume that (5) represents an adequate theory of the mapping between phrase structure configurations and dependencies, as conceived under the Form Dependency hypothesis. The next problem that needs to be addressed is represented by (apparently) long-distance dependencies. None of the examples of dependencies considered so far are of this type, though the ability to account for them is obviously crucial to a successful theory of grammar.

Suppose that elementary dependencies of the type considered so far can be composed into longer dependencies; for instance, the elementary dependencies (C, I), (I, V), (V, D) illustrated above can be composed into a longer dependency that reaches from D to C (or viceversa). We can assume that it is this kind of dependency that allows apparently long-distance relations to be expressed. I shall assume in particular that the operation of dependency composition works along the lines of (8):

(8) *Dependency Composition*
If (A_i, A_j) and (A_k, A_l) are dependencies, where $A_j=A_k$, then $((A_i, A_j), (A_k, A_l))$ is a dependency.

The notion of composed dependency can be exemplified with a simple case of wh-movement, as in (9), where *t* is of course a copy of *who*:

(9) Who did you see some friends of t

In the course of the discussion that precedes I have already introduced several types of elementary dependencies crucially involved in this example, as well as their mapping to phrase structure configurations. For instance, the dependency (*see*, *some*) expresses the theta-relation between V and its object; the dependency (t_{did}, *see*) expresses the saturation of the temporal argument of V by I, and so on. On the basis of these elementary dependencies and of (8), the composed dependency in (10) can then be formed:

(10) (who, did), (did, t_{did}), (t_{did}, see), (see, some), (some, friends), (friends, of), (of, t_{who})

Crucially, it is locality considerations that induce us to have recourse to composed dependencies of the type in (10). From the point of view of interpretation it is clear that it is elementary dependencies which are relevant. Because it is irrelevant for other principles of the theory and notably for the Principle of Interpretation, the notion of composed dependency seems then to be best introduced directly into the Mapping Principle, which expresses locality restrictions within the present theory, as a way of

satisfying it. Elementary dependencies remain the only relevant notion of dependency in all other respects. More specifically, we can modify the Mapping Principle so as to allow c-command and locality to be satisfied either by an elementary dependency (A, B), as in (5), or by a composed dependency of the form (A, ...), ..., (..., B), as in (11):

(11) *Mapping Principle*
Given a dependency (A_i, A_j)
a. A_i asymmetrically c-commands A_j, and there is no A_k such that A_k asymmetrically c-commands A_j and A_i asymmetrically c-commands A_k; or
b. (A_i, ...), ..., (..., A_j) is a composed dependency, and (a) holds of its members

It is worth pausing at this point to consider a question, almost completely avoided so far, concerning the derivational or representational nature of grammar. Suppose we take the question to be defined in the terms of Chomsky (1995), as follows: 'head movement is narrowly local, but several such operations can leave a head separated by its trace by an intervening head, as when N incorporates to V leaving the trace t_N, and the [$_V$ V-N] complex then raises to I leaving the trace t_V; the chain (N, t_N) at the output level violates the locality property ...; but locality is observed by each individual step'.

Consider then the present theory. The head movement sequence hypothesized by Chomsky (1995) would be represented by two subdependencies, namely (V, N) for N-incorporation into V and (I, V) for V raising to I. Assuming that the principles of grammar are defined not directly for Form Dependency, but rather for the object it creates, namely the dependency, the present theory can be characterized as representational. In general, however, properties that hold of single steps of derivations according to Chomsky (1995), notably locality and Last Resort, hold of elementary dependencies in the present grammar. Thus from this point of view the present grammar may as well be characterized as derivational. In the discussion that follows, therefore, I shall continue to simply disregard the question, which is arguably just terminological.

Summing up so far, I have argued that a theory replacing Move with Form Dependency is compatible with several conceptual claims of minimalism. The crucial question that arises at this point is whether such a theory allows for any new insights into the problems that minimalism leaves (partially) unsolved. What I shall concentrate on first is the problem of motivating both Merge and Move within a minimalist theory.

At first sight, the shift from Move to Form Dependency does not help in solving the problem. If anything, it magnifies it, since Merge and Form Dependency appear to be

entirely separate operations, forming entirely separate syntactic objects. Consider however selection, which has the interesting property that it is expressed by the formation of a dependency under the present theory, and by Merge in the theory of Chomsky (1995). In this case, the same syntactic objects can constitute the input both to Merge and to Form Dependency.

Consider for instance a V selecting a pronominal D. Abstracting away from the issue of labelling, Chomsky's (1995) Merge expresses the complementation relation through the formation of a set (V, D), where D is both a minimal and a maximal projection. Form Dependency expresses the same relation through the formation of an ordered set, i.e. a dependency, (V, D).

At least in the example under discussion, then, the object created by Form Dependency identifies with the object created by Chomsky's (1995) Merge plus some notion of order. Viceversa the object created by Chomsky's (1995) Merge identifies with the object created by Form Dependency minus the ordering relation. The overlapping between the two types of constructs obviously suggests that there cannot be two completely independent operations defining them.

Let me then explicitly introduce the assumption that Form Dependency represents the only rule of grammar, taking as its input lexical items and yielding dependencies as its output. Independently of the present work, Kayne (1995) suggests that Chomsky's (1995) Merge, which creates a symmetric, or unordered, object, is to be revised to an operation that creates an asymmetric, or ordered, object. Within the present system such an asymmetric Merge operation appears now to be identifiable simply with Form Dependency.

What we have construed above as the Mapping Principle can retain its validity, except that it must now be construed as ordering dependencies without reference to an independent structure created by Merge. If so, the notion of asymmetric c-command, crucial to the formulation of (5), must be an undefined primitive of the theory. This conclusion appears once again to be reached by completely independent work, e.g. by Frank and Vijay-Shanker (1995). Following once again the lead of Kayne (1995) and to avoid terminological confusion with the standard notion of c-command, I shall refer to the syntactic notion of order as syntactic precedence. The Mapping Principle can then be reformulated as the Syntactic Order Principle in (12):

(12) *Syntactic Order Principle*
Given a dependency (A_i, A_j)
a. A_i syntactically precedes A_j and there is no A_k such that A_k syntactically precedes A_j and A_i syntactically precedes A_k; or
b. (A_i, ...), ..., (..., A_j) is a composed dependency, and (a) holds of each of its members

While for the elementary (V, D) case discussed above, the dependency also defines a constituent, the question arises of how more complex constituents, i.e. constituents including more than two terminals, are defined. Once again, however, a first answer is already present in the theory, in this case in the form of the independently defined notion of composed dependency. Thus a constituent headed by A includes all B's such that there is a (composed) dependency relating A and B, as in (13):

(13) A *constituent* headed by A includes A and all B's such that there is a (composed) dependency relating A and B.

The partial definition in (13) presents the obvious problem that it excludes left branches and other islands from forming constituents with the main branch of the sentence. In fact, as I shall briefly mention in the next section, these same islands can be circumvented by parasitic gaps. If a weakening of the notion of (composed) dependency, to allow it to branch, is involved in these configurations, as suggested by Brody (forthcoming) and Manzini (1994a), then the same revised notion of dependency can presumably be used in the definition of constituent as well.

The last step in the argument that Form Dependency is in itself sufficient to express not only movement relations but also phrase structural concepts is to show that it has the power to express all and only the configurations described by suitably restrictive theories of phrase structure. No problem arises in the case of complementation, already discussed with the (V, D) example. In fact the present theory yields the classical X-bar theory result of allowing for all and only those complementation structures where the selecting head is asymmetrically ordered with respect to the selected head.

The truly complex case is represented as in all theories by the external argument configuration. Let us consider once again the typical nominative subject configuration. Let us assume that the nominative Case relation between D and I is represented by the dependency (D, I). The complexity that this relation presents is that D also normally takes N as its complement, heading the separate dependency (D, N); and I in turn normally is the complement of C in the dependency (C, I). Thus D is twice the head of a dependency, and I is twice its tail.

For present purposes, I shall simply assume that this situation is allowed, in that nothing in the principles considered so far, including crucially the principle of Order, disallows it. Notice that the lack of any simple or composed relation between N and I is exactly as desired; in classical terms the two positions do not c-command one another. Furthermore, we can take the lack of simple or composed dependencies between C and D to correctly correspond to the fact that D is an island for extractions, as I shall discuss in detail in the next section.

The question then arises whether external argument dependencies of the type of (D, I) are restricted along the lines of Kayne (1994) or give rise to multiple Spec's/ adjunctions. For the time being, I shall once again limit myself to the observation that none of the constraints posited so far, including in particular the principle of Order, yields Kayne's (1994) restrictions. In the same way, it should also be noticed that nothing in the principle of Order disallows multiple complement configurations. In this case the selecting head will be ordered with respect to the head of each complement, while the complements will be unordered with respect to one another. Here I can only postpone to future work discussion of the complex issue whether other principles of the theory, and in particular principles of linear — or phonological — order, further restrict these possibilities.

What is immediately relevant for the time being is that if what precedes is on the right track, the notion of dependency not only provides the basis for establishing relations between syntactic objects standardly captured by movement. It also provides the basis for a definition of constituent and hence effectively of phrase structure, that overcomes the dualism, never fully explained, between Merge and Move. Needless to say, there is no exact equivalence between the notions defined in terms of dependencies, in the present sense of the term, and their standard counterparts. We can then expect the choice between them to be an empirical matter. For what concerns phrase structure, only preliminary results have been indicated here. The next section however will be devoted to briefly showing that the present theory leads to distinct advantages in the theory of movement and in particular of locality.

3 Locality

The present proposal bears an obvious relation to that presented by Manzini (1992). In essence, Form Dependency represents a generalization of the mechanism for the formation of address-based dependencies. The present grammar, as well as Chomsky's (1995), is simpler than Manzini's (1992), which includes both address-based dependencies and ordinary chains. On the other hand, the double mechanism for

dependency/ chain formation allows Manzini (1992) to provide a unified solution for the three leading problems in locality theory introduced above in the initial section.

In this section these problems will provide once more the focus for the discussion. My main line of argument will be that the shift to Form Dependency brings with it an automatic solution for the strong islands and head-government problems, and allows for a solution to the weak islands problem that is not available under the approach of Chomsky (1995). In other words, the thesis of this article, namely that relations between sets of lexical properties, i.e. dependencies, form the basis of syntax, will provide the solution to the classical unification problem in locality.

A canonical example of strong islands is provided by subject islands, as in (14):

(14) *Who did [some friends of t] bother you

In the case of the well-formed wh-extraction in (9), we have seen that it is possible to form a composed dependency relating the wh-phrase and its trace, as in (10), under the Mapping Principle or equivalently the Syntactic Order Principle. By contrast, in (14) and similar examples I shall argue that no such dependency can be formed.

Let us consider why no composed dependency in (14) can connect the wh-phrase and the trace. As shown in (15), a dependency can clearly be formed connecting the trace to the subject D. Furthermore, we know from the discussion above that there is a nominative Case relation between the subject D and I, as in (16). And finally, the wh-phrase is of course connected to I along the main branch of the tree, and further down, as in (17):

(15) (some, friends), (friends, of), (of, t_{who})

(16) (some, t_{did})

(17) (who, did) (did, t_{did}) ...

It is clear that in order for the wh-phrase to be related to its trace, the subdependency in (15) and the subdependency in (17) must somehow be unified. It is equally clear however that the link in (16) between *some* and t_{did} cannot serve to this purpose, given the dependency composition mechanism in (8). Thus (14), or equivalently any other example of extraction from the subject, is illformed because of the impossibility of expressing the wh-relation through either an elementary or a composed dependency.

Of course, the problem of providing a connection between the two subdependencies in (15) and (17) would automatically be solved by the existence of a link between

some and *did* in C. However such a (C, D) dependency is simply excluded by the Principle of Interpretation, since there is no independent interpretive need for it. Thus it is crucially the interplay of the principles of Order and Interpretation, the two basic principles of the present grammar, that rules out strong islands.

Several empirical problems remain open after this first schematic illustration. One of them concerns the second major type of strong islands, namely adjunct islands. A preliminary problem to be solved in this case is represented by the phrase structure configuration in which the adjunct appears. In particular, since complex adverbials systematically appear to the right of the main branch of the sentence, they are generated as right adjuncts under classical X-bar theory. Kayne's (1994) conclusion that all adjunctions are left adjunctions apparently requires a movement analysis, or alternatively the assumption that the main branch of the sentence is indeed left-adjoined to the adjunct, as apparently proposed by Sportiche (1994).

Under the present theory it is clear that from an interpretive point of view, the main branch of the sentence functions as the external argument of the P's, such as *before*, *after*, *for*, etc, that head the adjunct. This external argument relation will be represented by dependencies of the form (I, *before*), and so on. Furthermore, the dependency-based definition of constituent provided above allows us to recognize for instance I and *before* as belonging to the same constituent, headed by I. At the same time, nothing under the present dependency-based conception implies that constituents are labelled. The question of left vs. right adjunction, which is just a question of labelling, then simply disappears.

If what precedes is on the right track, the analysis of adjunct islands can proceed along familiar lines. On the one hand, there is no direct connection between the head of the adjunct and the immediately superordinate head, which selects the main branch of the sentence. On the other hand, the external argument link between the adjunct and the main branch of the sentence is not sufficient to insure the formation of a composed dependency reaching from the wh-phrase on the main branch to its trace embedded in the adjunct.

The second major empirical problem with strong islands concerns the fact that they can be circumvented in so-called parasitic gaps configurations. Following Brody (to appear) and Manzini (1994a), these crucially involve the formation of a branching dependency. If so, the solution to the problem can consist simply in an adequate reformulation, and in particular a weakening, of the present mechanism for dependency composition. An analysis of this question however is once again beyond the scope of the present article.

Assuming that the problems noted do have a solution, along the lines just indicated, we can conclude that the theory advocated here provides a conceptually satisfying solution to strong islands. Indeed the present theory can predict them on the basis

simply of the primitive operations and principles of grammar, namely Form Dependency and the principles of Order and Interpretation. By contrast, in a theory such as Chomsky's (1995), strong islands do not follow directly from the syntactic definition of Move, including the relevant versions of locality and Last Resort, but are derived by some additional principle, namely a version of the CED.

The second problem in the theory of locality individuated in the discussion that precedes concerns the problem of head-government and of the effects imputed to it. These are illustrated by wellknown contrasts of the type of (18)-(19) in English, or (20)-(21) in French:

(18) *Who do you think that t left

(19) Who do you think t left

(20) *Qui crois-tu que t est parti

(21) Qui crois-tu qui t est parti

According to the generalization drawn by Rizzi (1990), the ungrammaticality of (18) and (20) is due to the fact that the trace in subject position is not properly head governed. I does not properly head govern t because it does not c-command it; C does not properly head govern t for lack of lexical features. As is well known, Rizzi (1990) argues that this problem is solved in (21) by agreement between I and C, of which *qui* is the spell-out. For, the agreeing C not only satisfies the structural conditions on proper head-government, but also has lexical (Agr) properties. The same schema of solution can furthermore be extended to (19) by claiming that the agreeing form of *that* is phonologically zero.

The problems raised by Chomsky (1995) for the notion of head-government have already been reviewed. As discussed in detail by Manzini (1992), a more specific problem for Rizzi's (1990) definition of proper head-government arises in connection with its use of c-command, since this does not apper to be reducible to the fundamental use of c-command in ordering dependencies and/or phrase structure. Finally, the requirement that heads must have some sort of lexical content in order to properly govern means that some definition of what counts as a lexical property must be drawn up; this is again undesirable to the extent that the notion does not have independent motivation in the grammar.

Let us then conclude that there are sound conceptual grounds for abandoning the notion of (proper) head government. Unfortunately, as already discussed, this conclusion leaves Chomsky (1995) without any apparent account for the facts in (18)-

(21). By contrast, as I shall argue, the key to a solution for (18)-(21) within the present model is implicit in the discussion of strong islands that precedes.

In discussing (14), we have seen that (16) is not sufficient to provide a link between the two subdependencies in (15) and (17); thus no dependency linking the wh-phrase and the trace can be formed, and extraction from the subject is excluded. It is evident, however, that on the same grounds, not only extraction from the subject, but also extraction of the subject is blocked by the present grammar. Thus (20) presumably includes subdependencies (*que*, *est*) expressing the interpretive link between C and I and (t, *est*) expressing the morphological link between I and the nominative subject. Furthermore (*que*, *est*) is linked upwards to the matrix V and eventually to the wh-phrase. Nevertheless, no composition is possible between the two subdependencies (*que*, *est*) and (t, *est*); this means that t cannot itself be connected to the wh-phrase, yielding a prediction of ungrammaticality. This prediction, which follows from the theory without stipulation of sort, is indeed verified by examples of the type in (20).

At this point, only half of the original head-government problem survives, namely that of explaining the grammaticality of (21). We have seen that according to Rizzi (1990), the *que-qui* rule, i.e. the rule responsible for (21), involves incorporation of Agr into *que*. Early generative acounts, such as Pesetsky's (1982), on the other hand could express the same rule as involving incorporation of t itself into *que*.

Within the present framework, the idea that incorporation of t to *que* takes place to produce *qui* is naturally restated by saying that a dependency (*que*, t) is formed which is morphologically interpreted as *qui*. It is not difficult to notice that precisely the formation of such a dependency solves the extraction problem for subjects. For, this dependency links the subject to C, from where a straightforward composed chain can reach the matrix C and ultimately the wh-phrase. In other words, by providing a morphological interpretation for it, the grammar licences a (C, D) dependency otherwise impossible under the principle of Interpretation. This dependency of course also satisfies the Order Principle.

It is worth noting that such an analysis of C-t effects contains the implicit prediction that all left branch violations can be circumvented if there is some dependency involving the head of the left branch and the immediately superordinate head. However in the case of subjects it appears that the only possible dependency of this type is precisely the one just reviewed involving a wh-subject, presumably because wh but not other properties can be incorporated into C. The net result is that *that*-t effects can be circumvented, but subject islands never can.

Needless to say, if we follow Rizzi (1990), the zero complementizer of English can be construed as some sort of abstract version of *qui*, thus explaining the contrast in (18)-(19) as well. This solution unfortunately does not account for other phenomena relating to *that*, such as apparently optional cases of *that* deletion and anti-*that*-t

effects in relative clauses. It is possible then that an account for the latter two phenomena will involve a revision of the analysis of *that*-t.

What is immediately relevant here however is that the Form Dependency model provides a schema of solution for the head government problem which has recourse only to Form Dependency itself and to inescapable language specific assumptions, such as the assumption that French provides a morphological interpretation, *qui*, for a (*que*, t) dependency. As already pointed out, other leading models either necessitate the addition of the notion of head-government to the grammar, or leave the empirical problem open.

Finally weak islands remain to be taken into consideration. The abstract structure of a wh-island violation is illustrated in (1) above; (22) is a concrete example of such a violation:

(22) *Quanti hanno già decidere a chi assegnarne
How many have they already decided to whom to give (of them)

Given that the particular Italian example chosen only involves complementation structures, the relevant syntactic representation is straightforwardly as in (23):

(23) [Quanti Q [devono decidere [a chi Q [assegnar [t ne] t]]]]

Given the grammar advocated here, the wh-phrase and its trace can be related through Form Dependency; in particular, the Order Principle can be satisfied by a composition of elementary dependencies, whose first member is the wh-phrase and the last member its trace. Such a composed dependency takes essentially the form in (24):

(24) (Quanti, Q), (Q, hanno), (hanno, deciso), (deciso, Q), (Q, assegnare) (assegnare, $t_{assegnare}$) ($t_{assegnare}$, t_{quanti})

Each elementary dependency in (24) is necessary and hence possible under the Principle of Interpretation. In particular, it is obvious that the last dependency is of the general form (V, D), where D is the internal argument of V; and all intermediate dependencies are also complementation dependencies of the general form (I, V) or (C, I) or (V, C) already examined. As for the initial dependency, of the general form (wh, Q), in it wh- in essence acts as the external argument of the interrogative C head, Q. In Chomsky's (1995) terms, this is a feature checking relation. In present terms, much the same can be true, assuming that checking holds some interpretive significance, even if only at the PF interface.

So far, then, it appears that there is a composed dependency able to express the relation between the wh-phrase and its trace in (22). But if so, the theory seems to be faced with a problem, since (23) is clearly ungrammatical. What I shall argue however is that far from representing a problem, the conclusion that (23) and the like are syntactically wellformed is entirely correct. In fact, wh-island violations represent violations of independently needed interpretive principles, essentially along the lines anticipated by Manzini (1994c).

Consider the composed dependency in (24) once again. The dependency as a whole contains not one, but two distinct realizations of the Q/wh operator, one in the matrix CP and another in the embedded CP. By contrast, only one variable is included in the dependency, corresponding to the trace. Remember then that according to a proposal originally advanced by Koopman and Sportiche (1983), a Bijection Principle holds of the (operator, variable) relation, to the effect that each operator must bind exactly one variable and each variable must be bound by exactly one operator. While binding of more than one variable by the same operator is arguably grammatical, being involved in parasitic gaps and similar phenomena, it appears that any attempt at binding one variable by more than one operator indeed yields no interpretation.

Of course, since within the theory proposed here the only interpreted objects are dependencies, it appears to be natural to state interpretive principles as ruling in or out dependencies of a given form. If so, the surviving clause of the Bijection Principle can take the form of a prohibition against two dependencies including one variable and two distinct operators, as in (25):

(25) *Uniqueness - I*
*$(Op_1, vbl_1), (Op_2, vbl_2)$
if $Op_1 \neq Op_2$ and $vbl_1 = vbl_2$

On the basis of (25) we can now correctly rule out the dependency in (24) and hence ultimately the example in (22), if we assume that the composed dependency in (24) implies two (operator, variable) dependencies, of the form (Q, t_{quanti}), with the matrix and embedded Q's involved. The conjunction of these two dependencies is then ruled out by (25), as desired.

In general, we can hypothesize that Relativized Minimality effects arise when the requirements imposed on long-distance dependencies by the Order Principle come together with biuniqueness principles. These include the Bijection principle, but also the original Biuniqueness of Vergnaud (1985), establishing a one to one relation between theta-roles and Cases.

As with the Bijection Principle, one of the two clauses of Biuniqueness does not appear to hold, since one argument, corresponding to one Case position, can bind

more than one theta-role, for example in instances of secondary predication. The other clause however appears to hold, and by analogy with (25) it can be expressed as a constraint against one theta-role forming a dependency with more than one Case, as in (26):

(26) *Uniqueness - II*
*(Case$_1$, theta$_1$), (Case$_2$, theta$_2$)
if Case$_1$ ≠ Case$_2$ and theta$_1$ = theta$_2$

It appears then that superraising violations, as exemplified above in (2), can be excluded by the interaction of the surviving clause of Biuniqueness in (26) with the familiar syntactic constraints on dependencies. Indeed the superraising violation in (2) involves establishing a relation between a theta-property and a Case property across another Case property. If the intermediate Case assigning/ checking head is not included in the composed dependency relating the raised D(P) and its trace, the principle of Order, which has the force of the HMC in this respect, is obviously violated. If on the other hand, the intermediate Case assigning/ checking head is included, we can assume that, though the principle of Order is satisfied, the Uniqueness principle in (26) is now violated.

If this line of argument is correct, weak islands and other Relativized Minimality violations simply disappear as a syntactic problem, being reduced to interpretive principles of some kind or other. Let us then reconsider in this light the fact that certain wh-dependencies are not sensitive to wh-islands. For instance, (27), which is identical to (22) except for the fact that movement takes along non wh-material, is clearly wellformed:

(27) Quanti posti devono decidere a chi assegnare
How many jobs must they decide to whom to give

If weak islands are now recognized as a purely interpretive phenomenon, a general schema of solution to the contrast between (22) and (27) based on interpretive properties becomes especially natural. Suppose in particular that Cinque (1991) is on the right track in claiming that in examples like (27) the dependency between the wh-phrase and its trace is not construed as an (operator, variable) dependency. If so, we fully expect that (25) will not apply to (27), thus predicting the absence of wh-island effects.

Needless to say, this schema of solution needs to be implemented by showing that a construal different from the (operator, variable) one can be provided for the wh-dependency in (27). On the other hand, alternative schemas of solution having resort

to interpretive properties are in principle also allowed by the present theory. By contrast, any such schema of solution is unavailable in the theory of Chomsky (1995) for the simple reason that the MLC, which derives weak islands, applies to each step of a derivation, and cannot have access to a full LF representation.

In a nutshell, the idea supported here is that weak islands are best treated as the result of the interaction of a purely syntactic principle of Order (subsuming locality) with interpretive Uniqueness principles, and not as the result of a primitive locality condition on Move and/or chains in the manner of Chomsky's (1995) MLC, or originally Rizzi's (1990) Relativized Minimality. If the discussion that precedes is on the right track, the present modular approach is to be favored on empirical as well as on conceptual grounds.

References

Baker, Mark and Ken Hale. 1990. Relativized Minimality and pronoun incorporation. *Linguistic Inquiry* 21: 289-297

Brody, Michael. 1990. Some remarks on the Focus field in Hungarian. *UCL Working Papers in Linguistics* 2

Brody, Michael. 1994. Directionality and dependency. *UCL Working Papers in Linguistics* 6

Brody, Michael. To appear. *Lexico-Logical Form*. Cambridge, Mass.: The MIT Press

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

Cinque, Guglielmo. 1991. *Types of A'-dependencies*. Cambridge, Mass.: The MIT Press

Frank, Robert and K. Vijay-Shanker. 1995. C-command and grammatical primitives. *GLOW Newsletter* 34

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

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, Mass: The MIT Press.

Kayne, Richard. 1995. Post-graduate lecture. Università di Firenze and Siena

Koopman, Hilda and Dominique Sportiche. 1982. Variables and the Bijection Principle. *The Linguistic Review* 2:185-213

Manzini, M. Rita. 1992. *Locality: A theory and some of its empirical consequences*. Cambridge, Mass.: The MIT Press.

Manzini, M. Rita. 1994a. Locality, minimalism and parasitic gaps. *Linguistic Inquiry* 25:481-508

Manzini, M. Rita. 1994b. Syntactic dependencies and their properties: strong islands. *UCL Working Papers in Linguistics* 6

Manzini, M. Rita. 1994c. Syntactic dependencies and their properties: weak islands. *Quaderni del Dipartimento di Linguistica dell'Università di Firenze* 5

Pesetsky, David. 1982. Complementizer-trace phenomena and the Nominative Island Condition. *The Linguistic Review* 1:297-343

Rivero, Maria-Luisa. 1994. Clause structure and V-movement in the languages of the Balkans. *Natural Language and Linguistic Theory* 12: 63-120

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, Mass.: MIT Press.
Roberts, Ian. 1994. Two Types of Head Movement in Romance. In *Verb Movement*, ed. N. Hornstein and D. Lightfoot. Cambridge: Cambridge University Press
Sportiche, Dominique. 1994. Adjuncts and adjunction. *GLOW Newsletter* 32
Travis, Lisa. 1984. *Parameters and effects of word order variation*. Doctoral Dissertation, MIT, Cambridge, Mass.
Vergnaud, Jean-Roger. 1985. *Dependences et niveaux de representation en syntaxe*. Amsterdam: John Benjamins