

Wp 4

**THEORETICAL LINGUISTICS PROGRAMME, BUDAPEST UNIVERSITY (ELTE)**

**TOWARDS PERFECT SYNTAX**

**MICHAEL BRODY**

**RESEARCH INSTITUTE FOR LINGUISTICS, HUNGARIAN ACADEMY OF SCIENCES**

**WORKING PAPERS IN THE THEORY OF GRAMMAR, VOL. 2, No. 4**

**RECEIVED: AUGUST 1995**



MTA Nyelvtudományi Intézet Könyvtára



0000333



# TOWARDS PERFECT SYNTAX

MICHAEL BRODY

THEORETICAL LINGUISTICS PROGRAMME, BUDAPEST UNIVERSITY (ELTE)  
C/O RESEARCH INSTITUTE FOR LINGUISTICS, HAS, ROOM 119  
BUDAPEST I., P.O. Box 19. H-1250 HUNGARY  
E-MAIL: brody@nytud.hu

WORKING PAPERS IN THE THEORY OF GRAMMAR, VOL. 2, NO. 4  
SUPPORTED BY THE HUNGARIAN NATIONAL RESEARCH FUND (OTKA)

THEORETICAL LINGUISTICS PROGRAMME, BUDAPEST UNIVERSITY (ELTE)  
RESEARCH INSTITUTE FOR LINGUISTICS, HUNGARIAN ACADEMY OF SCIENCES

BUDAPEST I., P.O. Box 19. H-1250 HUNGARY  
TELEPHONE: (36-1) 175 8285; FAX: (36-1) 212 2050





Wp 4

Nyelvtudományi Intézet  
Könyvtára

Lelőírt szám:

26 967/96



**Contents**

Projection and Phrase Structure	....	1
Perfect Chains	....	28







## Projection and Phrase Structure

### 1. Perfect Syntax

Consider a rather standard system of grammar in which the relationship between meaning and sound is mediated by two interpretive systems applying to some interface representation(s) generated by syntax.\* Suppose that these apply to the same representation, clearly a desirable additional assumption. The syntactic computation can be viewed as having then the task of composing this interface representation (say the level Lexico-Logical Form (LLF) of Brody 1995a), from elements provided by the lexicon.

A possible view is to deny the existence of syntax in this sense. One can maintain that there is no competence theory internal question as to how LLF structures are put together. Under such a view LLF structures are **given**, by grammar external systems, and the task of the grammar would be only to define a subset (the well formed instances) of the structures offered by some grammar external component.

Consider the conjunction of such a view with a strong version of the minimalist hypothesis according to which syntactic interface conditions reduce to "bare output conditions", ie. conditions forced on (L)LF representations by the interpretive systems applying to it. The conjunction of this hypothesis with the assumption that the question of LLF assembly is grammar external entails that there is no syntax at all that is part of human grammatical competence.

Obviously such a "no syntax" view does not resolve the question of how the structures that are the input to the interpretive systems are related to individual lexical items (LIs), it only shifts this problem into a different component of the mind. The issue of whether syntax exists is nevertheless empirical and we might hope to find evidence that bears on this matter. The position that syntax, in the sense of LLF assembly, exists as part of competence theory will be supported to the extent that such a system can provide explanations of (L)LF properties. The theory to be presented in this paper will provide some evidence of this kind.

Suppose that syntax, in the sense just characterized exists. There is then an empirical issue as to the nature of this system, which relates LIs creating the LLF representation. Optimally, this system should be near trivial: we would hope that apparent complexities are due to properties of the interpretive components. We might then expect to find a system that is significantly more perfect than the assembly system of the standard minimalist framework. Even if the Chain/Move relation is taken to be part of this system, there should be no **syntax internal** conditions on it (like Uniformity, Minimal Chain Link, C-command, Last Resort etc.), --cf. Brody 1995b and the discussion of Uniformity below. Furthermore there should be no representational-derivational duplications of (near-)identical concepts (eg. Chain and Move or the representational definitions of well-formed syntactic objects in addition to actual derivations), --cf. Brody 1995a and some discussion in section 5.2 below respectively. Such a more restrictive framework eliminates also the possibility of using representational-derivational distinctions like deletion (interface invisibility) vs. erasure (invisibility for the syntactic computation) that build on such duplications. Additionally we will expect to be able to dispense with economy conditions and the serious computational complexity that they create. Such requirements simply ensure that apparent imperfections in the assembly system follow from syntax-external considerations. Let us call the theory meeting them Perfect Syntax. In fact given recent advances in the minimalist framework, the apparently ambitious

---

\* This paper contains a revised version of part of the material in Brody 1994. "Brody 1995b" refers to the second paper in this issue in the first and to the first in the second.



program of Perfect Syntax seems quite reasonable. (See Brody 1995b for a more extended discussion of this approach.)

In this paper I shall discuss a system of LLF assembly that could be part of such a theory of Perfect Syntax and justify empirically some of its restrictive aspects. I first raise some problems in section 2 for the relational (contextual) definitions of projection levels of the standard minimalist framework. I provide an alternative system of phrase structure in section 3. Section 4 derives basic conditions of this system from a theory of the assembly of syntactic structures, providing evidence for the competence theory internal existence of this computation. I shall compare certain salient aspects of this theory with the corresponding properties of the standard minimalist framework arguing that the theory defended here is not only simpler but also more adequate in other ways.

Finally in section 5 I shall turn to the explanation of the Generalized Projection Principle, the condition whose major consequence is that selectional requirements and categorial projection must hold in the root positions of chains. I shall discuss a non-syntactic explanation which I will argue is superior to other recent accounts that give only a partial solution and assume the accidental conspiracy of unrelated principles.

## 2. Problems with Uniformity and the contextual definitions of projection levels

Chomsky 1994, 1995 puts forward the Uniformity condition for chains:

(1) A chain is uniform with regard to phrase structure status

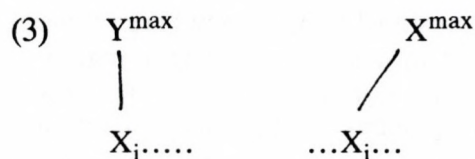
Here the "phrase structure status" of an element is its (relational) property of maximal, minimal or neither". Since intermediate projections are assumed not to be accessible to the syntactic computational system, and hence for chain formation, the Uniformity condition in (1) predicts that only  $[X^{\max}, X^{\max}]$  and  $[X^{\min}, X^{\min}]$  chains exist.

The Uniformity condition is necessary to achieve this result only in the context of the relational definitions of minimal, maximal and intermediate projections. If the projection level of a category is an inherent property, then the independently motivated assumption that chains consist of copies trivially entails this result. On the other hand given relational definitions like (2) it is easy to construct non-uniform chains.

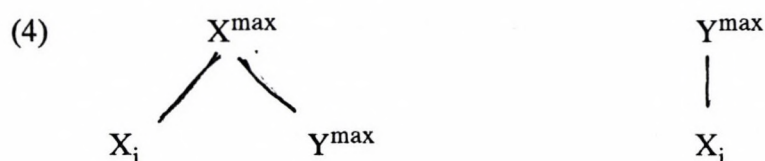
- (2)
- a. a maximal projection ( $X^{\max}$ ) is one that does not project further
  - b. minimal projections ( $X^{\min}$ ) are the lexical items themselves
  - c. intermediate projections ( $X'$ ) are elements that are neither maximal nor minimal

For example in (3) where an  $X^{\min}$  forms a chain with a copy that adjoined or substituted to an  $Y^{\max}$ ,  $Y_{\text{not}}=X$ , it will form a chain with an  $X^{\max}$ : Here the first  $X_i$  is an  $X^{\max}$  and the second an  $X^{\min}$ :





Or, if a non chain-root ("moved")  $X^{\max}$  merges with some category, say  $Y^{\max}$  and then  $X$ , rather than the target of the operation projects further, then it will form a non-uniform chain with a nonmaximal projection. In (4) the second  $X_i$  is an  $X^{\max}$ , the first is not:



On usual assumptions the structures in (3) and (4) are ill-formed: minimal projections cannot move to (form chains with) positions that are not word-internal ( $X^0$ -internal in Chomsky's (1995) terminology) and it is always the target of movement that projects (only chain-root positions can project). The impossibility of configurations like (3) and (4) has been thought of as providing evidence for the relational definitions of projection levels and the uniformity condition.

The behavior of clitics has been cited as additional empirical evidence for this system. According to the definitions in (2) a category can be both a minimal and a maximal projection: a lexical item that does not project further. Clitics appear to transparently instantiate this option, since they show properties of both minimal and maximal projections. As an  $X^{\min}$  they show up word internally, but they seem to be linked to argument positions that are maximal. Furthermore they can often form chains that ignore intervening heads, again suggesting (in the context of the head movement constraint) that they are maximal. Thus, given the relational definitions in (2), clitics might be treated as both  $X^{\max}$  and  $X^{\min}$  at the same time, accounting for their apparent double nature.

There are a number of problems with the relational definitions (plus Uniformity) system that seem to me serious. Let us consider them in turn.

i. Ordinary head movement chains that target word-internal positions are  $[X^{\max}, X^{\min}]$  by the contextual definitions, since the root of the chain is an element that does project further while the non-root word-internal member does not. Such chains must be allowed but they violate Uniformity. Chomsky (1994, 1995) suggests that there is a special component "WI" at LF, where "independent word-interpretation processes" apply. This then ensures that word( $X^0$ )-internally the principles in (1) and (2) do not apply. WI is "a covert analogue to Morphology" (1995, section 7.2, p9). But the reason for the existence of such a covert analogue of morphology, and thus the status of WI is unclear. Given the lack of evidence for such an additional module the WI hypothesis appears to amount to little more than a statement that head-movement targeting a word-internal position is exempt from the uniformity requirement. But if so, then we cannot say that Uniformity explains the impossibility of head-chains like (3) where a non-root member is word-external. The crucial distinction between good and bad cases here is the word-external versus word-internal contrast. Uniformity says nothing about this divide, which is simply stipulated.

Note incidentally that in the standard minimalist framework even the status of



"ordinary" morphology is rather unclear. Morphology must presumably be somewhere on the SPELLOUT branch. Since there are only interface conditions in the minimalist grammar, it would have to be at the PF level, which does not have the structure necessary for this component to operate. (In contrast in the framework of Brody 1993a, 1995a where spellout applies to the LF level, morphology can be identical to WI and its principles will hold at this level.)

ii. If relational definitions do not apply word-internally, then the evidence clitics appeared to provide for them disappears. Recall that the evidence was that clitics appear to have a different projectional level status word-internally and word-externally. Since grammatical cases of head movement make it necessary to exempt word-internal structure from these definitions, the word internal status of clitics becomes irrelevant, and thus cannot be used to support the system.

iii. The assumption that the relational definitions do not apply word-internally creates further problems. Consider the assumption that word-internal XP-adjunction is excluded in Morphology: "The morphological component gives no output (so the derivation crashes) if presented with an element that is not an  $X^0$  or a feature" (Chomsky 1995, section 7.2, p6). The question arises how morphology will be able to tell what is an XP inside a word if contextual definitions do not apply inside a word. Clearly some other characterization of minimal and maximal projection will be necessary. But the resulting system seems quite undesirable: why should we need two systems (one relational, one presumably not) to define projection levels? Differently put, why do we need contextual definitions of projection levels in addition to the apparently independently necessary inherent characterization?

iv. As we have seen above in connection with the structure in (4), Uniformity can be used in certain cases to ensure that the target rather than the "moved" (non chain-root) category projects. But Uniformity captures here only a small segment of a much larger generalization. Firstly it cannot ensure generally that categorial projection is always in the root positions of chains. The generalization holds also for  $X^0$  chains which involve word-internal positions. But Uniformity is relevant only for phrasal movement, word-internally different principles must apply as we have seen. Furthermore the generalization that categorial projection always holds in the root positions of chains is still only one aspect of a much larger generalization, the Generalized Projection Principle (GPP). There are a number of other "projectional" features that only chain-roots can project: theta roles, and semantic selection in general. Since the principle constrains also (quasi-)semantic properties like non-grammaticalized selection, it does not appear to be fully reducible to syntax. cf Brody 1995a and section 5 below for discussions. If the arguments below against the conspiracy account of the GPP are correct, then the fact that Uniformity gives a partial account of one aspect of the GPP is not an argument in its favor. If anything it is an argument against Uniformity since it seen to be redundant here.

v. An additional curious feature of the relational definitions plus Uniformity theory is that according to this system chains but not categories have to be uniform (recall that a non-projecting lexical item is both an  $X^{\min}$  and an  $X^{\max}$ ). This is of course logically possible: a chain-member may be multiply characterized, but all chain-members must have the same characterizations. But once we recall that characterization of an element as both minimal and maximal does not necessarily lead to contradiction and ungrammaticality, the Uniformity assumption seems to lose much of its intuitive appeal.

vi. In my view the most serious objection to the uniformity condition is that one would expect a well-designed theory of syntactic computation simply not to make it possible to violate this condition: the theory should not provide devices that can violate it. Now without



relational definitions there can be no Uniformity violations: chains are copies. The copy of an  $X^{\max}$  is an  $X^{\max}$ , the copy of an  $X^{\min}$  is an  $X^{\min}$ .

The assumption that chains consist of copies is an independently necessary assumption in minimalist frameworks, where representational conditions, like the binding theory can hold only at or beyond the interface level of LF. For example in order to rule out the principle C violation indicated by the indices in (5), the trace must be (at least a partial) copy at LF (and perhaps beyond):

(5) Whose<sub>x</sub> mother did he<sub>x</sub> like (whose<sub>x</sub> mother)

I conclude that the grammar should contain no contextual definitions of projection levels. Since chains consist of copies, Uniformity is unnecessary since there are no means to violate it, --the optimal situation.

### 3. A Minimal Theory of Phrase Structure

#### 3.1. The Principle of Phrasal Projection

Phrases and their heads share properties, like being an N(P) or V(P) etc. It is often assumed that the shared properties of the phrase are inherited from the head, syntactic categorial structure is projected from the lexicon. Let us express this by saying that phrases are projected by their heads. It seems that every phrase must share properties with some head, there are no "pure" phrases. If this is true then it is presumably true because phrases can only arise through projection:

(6) Principle of Phrasal Projection (PPP)

Every phrase is projected by a lexical item (LI) that it dominates

The step from "phrases are projected by LIs" to "all phrases are projected by LIs" seems highly natural to me although clearly it is not a necessary one. (Compare this with Kayne's (1995) approach in terms of his Linear Correspondence Axiom (LCA), where there is no relation between the fact that phrases have heads and the fact that phrases and their heads share features.)

Given Chomsky's (1995) general condition of Inclusiveness ("the interface levels consist of nothing more than arrangements of lexical features") the additional assumption in (6) seems unavoidable. Let us assume then that phrases are copies of features of lexical items. A lexical item is thus an  $X^{\min}$ , a phrase is its partial copy that dominates it. Ignoring intermediate projections assume, that all phrases are maximal. For the time being this last is only a simplifying assumption, made for the sake of presentation, --but see section 3.2 below for some discussion. The PPP in (6) seems to provide an optimal theory of syntactic structure. But the PPP as a general condition on the well-formedness of phrasal projection does not suffice; several additional assumptions appear to be necessary.

First of all it must be ensured that all and only non word-internal heads project a phrase, let us call this the extended structure preservation restriction:

(7) Extended structure preservation

- a. Every non word-internal head projects some phrase
- b. No word-internal head projects a phrase



As noted in section 2, Chomsky 1995 assumes that (7b) is a morphological condition: morphology does not tolerate phrases. Adopting the relational definitions of projection levels, he assumes instead of (7a) that a non word-internal head that has not projected is both minimal and maximal. Such elements thus can occupy specifier and complement and  $X^{\max}$ -adjoined positions, which are reserved for maximal projections. He then rules out a "moved" non-root  $X^{\min}$  in such positions using Uniformity. But as we have seen the account of why a non-root  $X^{\min}$  cannot appear here is stipulative and there are also a number of other reasons for not adopting a theory with the contextual definitions of projection levels. In addition the approach in terms of contextual definitions cannot capture the suggestive symmetry of (7).

Secondly the uniqueness of the relation between a phrase and a head also needs to be ensured, say as in (8):

- (8) Uniqueness  
Every phrase is projected by a unique LI

The uniqueness requirement ensures that a phrase cannot be projected by two heads. Thus (8) excludes examples like (9).

- (9) a.  $*[X/YP \ X \ Y \ ]$   
b.  $*[X/ZP \ Z \ [XP \ X \ ]]$   
c.  $*[X/YP \ [XP \ X \ ] \ [YP \ Y \ ]]$

(In the examples in (9) "X/YP" indicates a phrase that both X and Y have projected, ie. a phrase that shares properties with both.) Notice here, that Kayne's LCA predicts this result only for the special case when the two heads are both immediately dominated by the phrase, as in (9a). The LCA rules out this structure since contrary to its requirement there is no pair (C, C') of constituents related by asymmetric c-command such that C dominates X and C' dominates Y. (According to Kayne's theory the terminals dominated by X and Y will therefore violate the requirement that all terminals need to be ordered by an asymmetric c-command relation between categories dominating them.)

The LCA will remain silent, however, about cases where multiple categorial projection does not occur in a configuration where more than one head is immediately dominated by the offending phrase. For example (9b), a head complement structure (where XP is the complement of Z) and (9c), an adjunction configuration (with XP adjoined to YP), are both allowed by the LCA. (Z asymmetrically c-commands X in (9b) and XP asymmetrically c-commands YP and Y ordering the terminals appropriately, as required by Kayne's condition.) The uniqueness requirement on projection thus does not follow from the LCA, except in the special case of (9a).

A third condition, additional to the PPP is necessary to ensure the locality of the projection relation:

- (10) Locality  
if X projects  $X^{\max}$  then there is no category C such that  
 $X^{\max}$  dominates C, C dominates X and C is not a projection  
of X

(10) excludes configurations like (11a). (11b), where the lower YP may be interpreted as an intermediate level projection or as a segment of adjunction exemplifies (11a).



- (11) a. [<sub>XP</sub> [<sub>C</sub> X]]  
 b. [<sub>XP</sub> [<sub>YP</sub> X [<sub>YP</sub> Y ]]]

The PPP, requires that every phrase P has a head, namely the one that projected P. Together with the uniqueness and locality requirements and the extended structure preservation condition in (7), the PPP entails also that every phrase must have a unique head, ie.(12):

- (12) \*X, when X is not maximal and is immediately dominated by  
 $Y^{\max}$ , unless  $X = Y$

As stated in (12)  $X^{\min}$  cannot be a complement or a specifier of some other projecting head Y. An LI  $X^{\min}$  distinct from  $Y^{\min}$  cannot be immediately dominated by  $Y^{\max}$ , since if  $X^{\min}$  is not word-internal then it must project some phrase ( $X^{\max}$ ) by (7), and we can show that  $X^{\max}$  will intervene between  $Y^{\max}$  and  $X^{\min}$ . I shall do this by establishing that the assumption that  $Y^{\max}$  immediately dominates  $X^{\min}$  leads to a contradiction. We know that  $X^{\max}$  dominates  $X^{\min}$  by (4). Furthermore  $X^{\max}$  is distinct from  $Y^{\max}$ , the phrase projected by Y by (8). But  $Y^{\max}$  cannot intervene between  $X^{\max}$  and  $X^{\min}$ , by the locality requirement of (10). Hence,  $Y^{\max}$  cannot immediately dominate  $X^{\min}$  (it can only immediately dominates  $X^{\max}$  that in turn dominates  $X^{\min}$ ).

To sum up so far, the PPP expresses the idea that syntactic categorial structure is projected from the lexicon. The PPP states that all syntactic categories are related to the lexicon: they must either come from the lexicon or be projected by categories which do. That a phrase must have a head follows from the PPP, that is from the fact that all phrases are projected by their heads. Another assumption is the extended structure preservation requirement (7) according to which a precondition for a non word-internal lexical element to enter the structure is for it to project (create a phrase). That a phrase must not have more than one head will follow the PPP together with the assumptions of uniqueness in (8) and locality in (10).

Of course extended structure preservation, like uniqueness and locality are so far only stipulated and all three are in need of an explanation. Before going further in trying to understand why these conditions on phrase structure should hold (see section 4), I would like to make some comments on several concepts that current theories generally assume, but which the discussion has so far avoided.

### 3.2 Some remarks on adjunction and intermediate projections

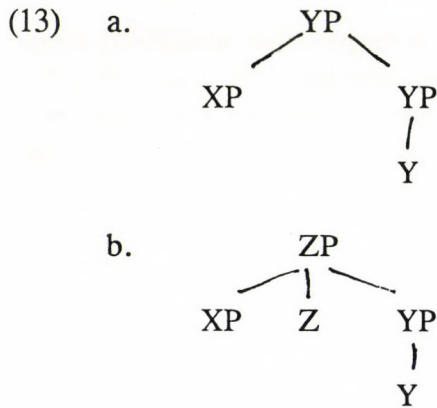
Notice first of all that the theory of phrase structure in the previous section is neutral with respect to the question of binary branching: a condition ensuring this may or may not apply in addition to the PPP and related conditions.

Current theories of phrase structure diverge from the simple picture which only contains the configuration where a phrase dominates a head and a number of other phrases in two major but related respects. An intermediate X' level is assumed between the head and the phrasal node and the configuration of adjunction is allowed in addition. What is the status of adjunction and of intermediate projection levels given the theory of section 3.1?

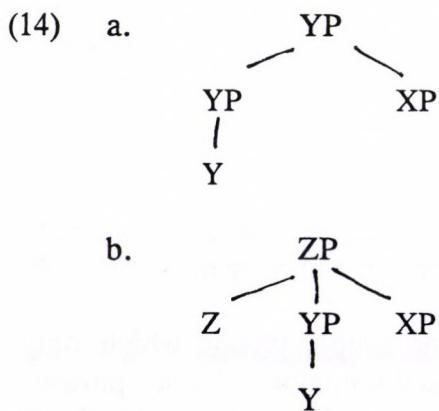
These two configurations can be reduced to one if, as proposed by Kayne (1995), the intermediate X'-level is treated as the lower segment of adjunction. It would be quite possible



to graft a segment-category distinction, and with it a theory of adjunction onto the theory of phrase structure as developed so far. But a simpler alternative might be to assume that there is no special adjunction configuration. Various arguments have been put forward that adjectives and adverbials, which have typically been treated as adjoined elements must in fact occupy either the head or the specifier position of some higher projection (Sportiche 1994, Cinque 1993, 1995). Under this option, instead of left-adjunction of XP to YP as in (13a), we will have the configuration in (13b) with the higher head Z. Z may or may not be invisible and/or transparent for selection (selectional requirements may be satisfied here by the lower head Y).



As for right-adjunction, this cannot exist in a strictly binary branching theory like that of Kayne 1995, where complements of embedded heads correspond to right-adjoined elements. As is well known, various tests suggest strongly that right adjoined constituents are in fact higher than a general condition like the LCA allows them to be (cf. eg. Williams 1994, Brody 1994). If structures are not necessarily binary branching, then these problems will not arise. Suppose that they are not. A possible alternative treatment of right adjunction might be to then take the element A adjoined to constituent B as an additional complement of a **higher** head (rather than of a lower one as in the binary branching account) whose preceding complement is B. Instead of structures like (14a) we will have (14b):



Like in the case of "left-adjunction", the higher head Z may be invisible and transparent for selection.

Consider next the question of intermediate projection levels. As noted one possibility

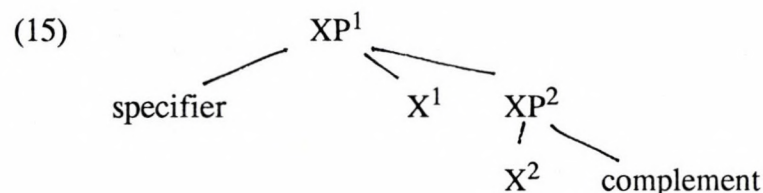


is to follow Kayne and treat the intermediate projection as a segment of adjunction. If however adjunction does not exist, then a different account is necessary. But the PPP, as it stands, allows a word to project more than one phrase. Given the way the locality condition is formulated in (10), a phrase does not have to immediately dominate the word that projected it, they can be separated by a phrasal node of the same type. Thus the system above allows non-maximal projections.

We could define the difference between a maximal and a nonmaximal phrase relationally: since nonmaximal phrases are not visible for the computation this will never cause the type of problem discussed in connection with relational definitions for all projections levels. Given the invisibility of non-maximal phrases, we can assume that no chains can be formed which contains such an intermediate projection as a member. Thus no Uniformity violation can arise.

Again let us consider briefly an alternative theory. Intermediate projections are not visible for the grammar. The best explanation of this fact would be if they did not exist at all. Let us suppose that they do not and eliminate the intermediate X' level. A word can then project only a single phrase. The question arises, how specifiers and complements can be distinguished. For many cases the checking configuration will provide the answer: the specifier is the element that undergoes checking. This will need to be extended to specifiers of those projections that instantiate adjunction in the impoverished system tentatively suggested above. But the specifiers of lexical categories probably do not participate in a checking relation with the lexical head. Here a different solution is necessary.

We can differentiate specifiers and complements of lexical heads without postulating either adjunction structures or the existence of categories that are neither word-level nor maximal projections by an analysis partly in the spirit of Larson's (1988) work. Suppose that we take a phrase to consist of an internal XP that includes the head and its complements and an external XP-shell that contains an empty head and the specifier or specifiers of X as in (15). The empty head  $X^1$  and the lexical head  $X^2$  are then taken to form a unit, -- a head-chain.



We could then take the specifier to be that sister of the higher head that does not contain the lower head, while the complement(s) would be simply the sister(s) of the lower head. Notice that the tree in (15) is only partly Larsonian, since although it involves an higher shell, it is not binary branching.

The solution, as it stands, inherits a general problem of Larson's empty shell approach. It is incompatible with the Generalized Projection Principle, which requires that categorial projection and the selectional properties of a head must be satisfied in the root position of its chain. This problem carries over to the analysis of the phrase in (15). In this case the subject is not in the same phrase ( $XP^2$ ) that contains the root position of the head-chain. The spec in (15) would therefore have to be selected from the position of  $X^1$ , not the root position  $X^2$  of the [ $X^1$ ,  $X^2$ ] chain. Furthermore the higher head  $X^1$  projects an XP, again in spite of not being in the root position of its chain.

One possibility is to assume that the higher head creating the "empty shell" is in fact



not empty but is itself an abstract lexical element, one that carries the appropriate categorial features and selectional requirements of the lexical item whose features are shared between a number of head positions (This modification of Larson's approach is suggested in Brody 1993a, 1995a, see also Koizumi 1993, and Collins and Thrainsson 1993 and also Chomsky 1994, 1995 for similar proposals and additional argument.) Multiple argument verbs under a Larsonian analysis would all require such a decomposition treatment. Let us apply this analysis to the present problem of eliminating the intermediate X'-level in terms of a structure like (15). If X is decomposed into X<sup>1</sup> and X<sup>2</sup> and categories standardly taken as sisters of X' and sisters of X are distinguished as sisters of X<sup>1</sup> and sisters of X<sup>2</sup> then also simple transitive and intransitive heads must decompose into two heads. Eg. the verb see would have to be composed of an agent selecting segment and a non-agentive SEE, something like the passive was seen.

(In Brody 1994 I raised an apparent problem for this approach to the question of the intermediate X'-level: with heads that assign no theta role to their subjects, specifier and complements could be distinguished only at the price of postulating a fully empty head. For example seem would have to decompose into a higher head that does not select its subject and which does not appear to contribute in any other way and a lower one which is exactly like seem. The problem arises however only if the expletive subject is generated VP-internally. If a verb like seem simply has no VP-internal subject, then there will be no question of how such a subject can be distinguished from the complements.)

#### 4. Assembly of syntactic structures

##### 4.1 Chain, Project and Insert

The discussion of phrasal projection (in section 3.1 above) has raised several questions. I would like to show that a version of the theory of the assembly of syntactic ((L)LF) structures proposed in earlier work provides straightforward answers.

This theory postulates three operations: Chain, Project and Insert. Chain forms chains by creating copies. (I assume that it may create multiple copies, to allow multi-member chains.) Project adds a phrase P to LI and establishes the relation: immediately dominates(P, LI). (Recall that I assume that P is simply a word-external copy of some features of LI.) Chain and Project are unordered and create what we may call, the "(syntactic) input list." (The concept of input list is different from, though related to Chomsky's concept of "numeration.") The input list then consists of several types of objects: (i) LIs, (ii) copies of LIs, (iii) phrases dominating LIs (LIPs) (iv) copies of LIPs.

Although some of the objects in the input list can be complex, they all involve a single lexical item. The input list thus can be taken as the normal form in which LIs are presented to syntax. The operation of Insert then applies to the elements of the input list. Insert establishes immediate dominance relations. (Notice that since a chain is a set of copies it is not a member of the input list, although members of the chain are members of the input list).

For a concrete example consider (16) with the simplified structure in (17):

- (16) Jean embrasse Pierre  
 (17) [IP NP V+I [VP (NP) (V) NP\* ]  
 (18) a. Chain V, V  
 b. Project NP\* > N\*, NP > N, VP > V, IP > I  
 c. Chain NP > N, NP > N  
 d. Insert all



- (19) a.  $V, VP > V NP^* > N^*, NP > N, NP > N, IP > I$   
 b.  $IP > NP, I > V, IP > VP, VP > NP, VP > NP^*$   
 (where  $X > Y$  means  $X$  immediately dominates  $Y$ )

If Chain applies before Project, it creates a head-chain as in (18a). If it applies to an element after Project applied to it, it creates a copy of the LIP, hence an XP-chain as in (18c). I assume that when Chain applies to LIPs, it creates a copy of the whole of the project relation, ie. it does not simply copy the phrasal node, but also the LI that projected it and the relation of immediate domination created by Project. Chain and Project in (18a,b,c) create the input list shown in (19a). Notice that the input list is not unstructured: it has two types of relations between its members: the copy relation and the immediate dominance relation. Finally Insert applies, relating elements in the input list by simultaneously establishing the further immediate dominance relations in (19b). Insert can only add relations, but cannot contradict hierarchical relations established by Project in the input list.

There are essentially two core concepts the theory is built on: the concept of copy and the structural notion of immediate domination. Both concepts are involved in the notion of projection: a projection of LI is a copy of a subset of the features of LI that immediately dominates LI. Only the notion of copy is involved in the Chain operation and only immediate domination in Insert. A major advantage of such a system is that the structure is built in one step, there are no intermediate syntactic structures, ie. no structures distinct from LF where lexical items are related to each other. (Notice that although the input list is structured, it is not a syntactic structure: all elements and all relations involve only a single lexical item.) The theory is thus able to explain the basic minimalist generalization that no conditions can hold on non-interface structures. This is because they do not exist

Numerous questions arise about this theory of one-step assembly of syntactic structure. Many of these are not specific to this system like for example what sort of word-internal structure should Insert establish. Take the second element of (19b): " $I > V$ ". Let us assume for this case that in addition to the word-external copy of (some feature of) LI (ie. the phrase), there are also non word-external projections. Such a projection of I (the highest one that is not a phrase) will then dominate V.

Another question that is only partly specific to this framework has to do with the notion of copy. Since XP-chains are formed by copying an  $X^{\max}$ , there must be a nondistinctness requirement on copies in chains to ensure that the same argument and selectional structure is inserted in all copies/members of XP-chains. For example we need to ensure that the principle C violation indicated in (20) can be ruled out at or beyond LF. This cannot be done if the chain-member in the lower (bracketed) position is simply the empty XP projected by the highest head of the antecedent (DP projected by which in (20a) and PP projected by to in (20b)).

- (20) a. \*Which claim that John<sub>x</sub> was asleep did he<sub>x</sub> deny (which claim that John<sub>x</sub> was asleep)  
 b. \*To John<sub>x</sub> he<sub>x</sub> gave a snake (to John<sub>x</sub>)

The status of chain-members as copies must be accessible to the post-LF interpretive systems in every minimalist framework: copies that are members of the same chain must be distinguished at least at LF from accidentally identical LIs and LIPs that are not chain related. Suppose then that structures in which two copies/chain-members dominate distinct elements cannot be interpreted, that these are not proper copies. The nondistinctness condition will thus



constrain the selectional requirements and therefore both chains members in (20a,b) will dominate the same elements. Take for example (20b). Here the chain is formed on the PP that was projected by the preposition to, ie. on "PP > to". The preposition in both copies selects a complement which then must be the same in both copies by the nondistinctness condition. The account is the same in the case of (20a), with the selectional requirements of heads applying recursively. The head of the highest DP in the copy selects an NP, which must have been projected by the noun claim given nondistinctness. This noun then selects a CP etc.

(In Brody 1994 I took "Near John he saw a snake" to be grammatical on the coreferential reading. If it is in fact not better than (20), then it needs no additional comment. If however I was correct in taking this reading here significantly better, then we can attribute this improvement to the option of not chain-relating the adjunct PP to the IP-internal position, --an option not available in the case of the selected PP.)

Notice that the copy of the lexical item involved in projection, serving as the phrasal node, appears not be subject to this interpretive non-distinctness condition: the head of a phrase never dominates elements like the complement and the specifier of the phrase, which the phrase contains. One possibility would be to take the copying involved in projection to be purely mechanical, the copy status of projection not being accessible to interpretation. But optimally we would not want to assume that only certain copies are taken to be copies by the interpretive components. In fact there is no need to make this assumption, since we can distinguish the two cases in terms of an independently motivated distinction. The copy operation involved in chains (at least in those corresponding to the "overt movement" relation, see section 4.2 below) targets a category together with its content. The copy relation involved in categorial projection targets the category only and ignores its content, in fact it probably targets only a subset of the category's features. For example the selectional or the phonological features of an LI will not be present on the LIP. In this respect categorial projection appears to be similar to the copy relation involved in "covert movement" type relations (see section 4.2). If this treatment of "covert movement" is correct, then the natural distinction having to do with the target of the copy relation is independently necessary.

Let us return then to the questions raised by the theory of phrase structure set out in section 3, namely (21):

- (21) a. Why does the PPP (each phrase is projected by a LI) hold only for all and only non word-internal heads (cf. (7))?  
b. Why is projection local (no intervening elements between LI and the phrase it projects, cf. (8))?  
c. Why is projection unique (each phrase is projected by a unique head, cf. (10))?

(21b) and (21c) receive an immediate answer, given the above theory of (L)LF assembly. Project applies before the syntactic structure is created (by Insert), and it applies separately to each head. Hence two LIs cannot project the same phrase and no 'foreign' projection can ever intervene between a head and its projections in the input list. Since Insert cannot modify the hierarchical relations established by Project, the conclusion carries over to fully formed syntactic representations.

As for extended structure preservation (7), recall that the impossibility of word-internal phrases, (7b), has been attributed to the fact that morphology does not tolerate such



constituents. The symmetry of (7) suggests an extension of this condition to (7a). Suppose we said that parallel to (7b), (7a) is due to syntax not tolerating non-phrasal elements. This would be an elegant modular solution, but unfortunately the condition is clearly incorrect: both phrases and nonphrasal elements (words) play a role in syntax. But let us reconsider this idea in the context of the system of LF assembly outlined. The modular solution is made available here by the separation of Project, where words play a syntactic role, and Insert, where they do not. So assume that Insert is modular in the relevant sense:

(22) Insert relates LIs to LIs (morphological application) and phrases to phrases (syntactic application)

(22) entails that all non word-internal heads must project a phrase. If an LI does not project a phrase then only morphological Insert can apply to it, hence it will be word-internal. (Since Insert cannot destroy the hierarchical relations established in the input list, LI cannot dominate a projecting head H: HP then could not immediately dominate H.) It also follows from (22) that there can be no word-internal phrases: again these could only arise by Insert non-modularly combining LIs and phrases.

The theory of (L)LF assembly involving Chain, Project and Insert was originally constructed as a system that can build syntactic structures in one step from input lists. Since it did not create intermediate syntactic structures it explained the non-availability of these. As I just showed, the theory entails also the three basic stipulated properties of the theory of phrase structure: extended structure preservation, uniqueness and locality. While the lack of intermediate structures is a property that the present theory shares with the "no syntax" view outlined in section 1, without some theory like the proposed one of how structures are assembled, the three basic conditions on phrase structure would remain stipulative. The account therefore provides evidence for the assumption that syntax, in the sense of an assembly system, in fact exists as part of the system of grammatical competence.

#### 4.2. F-movement and pied piping

The account of chain formation in Brody 1993a, 1995a, summarized and somewhat modified above incorporates what is in effect a "pied piping" hypothesis. Both head and XP chains are formed on an element that contains only a single lexical item. In the case of XP-chains this is the highest head of the phrase whose copies are the members of the chain. "Pied-piping" of the rest of chain, ie. filling out all the XP copies with material additional to this highest head is due as we have seen to selectional requirements applying recursively.

Chomsky 1995 presents a different theory of movement and chain formation which shares the general idea of pied piping with the above account. He proposes that movement can only take place to establish a checking relation, and for this only a feature F needs to move. Movement of categories occurs only in the overt component of the grammar and this is due to F-movement pied piping the whole category. Such pied piping in overt movement is forced by PF considerations. (There is an additional assumption that a set of features (formal features, FFs) are mechanically pied piped in both overt and covert movement.)

For example in "Whose book did you read" the +WH feature must move to establish a checking relation with the corresponding feature on the C node. It must pied pipe the word who, otherwise the PF features of this word would be scattered at PF, a state of affairs naturally taken as resulting in an ill-formed representation. The genitive 's must also be pied piped due to its affixal nature, thus whose must move together. But whose is not a syntactic



object, it is not a constituent. Hence the whole phrase whose book must move together.

Abstracting away from the difference between movement and chain-formation (cf Brody 1995a for discussion), we see that the two theories have much in common. Both accounts assume that chains are formed on a single element of the head or phrase that ultimately is the member of the chain. Under the present proposal this element is the head of the chain member, in Chomsky's theory it is the checking feature. The crucial difference appears to be that in the account defended here pied piping is due to LF requirements, whereas in Chomsky's theory it is a consequence of PF conditions.

There are reasons to prefer the LF pied piping approach. Notice first that if whose pied pipes book because whose is not a constituent, then it is not clear why which in (23) -- clearly a constituent-- similarly pied pipes book.

- (23) a. Which book did you read  
b. \*Which did you read book

More importantly, in the PF pied piping theory the question arises why pied piping does not take place only in the SPELLOUT component? Given a minimalist perspective it is particularly difficult to understand why a PF requirement should force complications in the syntactic computation. But the assumption of SPELLOUT pied piping does not seem to be correct: the position of the "moved" phrase has syntactic and semantic effects. To take an emblematic example consider the contrast in (23):

- (23) a. Mary wondered which picture of herself John saw  
b. \*Mary wondered when John saw a/which picture of herself

If anything beyond the +WH feature (or the formal features of the wh-word) remained in situ in syntax, then we would expect (23a) to behave syntactically and semantically in a parallel way to (23b). This is incorrect however. Thus the contrast between (23a) and (23b) would be impossible to account for on the SPELLOUT pied piping hypothesis.

On the other hand as pointed out in earlier work cited above, there is evidence for pied piping being LF driven. The adjunct-argument asymmetry in reconstruction (Lebaux 1989) falls out from the non-distinctness requirement and projectional requirements discussed earlier.

- (24) Which claim that John<sub>x</sub> made did he<sub>x</sub> later deny (which claim)  
(25) \*Which claim that John<sub>x</sub> was asleep did he<sub>x</sub> deny (which claim that John<sub>x</sub> was asleep)

The principle C violation in (24) where, the relevant name, John is inside an adjunct (the relative clause) is weaker than in (25) (= (20a)) where it is inside a complement clause. As we have seen, selectional properties together with the nondistinctness requirement force the name in the complement to present in the bracketed copy in (25), and a principle C violation results. In (24) in contrast no selectional requirement forces the presence of the relative clause and the nondistinctness condition also allows its absence in the (bracketed) copy. Hence there is a structure of this sentence on which no principle C configuration obtains. Clearly the PF-triggered pied piping account will not be capable of capturing such a distinction, which falls out under an appropriately constructed LF pied piping account.



Chomsky 1995 brings up another consideration: "The computation "looks at" only F [...], though it "sees" more. The elementary procedure for determining the relevant features of the raised element x is another reflection of the strictly derivational approach to computation." ( section 4.4, p27) Thus for example in (26) there is no question of determining where the WH-feature is located inside the complex wh-phrase pictures of whose mother since the computation looks at such features directly: pied piping of the rest of the phrase is only an additional matter.

(26) Pictures of whose mother did you think were on the mantelpiece

In reality however, the elementary procedure for determining the relevant checking feature is property of the pied piping theory. As we have seen a representational pied piping account is feasible (and also quite well-motivated), hence the question of derivationality does not seem relevant. Note furthermore, that in any case the pied piping theory does not seem to genuinely achieve a result here. The relation between the XP (in (26) the wh-phrase) and the checking feature F (in (26) the WH) remains mysterious also on the pied piping account. This is of course true of both the LF and the PF triggered version of the pied piping theory.

On the other hand the PF-triggered pied piping theory appears to create a genuine problem within the standard minimalist framework in that it creates a duplicate mechanism that appears conceptually and empirically unjustified. Consider a grammatical structure where movement without pied piping has taken place. This could in principle be due not only to the covert nature of the movement but also, as Chomsky notes (section 4.4, p23), to overt movement failing to pied pipe for whatever reason as for example in Watanabe's (1992) theory. There is no genuine evidence for making the theory more permissive in this inelegant way. See Brody 1995a for a critical discussion of Watanabe's theory. (The problem is in fact more general: in the versions of the minimalist theory that allow the SPELLOUT point to be distinct from (L)LF, empty categories can also be inserted both overtly and covertly.)

Let us consider also briefly the question of how overt and covert "movement" structures can be distinguished in the present framework. The simplest assumption is that the distinction does not pertain to syntax at all, that it is only a matter of SPELLOUT positions: in overt movement a higher copy, in covert movement a lower copy is subject to SPELLOUT. It seems to me that in a framework that assumes that there are no covert A'-movement relations there is little reason to depart from this simple hypothesis. If however there exist chains at LF corresponding to what used to be treated as covert A'-movement (cf Brody 1995a,b), then this will create problems for the simple SPELLOUT hypothesis. So for example if the relation between the wh-in-situ and the spec-CP position where it is interpreted is a chain relation, then spec-CP must not contain a full copy of the wh-in-situ at LF:

- (27) a. John wondered which pictures of himself Mary bought (which pictures of himself)  
b. \*John wondered which girl (which girl) bought which pictures of himself

If the spec-CP of the embedded clause contained a full copy of the wh-in-situ which pictures of himself, then we would expect (27b) to be on a par with (27a), the anaphoric element should be appropriately bound by the matrix subject. But this is incorrect, and this suggests



strongly that the higher position in the chain of the wh-in-situ must not contain a copy. In earlier work (Brody 1993a, 1995a) I treated these structures in terms of what I called "expletive-associate chains". Such chains expressed relations standardly treated in terms of LF-movement. In expletive-associate chains the chain-forming associate always remains in situ and the higher positions in the chain are not occupied by copies, but rather by an expletive element (or copies of this expletive). The expletive can carry features of the associate, --this accounted for various "agreement" effects (like checking of the WH-feature in covert wh-structures or subject verb agreement in there-associate structures etc.).

Chomsky's 1995 theory of covert movement as movement of formal features (FF-movement) only does not essentially differ from this proposal. If we abstract away again from the representational/derivational difference, the major difference we find is that FF-movement is head movement, whereas expletive-associate chains may be either head chains or XP chains. Without attempting to resolve the issue, I note that what evidence currently exists appears to favour the hypothesis that chains corresponding to covert movement relations can be phrasal.

(28) There arrived three men

Raising of FF(a man) in (28) to T violates the head movement constraint, as Chomsky notes. The assumption that FF can be phrasal would avoid this problem. There are then two options: either FF is an additional specifier of T or FF is identical to there which spells it out. (The second version is fully equivalent to the expletive-associate chain solution.)

## 5. Explanations of the Generalized Projection Principle

### 5.1. A non-syntactic account

The discussion of categorial projection would remain incomplete without the Generalized Projection Principle (GPP), a major and pervasive condition, one effect of which is the restriction of categorial projection to root positions of chains. Although the existence of deep structure as a distinct level of representation is quite dubious there are not many reasons to doubt the existence of the major generalization it expressed (Brody 1993b, 1995a, see also Chomsky 1993, 1994, 1995 for relevant discussion). This generalization, captured by the GPP refers not only to categorial projection but also to thematic selectional requirements, and in fact to syntactic and semantic selection in general. (I assume therefore that the GPP is a principle of the interpretive component.) All these requirements hold in the root positions of chains.

Thus for example a verb V raised to some higher functional projection, say C, never projects a VP here: categorial projection holds only in the root positions of chains. A V in C furthermore never forces the specifier and the complements of C to satisfy the selectional requirements of V: selectional requirements hold only in chain-roots. (For a potential well-defined set of principled exceptions see below.) I argued in earlier work that an appropriately formulated projection principle is both compatible with and necessary in a minimalist framework. In addition I attributed to the GPP the restriction against movement into a thematic position on the assumption that the GPP requires that selectional, including thematic, features not only must hold but also must be satisfied by root positions:

(29) Generalized Projection Principle

Projectional (categorial, thematic, selectional) features



must hold in root positions of chains, thematic and selectional features must also be satisfied by root positions of chains

Notice that in contrast to selectional projection, categorial projection apparently can be satisfied by non-root positions, namely by phrases in XP-chains.

In what follows I would like to summarize and somewhat revise the explanation of the GPP given in Brody 1995a. I shall then comment on the alternative (partial) explanation of the GPP proposed in Chomsky 1994 and 1995. To set out the explanation of the GPP I will concentrate on selectional features, ignoring for the moment the complication having to do with categorial projection being satisfied by non-root positions in XP-chains. Consider two chains that are to be related by a selectional feature F. Suppose that F must identify all positions of the chain to which it is assigned and that all positions of the chain whose member assigns F must be marked as having assigned F. This second requirement is also natural in the framework of copy theory: other members of the head chain whose LI has F are copies of LI and will therefore also have F. In other words I assume that the two chains will be properly related iff all members of both chains are appropriately identified as related in this way.

- (30) If a selectional (more generally projectional) feature F of a member of chain  $C_1$  selects (a member of) chain  $C_2$  then
- (a) all members of  $C_2$  must be identified as being selected by F and
  - (b) all copies of F on members of  $C_1$  must be identified as having been assigned

Let us say that an assignee position is selectionally identified if it has the appropriate selectional feature F while the assigner position is selectionally identified if it has some feature S indicating that proper assignment has taken place. Let us make also a simplifying assumption (I shall return to this below): that a head can only select and an XP can only be selected in a single position in a chain. Suppose finally that feature percolation in chains can only take place bottom to top, it is strictly upward directional. It follows that the selectional feature F must be assigned to the most deeply embedded position in the assignee chain, otherwise lower positions in this chain will not be selectionally identified. Similarly F must be assigned from the most deeply embedded position of the assigner chain, otherwise the feature S indicating the satisfaction of the selectional requirement F cannot percolate to all members of the assigner chain. All members (copies) of the assigner chain carry the selectional feature which can only be satisfied through percolation of S under the assumption that a selectional feature can only be assigned once in any given chain.

Chains where a non-root position is selected (including "movement" to theta positions) are now impossible: the selectional feature cannot percolate to the lower position of the chain, which thus fails to be identified. Conversely, no selection can take place from a non-root position either. A V for example raised to I or C now cannot select from the higher position of its chain since the information that this feature is satisfied could not reach the lower position of the chain.

The requirement that feature percolation in chains is strictly upward is in effect an equivalent of the derivational principle excluding lowering applications of Move. In a framework that assumes the rule of Move, a representation that is in violation of the GPP could have arisen in two ways. Either through raising in violation of the derivational



equivalent of the GPP prohibiting movement into a position that involves selectional features, or through lowering from this position. Downward spreading of the selectional features corresponds to lowering in a system incorporating Move. This needs to be excluded. Given this assumption the GPP reduces to the principle in (30) that all positions in a chain need to be selectionally (projectionally) identified. Thus while the GPP follows from fairly simple chain theoretical assumptions once the equivalent of lowering is excluded, the same explanation could not be translated into derivational terms in a system that assumes the operation Move. Excluding lowering rules would not help to explain why raising into a selected position is impossible.

Given the account so far, a selectional feature *F* on a member of a head chain can be satisfied in one of two ways. Either (a) directly, by assignment to some chain *C'* (in the root of *C'*) or (b) indirectly, through the upward percolation of the satisfaction feature *S*. The requirement that feature percolation can only take place to c-commanding elements restricts direct satisfaction to the root position, given the additional assumption that direct satisfaction can occur only once in any given chain. But while this uniqueness assumption is not unnatural by itself, it fits rather uneasily with the rest of the theory here. If all members of a chain carry the selectional features that need to be satisfied, it is not immediately obvious why the direct satisfaction of these should be restricted by the fact that they are members of the same chain.

We cannot simply dispense with the uniqueness requirement however. If direct satisfaction of a given selectional requirement in more than one position was allowed in general, then GPP effects would not follow any more. Consider a selectional feature which is assigned to two different chains from two different positions of its own chain. This must not be allowed since it would for example result in a *V* selecting an object in the *VP* and then selecting another one in its higher position in the chain, hosted by *I* or *C*. A more natural uniqueness requirement that still rules out the unwanted consequence would be to require that a given selectional feature must be assigned to a unique chain. Adopting this weaker condition, a selectional feature can now be assigned from multiple positions in an assigner chain and to multiple positions in an assignee chain as long as the assignee chain is unique. Thus a raised *V* can now select from its higher position as long as it selects a member of the same chain it selected a member of in its lower position.

In the case of experiencer predicates and several other related constructions there is evidence for exactly this type of multiple direct satisfaction of a thematic requirement (cf. especially Pesetsky 1995). The literature contains numerous arguments that the subject in (31) is an internal argument, cf. eg. Belletti and Rizzi 1988.

(31) This worries me

At the same time there are various indications like the possibility of passivization etc. suggesting strongly that the subject of (31) is external. Pesetsky resolves the conflict by allowing the same theta role to be assigned to more than one position in a chain. The availability of this option is exactly what follows from the present theory under the weaker version of the uniqueness hypothesis.

Let us finally return to the effects of the GPP for categorial projection. Recall that categorial projection is exactly like selection in that it is invariably initiated in the root positions of assigner chains. Our explanation of the GPP will then immediately predict this state of affairs once it is generalized from selectional features to cover also categorial



features, ie. features involved in categorial projection. Recall that categorial projection appeared to differ from selection in that a categorial feature can be received by non-root elements of an assignee XP-chain. For example all DPs in a DP-chain receive the categorial feature D from their respective D heads. But only one of these DPs is the root of the XP-chain. This appears to contrast with selectional features: a non-root position generally cannot be selected. But given the weakening of the uniqueness condition this generalization now has a set of systematic exceptions. As exemplified by (31) a chain may be selected in more than one position as long as it is selected by the same features. So categorial projection features in effect behave again just like selectional features: non-root members of XP-chains can receive a categorial projection feature (can be projected) because all members of XP chains receive exactly the same categorial projection feature.

The similar behavior of selection and categorial projection suggests that we should distinguish both of these features from checking features. I shall continue to refer to the former group as projectional features. I assume that both checking and projectional features relate chains, rather than categories, but they do so in a different way. Checking features are properties of chains. If such a feature is checked, then it is automatically taken to be checked in all copies (chain members) in the chain, independently of whether the copies are in a higher or lower position. A natural way of capturing this is to say that the checking feature is a property of the chain, rather than a property of the category from which it originates. Projectional features are different. Although a projectional feature also relates chains, it does not become a chain-property but remains the property of the category that is lexically specified to carry it. The difference presumably is another manifestation of the difference between grammaticalized "formal features" and other semantic features. (Notice that the similarity between selectional features and categorial projection casts some doubt on the idea that categorial features can serve as formal checking features. The features participating in categorial projection appear to behave like other non-formal semantic selectional properties and unlike formal checking features.)

Let us close this discussion of the GPP with an illustration of the interaction of this principle with the theory of (L)LF assembly. Consider a "moved" non chain-root head H inserted/substituted in some phrase, XP. Given the modularity of Insert, H must project a phrasal node HP, Insert can only combine a phrase with a phrase. There are two options to consider: the projected phrase HP may be either internal to the phrase, XP to which H substituted or it may force the category label of the target XP to be (H)P. Both options must be excluded. They both are excluded by the GPP: projection is restricted to root positions. Thus the GPP and the modularity of Insert together ensure that "moved" non chain-root heads must invariably be head-internal. It is important to see that there is no direct contradiction between modular Insert and the GPP. They only create a contradiction for word-external heads in non chain root positions. Insert requires all word-external heads to project a phrase and the GPP restricts all projection to originate in root positions of chains. Hence word-external heads that are not in root positions can neither project nor not project: they cannot exist.

## 5.2 Conspiracy theories of the GPP

Chomsky 1994, and Chomsky 1995 provide different explanations of certain effects of the GPP. The effects of the GPP fall into two classes. One effect is that (i) categorial projection always holds in the root position and another consequence as we have seen is that (ii)



selectional (thematic) properties also hold in and are generally (apart from cases like (31)) satisfied by root positions.

I shall focus on the issue in (i) first and start with the solution in Chomsky 1994. Here the generalization that categorial projection always takes place in root position results from conditions answering four distinct questions:

- (32) a. Why cannot a head in a substituted nonchain-root position project?  
 b. Why cannot a head in an adjoined nonchain-root position project?  
 c. Why cannot an XP in a substituted nonchain-root position project further?  
 d. Why cannot an XP in an adjoined nonchain-root position project further?

**(32a,b), --heads in non root position cannot project:** This is because the HMC would force such a raised head  $\alpha$  to substitute into or adjoin to the  $\alpha P$ , the phrase  $\alpha$  itself projected. This is prohibited by the fact that such "self attachment" would create an ambiguity: in such structures if the category/segment ( $\alpha P$ ) that dominates  $\alpha$  in the non root position can be taken to have inherited its label from  $\alpha$  in the non root position or from the  $\alpha P$  that  $\alpha$  projected in its root position. It is not clear however why the ambiguity of "self attachment" cases of adjoined and substituted projecting head should create a violation.

There is an additional case of adjoined non-root heads to consider for which the prohibition concerning self-attachment is not relevant, --namely where the head  $\alpha$  in the non-root position is adjoined to another head. This is the usual configuration of head chains and thus cannot be excluded in general.  $\alpha$  in this adjoined non root position of course cannot project either. But it is not clear what excludes here the configuration where the moved element projects instead of the head to which it adjoined.

**(32c), --substituted non-heads cannot project:** Turning to the question of why non heads ie. phrases cannot project further in non chain root positions, the substitution case is ruled out by the principle of Greed. Greed states that "Move raises  $\alpha$  only if morphological properties of  $\alpha$  itself would not be satisfied in the derivation". In a configuration like (33) if  $XP^*$  raises to K and then projects  $XP+$ , then  $XP^*$  ceases to be a maximal projection, given the relational definition of projectional status.

- (33) \* [ $XP+$        $XP_x^*$  [ $K$      $t_x$     ]]

$XP^*$  then will be "invisible for the computational system", which only sees non projected elements and maximal projections and therefore cannot "enter into a checking relation"

But the raised XP in (33) could satisfy Greed before it projects (cf.: "Adjunction to X' by merger does not conflict with the conclusion that X' is invisible to [the computational system of the grammar]; at the point of adjunction, the target is an XP, not X'." p.32.) So at the "point of substitution" the raised element is a maximal projection, not an X'.

There is an additional reason given in Chomsky 1994 for the ungrammaticality of (33), namely the uniformity condition, This also rules out (33) since here the trace of  $XP^*$  is maximal (by hypothesis) but  $XP^*$  is not. But as we have seen in section 2, there are strong reasons to reject the uniformity condition.

**(32b,d), --adjoined heads cannot project and non-heads (phrases) cannot project**



**further:** An adjoined element that projects would create the following configuration:

(34) \* $[\alpha \quad \alpha_x \quad [K \quad t_x \quad ]]$

Chomsky assumes that the two-segment category in adjunction involves two elements that have the status of a category: the lower segment and the two segments together. On this assumption Full Interpretation is violated in (34). Whichever of the two categories,  $\alpha$  or the two-segment element  $[\alpha, \alpha]$  is taken to be the head of the chain whose root is  $t_x$ , the other element receives no interpretation at LF and thus violates FI. Chomsky "conclude[s] that the target must have projected". Taking  $[\alpha, \alpha]$  to be the head of the chain is ruled out additionally by the UC. (This seems to work only where  $\alpha$  is non-maximal, if  $\alpha$  is maximal then so is  $[\alpha, \alpha]$ , hence the UC is not violated.)

The fact that in adjunction structures there is only one LF role for the two segment category  $[K, K]$  and the category corresponding to the lower segment,  $K$  is a general problem in adjunction however, whichever category projects. (This appears to be recognized in Chomsky 1995.) To allow adjunction to heads Chomsky invokes WI: the relevant restrictions again do not hold word internally. For non-heads he suggests that this fact essentially restricts adjunction to nonthematic categories (plus some other restricted cases, see the appendix below). But if WI can neutralize the problem when a minimal projection  $\alpha$  adjoins to  $X$  and  $X$  projects, WI will also neutralize the problem if  $\alpha$  projects. Hence the conclusion that the non root element cannot project does not follow.

Similar comments hold for nonminimal projections. If a configuration in which the target of adjunction is in a nonthematic position is permitted because no problem arises with FI, then in the same kind of position the adjoined element should be able to project without violating this principle. This again is probably an incorrect result.

Additionally, the assumption that there are exactly three elements in adjunction structures with the status of a category seems somewhat stipulative. Even granting that assumption, further questions arise. For example it is not clear why  $\alpha$  and  $[\alpha, \alpha]$  could not jointly serve as the antecedent of the trace.

Chomsky 1995 approaches the generalization that phrasal projection always takes place from chain-root positions somewhat differently. He rejects earlier formulations of Greed, and tentatively assumes that word-external adjunction does not exist (cf. also Brody 1994 and above). He bases much of the argument on the hypothesis that movement can only take place to 'immediately' establish a checking configuration. This hypothesis has a somewhat dubious status as it is in conflict with apparently well-established cases of successive cyclic movement where intermediate landing sites involve no checking (eg. adjunct wh-movement in English). Given this assumption however, word-external head movement cases will obey the generalization in (i), since if a moved head  $\alpha$  projects, it necessarily establishes a head complement relation with its target  $K$ :  $[\alpha \quad \alpha \quad K]$ . Checking relations can only be established in spec-head configurations. (Notice though, that it is only because non-branching projections do not exist in this system (in contrast to the theory of assembly adopted earlier where they do) that the conclusion that  $K$  is the complement of  $\alpha$  follows. In  $[\alpha \quad [\alpha \quad \alpha] \quad K]$ ,  $K$  would be a specifier.)

As before, the account does not extend to the word-internal domain. But the generalization (i) that a moved head does not project of course does. In the word-internal domain Chomsky has to distinguish two cases: structures where the moved head  $H'$  adjoins to some other head  $H$  that projects a phrase  $HP$  and structures where  $H'$  adjoins to some head  $H''$ , where  $H''$  itself is adjoined to  $H$  or to some other head adjoined to  $H$ . In the former case



if H' projects we get (35) since HP must have been projected before H' raised and projected.

(35) [HP [H' H' H]]

Chomsky proposes that the resulting structure is not well-formed: HP has no appropriate head. In effect he proposes a recursive filter (his (5)) which defines well-formed syntactic objects and rejects (35). But this seems to be an extremely dubious move. In a derivational theory the derivation necessarily provides a recursive definition of well-formed syntactic objects, no filter should duplicate this function. The independent evidence for such a filter is rather weak, but even if it was strong it would be difficult to be convinced of the necessity of such a filter: a well-designed theory just should not make it possible to violate it. (Hence in such a theory the GPP effect must be derived differently.)

Further conditions are required if H' illegitimately projects after raising and adjoining to some head H'' that is itself adjoined to a projecting head H. Here we get a structure like (36), where the considerations just reviewed are not relevant.

(36) [H [H' H' H'' ] H]

Turning to the question of why a moved XP cannot project further, the formulation of Greed that was taken to be relevant earlier is rejected here, and the impossibility of (33) is attributed to the uniformity condition, which I found numerous reasons to question in section 2 above.

Chomsky tentatively considers also another approach according to which the checking relation is asymmetrical and requires the checked element to be in the spec of the checking head. This would not be satisfied in (33). Here the head of K is taken to be the checking head and XP the checked element but XP is not a spec since its head projects the phrase, --instead K is the spec of XP. Chomsky notes several problems with this solution, but in any case it seems to contribute little: if checking is asymmetric in the way suggested then the question still remains: why cannot a checking relation be established in (33)? K cannot be the checking XP now, but X could in principle check K, unless we stipulate that the target of movement must be the checker. But that is barely different from the original problem to be explained: the target must project.

Let us consider next the solutions in Chomsky 1994 and Chomsky 1995 for the other GPP effect: that thematic properties must hold and be satisfied by chain roots. Chomsky 1994 does not discuss the problem of why the selectional (thematic etc.) requirements of heads have to hold in the root position of their chain, so I shall put this aside for the moment. The paper provides an account of the fact that thematic requirements must be satisfied by chain roots.

Consider the the hypothetical verbs HIT and BELIEVE that assign a theta role to their subject but no Case to their object. The GPP explains why such verbs cannot exist given the independently motivated assumption that the Caseless object position in structures like (37) must form a chain with the subject position: such a chain involves a non-root theta position.

(37) a. \*John [VP t' [HIT t ]]  
 b. \*John [VP t' [BELIEVE [t to VP]]]

In derivational terms the generalization translates as a ban on movement to thematic positions.



Chomsky 1994 attributes this also to the principle of Greed. The DP **John** in (37) cannot raise to spec-VP to pick up the unassigned theta role, since it does not need to do so to satisfy its own requirements.

Even if the DP originates in a non-theta position Greed would prevent raising to a theta position on the assumption that "the need for a theta role is not a formal property, like Case, that permits "last resort" movement". This explanation may not be general enough. The prohibition against movement to theta positions holds also for theta positions that are at the same time also Case positions. To illustrate, consider the hypothetical preposition IN, which is like "in" except that it does not assign Case. This should allow a structure like (38a).

- (38) a. I gave John the study IN t  
 b. cf: I gave John a book in the study

The GPP predicts that such structures are ungrammatical and that therefore prepositions like IN cannot exist. The explanation based on Greed does not have this consequence, unless not only structural Case positions but all Case positions are taken to be systematically distinct from theta positions.

Secondly the Greed based account allows movement to a theta position when this is made necessary by some other principle. This again seems to be an incorrect prediction. For example Relativized Minimality /Minimal Chain Link Condition (MCL) can force movement through a theta position in a derivation in which a later step satisfies Greed. To see this consider first Chomsky's analysis of the ungrammaticality of (39):

- (39) \*John reads often books

- (40) [<sub>VP</sub> John [<sub>V'</sub> v [<sub>VP/2</sub> often [<sub>V'</sub> reads books]]]]

He suggests that (39) has the structure in (40), and this is ruled out since the adverbial in spec-VP/2 prevents raising of the object **books** to spec-AgrO. "Note the crucial assumption that the subject **John** is in [SPEC, VP]... otherwise that position would be an "escape hatch" for the raising of **books**" p.33.

Consider in this light (41), that contains the verb HIT that assigns no accusative but is otherwise like **hit**. In (41) movement lands in a theta position:

- (41) John [<sub>VP</sub> t [<sub>VP/2</sub> often HITs t ]]

Here the DP **John** must raise outside the VP in order to get Case. But then as we have seen in the case of (39), relativized minimality/MCL forces it through [SPEC, VP] where it can pick up the subject theta role. Thus the nonexistence of a verb like HIT is not predicted.

As noted, Chomsky 1995 assumes that every application of Move must establish a checking relation and rejects the earlier formulation of Greed. He provides here a different account of why movement cannot land in thematic positions, one that is meant to generalize to answer also the question why all thematic selection holds in chain-root positions: "With regard to assignment of theta-role, the conclusion is natural in the Hale-Keyser theory. A theta role is assigned in a certain structural configuration;  $\beta$  assigns that theta-role only in the sense that it is the head of that configuration... Suppose  $\beta$  raises, forming the chain CH=( $\beta$ ,...,t)." "...the chain CH is not in a configuration at all so cannot assign a theta-role. In its raised position,  $\beta$  can function insofar as it has formal features: as a Case assigner or



a binder. But in a configurational theory of theta relations, it makes little sense to think of the head of a chain as assigning a theta role." (section 6, p1)

First of all it is not clear that it is any more natural (and it is certainly not independently motivated) to assume that a chain is "not in any configuration at all", rather than take it to be in multiple configurations. But more importantly perhaps the question is why "it makes little sense" to think of the head of a nontrivial chain as a theta-role assigner? In particular why it makes little sense to think of the head of the chain as a theta-role assigner when apparently the root of the chain can be a theta role assigner: "The trace  $t$  remains in the structural configuration that determines a theta role and can therefore function as a theta-role assigner..." It is true that in its raised position  $\beta$  cannot assign a theta role, or more generally activate any of its selectional features. But surely  $\beta$  is in **some** configuration in its raised position.

So what we in fact need is a characterization of configurations where selection is possible and those where it is impossible. An obvious possibility would be to assume that only categorially projecting heads can select. This will work once categorial projection itself is restricted to chain-roots. Although perhaps not unnatural it is clearly an additional stipulation that should be unnecessary. It is, if categorial projection and selection are both cases of semantic (non-formal) feature assignment constrained directly by the GPP as argued in section 5.1.

"With regard to receipt of theta-role, similar reasoning applies. If  $\alpha$  raises to a theta-position TH forming the chain  $CH=(\alpha, t)$ , the argument that must bear the theta role is CH, not  $\alpha$ . But CH is not in any configuration, and  $\alpha$  is not an argument that can receive a theta-role." Again it is not clear why the chain is not in any configuration. But independently of this, the paragraph is difficult to interpret. If the chain and not  $\alpha$  must bear the theta-role and the chain cannot, then we have a violation of the theta criterion also in the grammatical cases where the chain would receive the theta-role in its root position. Given minimalist assumptions  $\alpha$  cannot receive the theta-role before it moves, this would be resurrecting D-structure (on the reasons for rejecting D-structure cf. Brody 1993b, 1995a, Chomsky 1993). Thus we seem to be left without an account of the selectional effects of the GPP. Recall also that the discussion of these effects in Chomsky 1994, 1995 does not address the recalcitrant exception exemplified in (31) above. (Boskovic 1993 points out other cases apparently problematic for the GPP. See however Brody 1995a for an alternative analysis of these.)

In sum there are numerous uncertainties and unsolved problems associated with Chomsky's approach to the GPP effects. But the most important general objection is probably the familiar one with this type of approach that accounts for a major and simple generalization in terms of a complex conspiracy of principles. Such an approach appears plausible where the generalisation in question has a set of exceptions in need of an explanation: then the conspiracy can explain why the apparent generalization is true exactly for the cases for which it is. But a conspiracy account of a major and apparently unexceptional generalization makes the implausible and methodologically objectionable claim that the generalization is a --highly improbable-- accident.



## Appendix

Chomsky 1994, 1995 suggests a theory that radically restricts word-external adjunction but (in 1995 tentatively and partially only) retains this configuration in cases where the target has no theta role (expletive-associate chains) or where in his derivational system the adjunct is not present at LF (intermediate traces deleted by LF and "semantically vacuous" scrambling where LF reconstruction eliminates the scrambled element). These cases do not seem to provide strong motivation for retaining the configuration. LF adjunction of the associate to its expletive chain-mate is a problematic and probably unnecessary operation. The agreement facts, which constitute the main evidence for this operation can be treated without actual displacement of the associate, cf. Brody 1993a, 1995. This is also recognized in Chomsky 1995, --cf. section 4.2 above)

The necessity of adjoined intermediate traces in non-uniform chains is equally moot (cf. eg. Manzini 1992). Notice that the best and perhaps only strong evidence for their existence has to do with reconstruction effects (Lebaux 1989, Barrs 1986). For example in (42) binding of the anaphor appears to be licensed from the position that is internal to the intermediate trace/copy:

- (42) Which picture of herself did John think Mary told t Bill  
to buy t

But in the minimalist framework such evidence supports the copy theory of movement. Given the basic assumption of this framework, that conditions (like the binding theory) hold only at and/or beyond the interface levels, evidence like (42) shows that contrary to Chomsky's suggestion in this context, intermediate traces/copies must be present at LF (cf. Chomsky 1993, Brody 1995a).

As for scrambling, an alternative treatment of radical reconstruction may be to consider it to be stylistic displacement, ie. taking place in the SPELLOUT component cf. Aoun 1995. Chomsky suggests that LF reconstruction will provide an account of the contrast he finds between English topicalization cases like (43a) and (43b). Assuming that these work like scrambling, the expectation is that (43a), the adjunct case, is worse since forced reconstruction in this example will create a configuration that violates principle C. Since the fronted phrase "which pictures of John's brother" is not an adjunct in (43b), this example will not be similarly excluded:

- (43) a. Pictures of John<sub>x</sub>'s brother, he<sub>x</sub> never expected that I  
would buy  
b. Which pictures of John<sub>x</sub>'s brother did he<sub>x</sub> expect that I  
would buy

But forced reconstruction of adjuncts would lose the explanation of the contrasts like those between (44a) and (44b) or (44c) and (44d), which depend precisely on forced reconstruction of the selected argument but not of the unselected category internal to the fronted phrase, --cf. Lebaux 1989, and Chomsky 1993, Brody 1995a, for different ways of instantiating this idea. (Notice that no configurational distinction needs to be assumed to capture the difference, the selected/nonselected contrast suffices.) Now if topicalization is adjunction and adjuncts are reconstructed then (44a,b) should contrast with (44c,d) rather than (44a) and (44c) with (44b) and (44d):



- (44) a. Which claim that John<sub>x</sub> made did he<sub>x</sub> deny  
b. ?\*Whose claim that John<sub>x</sub> was asleep did he<sub>x</sub> deny  
c. The claim that John<sub>x</sub> made about this, he<sub>x</sub> later denied  
d. ?\*The claim that John<sub>x</sub> was asleep all day he<sub>x</sub> later denied

Thus the evidence for an account involving forced reconstruction of elements adjoined to semantically nonvacuous categories seems unconvincing. Its support for the more general claim that word-external adjunction exists in syntax (under restricted circumstances) is therefore weak.



## References

- Aoun, Youssef. 1995. Interception and reconstruction. Talk presented at the Vienna Round Table Colloquium
- Barrs, Andrew. 1986. Chains and anaphoric dependencies. Doctoral dissertation, MIT, Cambridge, Mass.
- Belletti, Adriana and Luigi Rizzi. 1988. Psych-verbs and Theta-theory. Natural Language And Linguistic Theory 6:291-352.
- Boskovic, Zeljko. 1993. D-structure, Theta Criterion and Movement into Theta Positions. Ms., University of Connecticut and Haskins Laboratories.
- Brody, Michael. 1993a. Lexico-Logical Form. Ms., University College London
- Brody, Michael. 1993b. Theta theory and arguments. Linguistic Inquiry 24:1-23.
- Brody, Michael. 1994. Dependence and Phrase structure. UCWP vol.6. University College London
- Brody, Michael. 1995a. Lexico-Logical Form. A Radically Minimalist Theory. Cambridge, Mass.:MIT Press.
- Brody, Michael. 1995b. Perfect Chains. Ms., University College London, to appear in Liliane Haegeman ed. Handbook of Syntax, Kluwer
- Cinque, Guglielmo. 1993. On the evidence for partial N movement in the Romance DP. ms., University of Venice
- Cinque, Guglielmo. 1995. Adverbs and the universal hierarchy of functional projections. Tromso GLOW abstract
- Collins, Chris and Hoskuldur Thrainsson. 1993. Object shift in double object constructions and the theory of Case. In MITWPL vol.18. MIT, Cambridge, Mass.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. MIT Occasional Papers in Linguistics, MIT, Cambridge, Mass.
- Chomsky, Noam. 1994. Bare Phrase Structure. ms., MIT, Cambridge, Mass.
- Chomsky, Noam. 1995. "Chapter 4". ms., MIT
- Kayne, Richard. 1995. The Antisymmetry of Syntax. Cambridge, Mass.:MIT Press.
- Koizumi, Masatoshi. 1993. Object Agreement Phrases and the Split VP Hypothesis In MITWPL vol.18. MIT, Cambridge, Mass.
- Larson, Richard. 1988. On the double object construction. Linguistic Inquiry 19:335-391.
- Lebeaux, David. 1989. Relative clauses, licensing and the nature of the derivation. Ms., University of Maryland.
- Manzini, Maria Rita. 1992. Locality. Cambridge, Mass.:MIT Press.
- Pesetsky, David. 1995. Zero Syntax. Cambridge, Mass.:MIT Press.
- Sportiche, Dominique. 1994. Adjuncts and Adjunction. Vienna GLOW abstract
- Watanabe, Akira. 1991. Wh-in-situ, Subjacency and chain formation. Ms., MIT.
- Williams, Edwin. 1994. Thematic Structure in Syntax. Cambridge, Mass.:MIT Press.



## Perfect Chains

### 1. Towards Perfect Syntax

#### 1.1. Duplications, conditions on chains, economy

In the Principles and Parameters theory representational conditions on various levels carried most of the burden of restricting syntax. In the minimalist framework representational conditions will only hold at interface levels, the only levels that exist. Let us assume the strong version of the minimalist hypothesis according to which syntactic interface conditions are "bare output conditions", that is conditions forced on syntax by the interpretive systems that are fed by the syntactic computations. Suppose furthermore that most of the effects of representational conditions of the Principles and Parameters theory turn out to be the effects of either bare output conditions or of conditions holding within the syntax-external systems. If so, then we may expect the syntactic computation, whose task is to assemble interface representations from a set of lexical items, to be near-trivial.

In any case the syntactic system may turn out to be significantly more trivial than in standard minimalist theories. One area where simplification is achievable is the derivational duplication of representational concepts. A central case is the independently motivated interface notion of Chain which captures the same relations as the derivational rule of Move, making the latter redundant and in fact untenable in a restrictive system. Notice that eliminating Move in favor of Chain does not necessarily make the theory strictly non-derivational: see the one step derivational theory of Brody 1995a,b. It will however necessarily make the architecture of the mapping system between lexical items and the two interfaces radically simpler: since there are no intermediate structures between the lexical input and the semantic interface, this level has to serve as the input to the SPELLOUT component. There is therefore only a single syntactic interface level, say the level of Lexicological Form (LLF) of Brody 1993a, 1995a, which is the input to both semantic interpretation and the SPELLOUT component.

There are various other related distinctions and duplications that seem equally dubious. For example Chomsky 1995 proposes a representational definition in addition to the derivational system of interface assembly (in effect an additional definition) of what counts as a well-formed syntactic object (cf. Brody 1995b for some discussion). Or take the additional distinction he makes between deletion (interface invisibility only) and erasure (essentially invisibility also for Move), where erasure occurs only if this would not violate the representational duplicate definition of well-formed syntactic object. Such duplications, and distinctions that build on these duplications should have no place in a restrictive system of syntax.

Another area where we might expect the system of interface assembly to be radically simplified has to do with economy conditions. Strong empirical argument is necessary to motivate the undesirable complication of the system of assembly which would make it powerful enough to compare derivations, --especially in view of the attendant computational complexity. It does not seem unreasonable to think that economy conditions will be eliminable without the introduction of any additional apparatus. More recent versions of economy conditions that "compare locally" seem to represent no major improvement in computational complexity. Suppose that at any stage in the derivation only the possible continuations of the derivation already constructed are compared. But clearly, for any



comparison of full derivations we can construct an equally complex comparison of "continuations", simply by embedding the appropriate structure. Notice that in order for the comparison to be meaningful, it cannot be restricted to look only at a single step in the relevant derivations. The derivations in which a step or a series of steps will be compared must all be carried out fully, if only to check that they converge, given the assumption --an inevitable one as Chomsky 1995 explains--, that economy conditions select only among converging derivations.

A third set of concepts and conditions that should have no place syntax-internally in a restrictive system of interface assembly pertain to the Chain/Move relation. Assuming that the relation is part of syntax proper, various conditions on this relation like Uniformity, c-command, Last Resort, Minimal Chain Link and Procrastinate should still be attributable to the syntax external systems (either as matters internal to the interpretive components or as properties imposed by these on syntax, ie. bare output conditions) or eliminable in some other way. Uniformity was discussed in this light in Brody 1995b, treatment of the other minimalist conditions on Chain/Move will be the main subject of the present paper.

## 1.2 Interactions with the interpretive components

In Brody 1995a I proposed an interface assembly system based on the notions of copy and immediate domination that complies with these restrictions. I used the term perfect syntax to refer to the theory that postulated such a near-trivial assembly system. There are two additional general constraints related to the interaction of the assembly system with the interpretive components that the standard minimalist framework does not incorporate, but we might expect perfect syntax to meet. One of these has to do with the nature of the interaction between syntax and the interpretive systems. In the standard minimalist framework it is suggested that imperfections in syntax are due to the effect of the interpretive components. Bare output conditions will be satisfied optimally, but this may necessitate departures from perfection. In particular it is often suggested that the fact that syntax has the Chain/Move relation is an imperfection due to output conditions. Chomsky 1995 contains also the further suggestion that the optimal operation of Move is covert (feature-movement only) and overt movement of categories is again due to interpretive (PF) pressures. These pressures could have to do for example with strong checking features holding at PF (although this assumption is rejected in Chomsky 1995) or with the necessity of moved features pied piping full constituents to avoid PF crash.

The idea that the components interacting with syntax force additional imperfect operations on this system seems to raise problems. Without further assumptions, whose nature seems unclear, for the external systems to cause the addition of imperfect operations to syntax it is necessary that perfection in syntax and the demands of the external system be in conflict. But it seems difficult to think of a case where the external demands could not be met in some alternative way without sacrificing perfection. If so, then given usual methodology, we would expect the system to choose the more perfect option. To look at the actual proposals concerning Move, consider the hypothesis that the existence of this operation is forced by the bare output condition of full interpretation. This requires certain features to be checked at LF by features of lexical items generated elsewhere. But if Move is an imperfection, there are in principle many ways in which it could be avoided without violating full interpretation. For example the checking features could freely delete or they could be generated in a position that is accessible to the lexical item without movement etc. It is not easy to see why all the



alternative options should be in principle unavailable.

(Notice incidentally that the proposal that Move is forced by bare output conditions interacts only with some redundancy with Chomsky's (1995) assumption that strong features are intolerable in the derivation. Given the additional distinction between +/- Interpretable checking features, Move will sometimes be forced by full interpretation (to eliminate weak non-interpretable features of hosts) and sometimes by both derivation-technological reasons and full interpretation (to eliminate strong non-interpretable features and (non-interpretable) "strength" of interpretable features). See below in sections 2 and 4 for related discussion.)

Similar comments apply to the question of overt Move. If overt movement of categories within the mapping between lexical items and (L)LF could be avoided for example by eliminating strong features, or by restricting overt movement to the SPELLOUT component. (See Brody 1995b for evidence that this latter possibility is not what actually happens.) Many other options are imaginable, which seem extremely difficult to rule out on principled grounds.

Let us avoid such problems and assume that the syntax-external components cannot force imperfect additions to syntax. This is then the first general condition promised above, related to the interaction of syntax with the interpretive systems. There will still remain at least two potential sources of imperfections. One can imagine that there might be cases of apparent imperfection: where syntactic perfection cannot be observed due to the limited expressive power of the interpretive modules. Non-apparent imperfection might also arise however, if the interpretive modules can cause the syntax internal imperfect instantiation of a perfect syntactic property or relation by not being able to interpret a perfect instantiation.

Returning to the concrete case of Chain/Move, this is simply an instance of the copy relation. It is not necessary to take this relation to be an imperfection, indeed within the assembly system of Brody 1995b this would not even be possible. In this theory the copy relation is involved also in categorial projection and is the only basic concept apart from the hierarchical notion of immediate domination. As for chains corresponding to "overt movement" of a full category, again this is simply the copy relation, where only the highest copy is visible for the SPELLOUT component. In the case of "expletive-associate" (Brody 1993a, 1995a) or "feature-movement" (Chomsky 1995) chains, the copy relation is not between full lexical items, but between features. See Brody 1995a,b for some discussion of why it may not be correct to eliminate syntax internally the distinction between chains where the copy relation is between full lexical items and chains where it is between features, by treating "covert movement" relations as full copies with a lower SPELLOUT position.

Given the distinction between feature copy chains and full category copy chains, it seems possible to view one of the two instances of the relation as an imperfect version of the other. Suppose that in the spirit of Earliness (Pesetsky 1989) or Transparency (Brody 1995a), we take the full categorial copy to be the default operation, and consider copying a proper subset of features as a degenerate case.

Suppose further that we assume that "covert movement" relations involve head-internal ( $X^{\min}$ -internal) feature-copies. This latter option will be taken only when the former is impossible. We can assume that the option of copying a full category is conditioned by the morphological properties of the head with which the copy will establish a checking relation. A head (more precisely a checking feature of a host head) may or may not have the ability to establish a head-external checking relation. (In fact we may need to distinguish checking features that license head-external but word ( $X^0$ )-internal relation, ie. an "overt"  $X^0$ -chain and features that license also word-external checking ("overt" XP-chains). Further, there is



evidence for three word-external choices: zero, one or multiple word-external checking relations.)

Licensing a head-external checking relation is matter of morphology, internal to the SPELLOUT component. When the relevant checking feature of the head is "weak" and does not license a head-external relation then only a "feature movement" chain can surface. In such a case a full categorial copy chain could not be expressed by the morphology. So this view dispenses with the strong/weak distinction in syntax and places it fully in the morphological component. Notice that it also eliminates the strong/weak distinction, in the sense of collapsing it with an independently necessary parameter which determines the number of specifiers (more precisely elements in the checking domain) of a given functional head. The analysis thus provides a case where the limitations of the interpretive, morpho-phonological component (word/head-external checking is not licensed) prevent the full expression of the perfect syntactic relation (copy of full lexical item) and force an imperfect instantiation (copy of features only, as these can be checked head-internally).

I argued so far for the assumption that the interpretive components cannot directly influence the syntactic computation, although they can filter its output. In this respect the semantic/conceptual and the SPELLOUT component behave similarly. There is however also a conspicuous difference between them: syntactic elements are generally identical to elements to which the semantic/conceptual system needs to refer. Suppose that the discussion of the strong/weak distinction above is on the right track and "strength", ie. the licensing of the head-externality of checking relations is not visible syntax internally. Then little reason remains to assume that syntax ever needs to make use of elements or features that have only morpho-phonological but no semantic/conceptual interpretation. This leads directly to the second general constraint pertaining to the interaction of syntax with the interpretive components. The perfect (L)LF assembly system should be constructed using only elements that the semantic/conceptual interpretive system provides. This condition, which I shall refer to as radical interpretability, is natural, and I will give some indication below that it may be tenable empirically.

Radical interpretability is related to the condition that requires syntactic representations not to contain non-interpretable features (or features without "effect" on the interpretive components), --call this interpretability. But the requirement that all features must have either semantic or morpho-phonological content is not only weaker but also stronger in one respect than radical interpretability. In contrast to interpretability, radical interpretability requires all syntactic elements to be (semantically) interpretable, but not necessarily actually interpreted in a given structure. The latter requirement seems too strong. As just noted, in perfect syntax there can be no distinction between interface invisibility and invisibility for the computational system. But the stronger construal of radical interpretability would rule out even the concept of interface invisibility necessary for example for checking theory (cf. section 4 below): if a syntactic feature is invisible at the interface it would receive no interpretation.

Although there seems to be no difficulty in taking the existence of the copy relation as such not to be a departure from perfection, as noted earlier the minimalist framework assumes several a priori unexpected conditions on this relation, which clearly could not be part of the perfect assembly system. In what follows I will argue that those conditions on Chain/Move that do not dissolve on closer examination should be thought of as either constraints on the interpretive components or as bare output conditions. In the next section I shall start with the c-command and the cyclic properties of Move. After critically examining two approaches in the standard minimalist framework to eliminating these stipulations



(sections 2.1 and 2.2), I shall turn to an improved analysis that is made possible by the assumptions of perfect syntax (section 2.3). Section 3 will discuss the MLC and section 4 looks at checking theory.

## 2. Strong features, c-command and the cycle

### 2.1. Weak vs strong checking features

In earlier versions of the minimalist framework it was assumed that weak checking features have to be checked by a moved element by LF and strong features by PF. This ensured that both weak and strong checking features forced movement to occur and strong features furthermore forced movement to occur overtly. Chomsky 1995 makes somewhat different assumptions about strong features. He argues that strong features not only force overt movement, they also ensure the cyclicity of this operation. He suggests that strong features once Merged must be checked/eliminated "quickly" by Move or Merge. "Quickly" is defined essentially as in (1):

- (1) The derivation terminates if an element H with a strong feature is in a category not headed by H

(1) entails the cycle for overt movement. Movement targeting a position in a given phrase P (and checking a strong feature of its head H there) must now precede movement targeting a higher position outside P. This is because a P-external position will necessarily be in a phrase not headed by H. Hence the strong feature of H will have survived in a phrase not headed by H, contrary to (1). (1) also ensures a form of c-command (m-command) for overt movement between the moved category and its trace. In order to overtly move  $\alpha$  to a position that is not in the minimal domain of a phrase that dominates  $\alpha$ , to let  $\alpha$  check a strong feature there, a structure must have been built which violates (1).

These appear to be welcome consequences: if possible we clearly do not wish to stipulate either c-command (which is only one of infinitely many possible structural relations) or the cyclic property for Move. This approach to the cycle and to c-command however does not seem promising for three reasons. First (1) appears to lack genuine independent motivation. Secondly its explanation, to be discussed below, is untenable and therefore (1) remains stipulative. Thirdly (1) would not be general enough. I shall take these points one by one, starting with the question of independent motivation. This involves mainly the question of strong features forcing overt movement.

Given the assumption that strong features cannot be inserted covertly, which Chomsky takes to be the consequence of wider considerations, (1) is taken to entail that strong features force overt movement, since the strong feature on a head H must be eliminated before the phrase H projected is merged with some other element. But covert movement will occur only after the whole structure has been assembled (after the SPELLOUT point). One problem here is that if H is the highest head in the tree then the theory predicts that its strong feature, introduced overtly, can be checked covertly. This is incorrect: a strong +wh feature on the root C for example cannot be satisfied by covert Move.

Notice that given Chomsky's assumption that strong features can only be satisfied by categories (section 2.1, p11), the strong +wh feature on the root C would force movement of a category rather than a feature, whether it is satisfied overtly or covertly. The PF pied-piping theory of overt movement includes a least effort type assumption that entails that



covert movement does not involve categories (section 4.4). But the least effort assumption is not strong enough to rule out category movement in this case since it requires nothing more than features to move, only when movement of nothing more is forced. But here movement of a category is forced by the strong feature under the assumption that strong features can only be satisfied by categories.

Note also that in any case the conjunction of the assumption that strong features can only be satisfied by categories with the PF pied piping theory creates a redundancy. The fact that overt movement must involve categories is now entailed by both, suggesting that at least one of the two ideas should be modified. Suppose we reject the PF pied piping theory (see Brody 1995b for arguments) and assume for the sake of argument that strong features must be satisfied by categories, whereas weak ones can also be satisfied by features. This should suffice then as a strong/weak distinction, there should be no need to duplicate this in terms of a derivational property of quick elimination of the feature. The distinction does suffice in a theory where (L)LF is the input to SPELLOUT, and therefore the question of invisible (covert) movement of categories does not arise. (For reasons not strictly relevant to the discussion of (1), and as noted in section 1 above, I assume a somewhat different theory of the strong/weak distinction, see also below in section 4.)

The direct empirical evidence for (1) is that it might allow adjuncts intervene between the two elements of the checking relation as eg. in (2). If adjunction does not close off a projection, then the strong feature of the inflectional head can be checked by the subject, still within the projection of this head. The dubious status of adjunction configurations in general (eg. Chomsky 1995, Brody 1994, 1995b), and in the particular case of adverbials (Cinque 1995), weakens this point considerably. If such adjuncts are heads or specifiers, then the argument will not go through.

(2) John probably has left already

The result is quite questionable also on directly empirical grounds: many clear cases of checking configurations require adjacency that does not tolerate intervening adverbials, for example wh-checking in English or focus in Hungarian:

- (3) a. Who (\*suddenly) did Bill discover  
b. Kit (\*tegnap) látott Mari  
Who+acc yesterday saw Mary

Let us next look at the question of whether (1) can be attributed to some more general consideration. In Chomsky 1995 it is claimed that (1) follows from (4), where (4) itself is claimed to be a consequence of other considerations.

(4) Nothing can join to a non-projecting category

"Non-projecting" here cannot mean a category C that is embedded in some phrase not headed by C. This is because in the standard minimalist framework under this interpretation (4) would exclude covert movement. So "non-projecting" in (4) must mean a category that does not project as a result of the joining operation, ie. (4) expresses an extension of the generalization that it is always the target of Move that projects (cf. Chomsky 1994, 1995, Brody 1995a,b for different approaches to this principle). But this reading is irrelevant since the requirement that the target projects does not entail (1). If Move lands in an embedded



position, its target could still in principle project.

It appears that we must choose the first reading of (4) and in order to distinguish weak and strong features we need to restrict it to overt movement:

(5) Nothing can overtly join to a non-projecting category

But although (5) does not refer to feature strength, it is just as much a stipulation as (1). It is simply a statement of the overt cycle which (1) therefore failed to explain.

Let us turn to the third set of problems with the explanation of the cycle and the c-command property of Move in (1): even if the principle was otherwise tenable, it would not be general enough. (1) stipulates that overt movement is cyclic. How about Merge? As Chomsky notes "it applies at the root only". Embedded Merge would be more complex than at the root. "Any such complication (which could be quite serious) would require strong empirical motivation." (Ch4.3, p11). Of course this does not entail the extension requirement for Merge, the remark just explains why a theory that entails it would be desirable. (A problem for the standard minimalist theory is that the same consideration applies to any embedded operation, overt or covert. Thus it applies also to covert object shift which is allowed in the minimalist theory.)

Thus in the theory based on (1), the c-command property of Move and the cycle are ensured only partially, and by a conspiracy of stipulations. (1) entails the cyclicity of overt Move, the stipulation in (5) would entail the extension requirement for overt Merge, --further conditions will be necessary to rule out covert countercyclic Move and covert Move to non c-commanding positions. If overt embedded Merge is taken to be ruled out for the same reason as overt embedded Move (as in (5)), then still further stipulations are necessary to rule out covert embedded Merge.

Chomsky also appears to note that the explanation of cyclicity and c-command based on (1) is not a full solution. He observes that "it would be interesting to strengthen [the] conclusion: to show ...that overt targeting of an embedded category (hence lowering and non-cyclic raising) is not possible, hence a fortiori not necessary" (section 4.1. p.16). But the remark seems to me to somewhat misstate the issue which does not have to do with the modality of the restriction but with its generality.

## 2.2. C-command by Merge

Crediting Kawashima and Kitahara (1994) and Eric Groat (p.c.) Chomsky briefly sketches a more general solution to the problems of cyclicity and c-command. This is based on the theory of Epstein (1994) where

(6)"C-command is just the relation that holds between  $\alpha$  and elements of  $\beta$  when  $\alpha$  is attached to  $\beta$  by Merge or Move" (section 4.1, p16)

Hence an embedded operation that attaches  $\alpha$  to  $\beta$  will establish no c-command relation between elements in the tree higher than  $\beta$  and  $\alpha$ . Given the further assumptions that all terminals must be ordered at PF and that terminals are ordered at PF only by c-command relations holding between them (or between categories dominating them) at LF (Kayne 1995), such embedded operations will be prohibited in the overt syntax. Thus overt Merge and Move must be cyclic. Similarly overt Move to a non c-commanding position, whether lowering or "sideways" (ie. where no c-command relation is established between  $\alpha$  and its trace) is







does not affect PF material. That "covert Move" relations do not affect the placement of the category and its phonology anywhere in syntax was argued in Brody 1993a, 1995a, see also Chomsky 1995, Brody 1995b. Of course the assumption makes it also unnecessary to consider the SPELLOUT point in the syntactic derivation to be different from LF.

The assembly system of Brody 1995a,b which satisfies the strictures of perfect syntax is not cyclic: the elements of the input list (a concept related to the notion of numeration) are joined simultaneously in one step. The cycle is incompatible with the requirement that the syntactic derivation create no intermediate structures between the input list and the interface level LF (Brody 1995a). But the question of how to ensure the cycle may be a pseudo-issue if the cycle in fact is unnecessary. Cyclic effects can be noticed only where the application of some constraint can be illegitimately avoided by a noncyclic derivation. But if the constraints in question in fact apply to or beyond the syntactic output representation (the expected case in perfect syntax, where they will be constraints on the interpretive components or bare output condition) then it will generally be impossible to avoid them through changing the derivational history.

A typical case is the late insertion of intervener type cycle violations for the minimal link condition (MLC). But if the MLC (or any other condition with the relevant effect) applies to the fully formed LF representation (cf. section 3 below), then the derivational history of the structure will be irrelevant: the effect of the MLC cannot be avoided at LF which necessarily contains the intervening elements. In (9) for example the the intervening subject *it* will necessarily be present between *John* and its trace.

(9) \*John seems it is certain *t* to go

(10) \*Who were [pictures of *t*] bought [pictures of *t*]

Similarly, the subject island violation of (10) cannot be voided by first applying wh-movement and then passive if the presence of a subject-internal trace is determined on the basis of the output representation.

Thus no cycle should be necessary in perfect syntax given the general architecture of the theory, where constraints apply to the output of the assembly system. There will be then nothing to explain: the optimal case. The question of excluding lowering operations may similarly turn out to be pseudo-issue due to the minimalist duplication of the concept of Chain by Move. In perfect syntax lowering and raising cannot be distinguished: chains are neutral with respect to this dimension. Again the situation is optimal: the theory is designed in such a way that there is nothing to explain. (The question of hierarchical directionality does arise in the treatment of the Generalized Projection Principle, but this is demonstrably a matter for the interpretive component since this principle constrains also nonsyntactic features. Cf Brody 1995a,b.)

All that remains then is the question of chains connecting positions that are not related by c-command. One possibility would be to exclude such chains by adopting a version of the Epsteinian solution to the assembly system of perfect syntax. We could require that there must be a syntactic relation at LF between the members of a given chain and assume that the only syntactic relation that exists is the one created by the operation of Insert (cf. Brody 1995a,b). Thus c-command is just the relation that holds between  $\alpha$  and elements of  $\beta$  on one hand and (irrelevantly for the present problem) the head of  $\beta$  and elements of  $\alpha$  (including  $\alpha$ ) on the other, when the operation Insert joins  $\alpha$  and  $\beta$  by making  $\beta$  immediate dominate  $\alpha$ . Alternatively, if ellipsis involves chains, as suggested by Chomsky 1995, then chains whose members are not in a c-command relation in fact exist. The question then shifts to



differentiating ellipsis chains from others in terms of c-command, --presumably a problem in the interpretive component and not in the assembly system.

This raises a different though related issue of whether all relations involving c-command reduce to the chain relation or whether the interpretive component will need to refer additionally to c-command. I will not attempt here to go beyond noting the issue. The assumption that chain-members are not forced to be in a c-command relation by syntax assumes that the interpretive system has access to this relation. On the other hand the obvious counter-example of anaphora to the more restrictive hypothesis, according to which all c-command relations reduce to chain-relations can be fairly straightforwardly accommodated. Thus ordinary anaphors have been treated as involving a chain/move relation, while bound (variable) anaphora has been argued to require precedence rather than c-command (Williams 1994, Brody 1994, see also Hornstein 1994 for a contrary view and some discussion below).

### 3. The Minimal Link Condition (MLC) and covert A'-chains

The MLC in Chomsky 1995 restricts elements moving to a given target K to the closest one among those that have the property that they can enter into a checking relation with K. This version of the MLC covers cases like superraising (11b), superiority (12b) and wh-island violations (13b).

- (11) a. \*\*It seems t(it) is certain John to meet Mary  
b. \*\*John seems it is certain t(John) to meet Mary
- (12) a. Who t(who) saw what  
b. ?\*What who saw t(what)
- (13) a. \*Who did you wonder [<sub>CP</sub> t(who) [<sub>IP</sub> t(who) gave this book to whom]]  
b. ?To whom did you wonder who gave this book t(to whom)

The MLC does not allow these derivations since in each case there is a nearer element to the target of movement that has appropriate checking features. In (11b) this is the expletive subject *it*, in (12b) and (13b) the wh-phrase *who* in subject and spec-CP position respectively. The MLC would therefore allow the derivations in the a. examples in (11) through (13). Such a derivation happens to give a grammatical result in (12), but crashes in (11) and (13). Chomsky points out that under such a formulation of the MLC it could not be an economy condition: if crashing derivations could block a converging one then presumably no operation would ever take place. So he takes the MLC to be part of the definition of Move.

The restrictive assembly system of perfect syntax aims to avoid both economy conditions and stipulated conditions on chains/movement like the MLC. There are also empirical reasons for questioning the generalization the MLC expresses. Superraising, Superiority, Wh-island violations appear to very have different (un)grammaticality status, raising initial doubts about a principle that treats them uniformly. Judgements range from the completely hopeless superraising case through the intermediate superiority effects to the only mildly deviant and sometimes fully grammatical wh-island violations. As is well known, the wh-island cases improve probably to full grammaticality with infinitival complements as for example in (14). Since the MLC cannot be made sensitive to the presence of tense in any obvious way, it will rule out also such cases.

- (14) Which crimes does the FBI know how to solve



Another problem with the MLC is that it is not compatible with covert A'-movement: covert A'-movement of wh-in-situ and similar elements would regularly violate the MLC, as exemplified in (15) where the wh-in-situ should be understood with matrix scope.

(15) Who wondered who t bought what

Chomsky 1994, 1995 assumes that A'-movement can take place only in the overt derivation, but it is not clear why A- and A'-movement/chains should differ in this way. Furthermore the assumption makes some of the standard evidence for covert A'-movement/chains puzzling. In-situ neg-phrases are sensitive to complement non-complement divide as was first observed by Kayne 1981 (see also Jaeggli 1981, Rizzi 1982, Longobardi 1991 for the same effects in other Romance languages). This is parallel to overt A'- chains as exemplified in (16) and (17)

- (16) a. Who did you say that Mary saw  
 b. \*Who did you say that saw Mary
- (17) a. Jean n'exige que Pierre voit personne  
 J not require that P has seen no one  
 b \*Jean n'exige que Pierre soit arrete  
 J not require that P be arrested

Or as argued first by Longobardi, the relation between the in situ neg-phrase and its scope (marked by non) shows subjacency effects:

(18) Complex NP Constraint

- \*Non approvarei la tua proposta di vedere nessuno  
 'I would not approve of your proposal of seeing anybody'

(19) Sentential Subject Condition:

- ?\*Chiamare nessuno sara possibile  
 'To call no one will be possible'

(20) Adjunct Condition:

- a. \*Non fa il suo dovere per aiutare nessuno  
 'He does not do his duty in order to help anyone'  
 b. \*Per ottenere nulla ha fatto il suo dovere  
 'In order to obtain nothing has he done his duty'

As observed by Watanabe (1991) wh-in-situ in Japanese type languages also has properties that parallel those of overt movement/chains. (21) is an example, where the wh-in-situ within the wh-island creates a degraded grammaticality status:

- (21) ??John-wa [Mary-ga nani-o katta [ka dooka]] Tom-ni  
 top nom what-acc bought whether dat  
 tazuneta no  
 asked Q  
 'What did John ask Tom whether Mary bought t?'

Hornstein (1994), who assumes the MLC and rejects covert A'-movement attempts to neutralize Longobardi's examples by assuming that they involve A-movement. This does not



appear to lead to an easily tenable position however for two reasons. First as Longobardi made clear, the relation has typical A'-chain locality properties, in particular it can escape from embedded tensed clauses:

- (22) Non approverei che tu gli consentissi di vedere nessuno  
'I would not approve that you allow him to see anybody'

Secondly, as (21) shows, covert wh-relations show similar effects. But wh-chains are prototypical A'-relations.

A more plausible theory that would neutralize the evidence of the examples in (18) through (21) would claim that all syntactic A'-movement is subject to the MLC and that syntactic A'-movement occurs only to satisfy some checking feature of a +wh/+Neg head. This point is neutral with respect to whether the movement of the in situ neg-phrase in (18) - (20) and the wh-phrase in (21) involves overt movement (of, say, an empty operator as in Watanabe 1991) or covertly (say, as movement of formal features, as in Chomsky 1995). The crucial distinction would be between in situ elements that need to move to satisfy some checking feature and in situ elements that do not have similar motivation to move. The two groups correspond to the primary and the secondary or parasitic wh/neg-chains respectively of Brody 1995a. The wh-in-situ not constrained by the MLC in (15) would belong to the group of secondary relations, while the in situ elements exhibiting movement characteristics in (18) through (21) would be primary ones. Could it be claimed then that only primary relations involve syntactic movement/chains, and hence only these show movement/chain characteristics including the MLC?

Such a claim would be incompatible with the checking theory to be discussed in section 4 below, in which no distinction can be made between primary and secondary chains with respect to their behavior in checking relations. There is also direct empirical evidence that appears to show that the claim that only primary relations involve syntactic movement/chains would not be correct: secondary relations also show island effects. Neg-phrases participating in secondary relations show parallel locality behaviour to parasitic gaps. As Kayne 1983 showed, although parasitic gaps can be separated from the primary gap by an island, within this island they show movement diagnostics with respect to further islands. As pointed out by Longobardi 1991, secondary neg-phrases appear to do likewise. This is exemplified in (23)-(24).

(23) Parasitic neg-phrases:

- a. (In a Sentential Subject)  
(?) Chiamare nessuno [secondary neg-phrase] servirà a niente, ormai  
'To call nobody will do any good now'

- b. (In an Adjunct)  
Non fa niente per aiutare nessuno [secondary neg-phrase]  
'He does not do anything in order to help anyone'

- (24) a. Adjunct island inside Subject island:  
\*Partire per incontrare nessuno [secondary neg-phrase] servirà a niente  
'To leave in order to meet no one will do any good'
- b. Adjunct island inside another:  
\*Non fa niente per scoprire la verità indagando su



nessuno [secondary neg-phrase]  
'He doesn't do anything in order to discover the truth  
by investigating anyone'

In Brody 1995a I argued for extending the analysis in terms of parasitic chains to wh-relations. Chomsky 1986 drew the conclusion about parasitic gaps on the basis of such locality evidence that they must involve movement. If we draw the same conclusion here about secondary neg-and wh-relations, then the relation of the in-situ wh-phrase in (15) to its scope position must involve a chain/Move relation. The analysis creates difficulties then for the MLC, since in examples like (15) with matrix scope for the wh-in-situ, this condition will be violated.

Various further problems for the MLC arise from properties of superiority effects. These can be alleviated in several types of constructions including (25), where the wh-in-situ is associated with the matrix wh-phrase (cf. Lasnik and Saito 1992) and (26) where the wh-in-situ is d-linked (cf. Pesetsky 1987)

- (25) Who wondered what who bought  
(26) Which book did which girl buy

From the perspective of the MLC, such cases are difficult to understand and therefore they raise doubts about this condition. The alternative quasi-semantic treatment of superiority based on work by Chierchia (1991), put forward independently by Williams (1994) and Hornstein (1994), on the other hand is able to make sense of such apparent counterexamples.

The analysis, which relates the phenomenon to weak crossover, claims that a wh-in-situ is or contains an element whose interpretation is dependent on the A-position copy of the wh-phrase in spec-CP. Superiority violations arise when this dependency is illegitimate. For present purposes it does not matter if this is because the antecedent of the dependent element D must precede D (Williams 1994, Brody 1994) or because it must "almost c-command" (Hornstein 1994) D. Such an analysis captures the three way parallel between the crossover and the superiority violations in (28) and (29) and the lack of pair-list interpretation in (30) (cf. especially Hornstein 1994 for details and much additional argument):

- (28) What did who buy (what)  
(29) Who did his father meet (who)  
(30) What did everyone say (what)

In none of the three cases does the trace of the wh-phrase in spec-CP c-command or precede the dependent element (who, his and everyone respectively). Furthermore in (25) the trace of the matrix wh-phrase does c-command and precede the wh-in-situ, thus the latter element can be associated with the matrix wh-phrase without incurring a superiority violation. (26) also becomes understandable on the assumption that a d-linked wh-phrase need not be a dependent element. (Notice that we cannot meaningfully make a similar exemption from the MLC for d-linked wh-phrases by allowing these to be crossed. The MLC is a formal no-crossing requirement not directly sensitive to the interpretive status of the elements involved.)

To summarize so far, the island behavior of neither overt nor covert A'-movement conforms to that predicted by the MLC. Furthermore the condition incorrectly predicts superiority violations in various cases and is incapable of capturing the similar behavior of weak crossover, pair-list reading and superiority constructions. Discounting superiority and



wh-islands the major remaining effect of the MLC is the exclusion of superraising. Here again there are alternatives. One possibility might be to restrict theta role percolation in the spirit of Williams 1994. Brody 1995a,b argues for an interpretive mechanism of feature percolation in chains, that involves thematic roles alongside non-syntactic selectional features. Suppose that NP-traces can receive but cannot directly percolate (transmit) theta roles to higher chain members: they can do this only via the subject-predicate relation under a stricter locality condition. Let's assume for concreteness that the theta role can be inherited from the NP-trace vertically by the smallest predicate VP that includes the trace, and which can in turn assign it to its subject via the predication relation. Vertical transmission is possible only from from (Caseless) NP-traces. This may be a subject trace as in raising or an object trace, as in passives and ergatives. This will allow successive cyclic chains but no superraising in English.

In (11b) for example the theta role of the most deeply embedded verb, meet will be assigned to the VP-internal NP-trace from which it will percolate to the predicate VP t meet Mary. This assigns the theta role to the subject, again a Caseless trace that allows further percolation to the next predicate. The subject of this predicate is the expletive it however, so this receives the theta role. No further percolation can take place and the matrix subject John will remain without a theta role. The account predicts also that superraising will be possible in multiple subject languages (Ura 1994) on the assumption that in these a predicate can sometimes enter multiple predications relations.

The MLC thus largely dissolves, as perfect syntax leads us to expect it would: most of the phenomena in its scope turns out to involve interpretive relations that should be constrained within the interpretive component. Dependency, in the sense used here in the account of superiority and related phenomena is clearly a semantic relation, and so is the mechanism of theta role percolation. This is just a particular case of percolation of selectional features, only a subset of which are syntactic, --cf. Brody 1995a,b.

Numerous questions remain in the area of locality phenomena, which a fuller treatment would need to address. Let us look at one case that might appear related to the MLC: the well-known wh-island effect that shows up even in constructions like (14). This is the so called scope reconstruction, --in fact a dependent reading of the wh-phrase on a quantifier c-commanding its trace. This reading does not obtain even in otherwise grammatical wh-islands (on "scope reconstruction" cf. e.g. Longobardi 1984, Rizzi 1990, Williams 1994, Hornstein 1994). Thus (31a) can be construed as a question asking about each individual which book that individual read, where they all potentially read different books. Such a construal is possible also where the quantifier is in an embedded clause as in (31b). (31c) on the other hand only has a reading "which book is such that you wondered whether everyone read that book", ie. on which everyone read the same book. Similar judgement obtains with the infinitival embedded clause in (31d).

- (31) a. Which book did everyone read  
b. Which book did you believe everyone to have read  
c. Which book did you wonder whether everyone read  
d. Which book did you wonder whether to believe everyone to have read

Hornstein proposes that the MLC is responsible for this difference: it always prevents extraction from the wh-island. (14) will then involve an island internal empty operator movement (cf. Cinque 1990) construed with the matrix wh-phrase. Dependent reading of the wh-phrase will be impossible since the MLC effectively prevents "reconstruction" of the wh-



phrase into the island: the trace in the island will be the trace/copy of the empty operator and not that of the *wh*-phrase. This account raises numerous questions. First of all, if our discussion so far is correct, the standard minimalist version of the MLC has little independent motivation, hence invoking it here would be quite stipulative. Other problems have to do with the empty operator: what is its landing site, what checking relation triggers its movement (there are no plausible candidates), etc? Yet others have to do with the scope of the phenomena. Since the effect shows up in other types of islands, Hornstein suggests assimilating all relevant island violations to *wh*-islands. If all islands have a filled A'-spec then crossing these will be ruled out by the MLC (note 26, p.181). This would make it necessary to postulate a filled A'-spec (in addition to postulating a head that is necessary to check the empty operator) also in complex NP constructions and adjuncts where the same effect obtains. The marginal structures in (32) cannot have the "scope reconstruction" reading either.

- (32) a. ??Which book did you deny the fact that everyone read  
 b. ??Which book did you meet Mary before everyone read

But there is no evidence for any of the three empty elements that this account needs to postulate (empty blocking A'-spec, empty Operator, empty head to check the empty Operator). Thus given minimalist assumptions, it is quite unlikely that the MLC is responsible for "scope reconstructions" effects. (This is not to say that it is inconceivable that some notion of crossing is at issue in the examples in (31), although (32) raises some doubts even about this much weaker claim. In the context of Rizzi's (1990) relativized minimality it was reasonable to put examples like (32) aside as due to some other notion of barrier, since relativized minimality captured a rich generalization in that it constrained also the behavior of adjuncts. But this is not true of the MLC. One important respect in which the MLC differs from relativized minimality is that it is taken to refer to argument-type elements. Thus if it was taken to constrain also adjuncts, it would predict the same type of violation: correctly or not, it cannot capture the adjunct argument asymmetry with respect to extraction from islands.)

#### 4. Checking theory

##### 4.1. Bare checking theory

Let us start with subject verb agreement. The checking theory of Chomsky 1993 assumes that in addition to agreement features appearing on the verb and the subject, mediating features occur on the agreement node. In "John hits Bill" ...The  $\phi$ -features appear in three positions in the course of the derivation: internal to John, internal to hits, and in  $AGR_S$ ". In effect the mediating features are present in duplicate: "AGR must in fact have two kinds of features: V-features that check V adjoined to AGR and NP-features that check NP in SPEC-AGR." The mediating features delete when checked, so "at PF and LF the  $\phi$ -features appear only twice, not three times: in the Noun Phrase and verb that agree." (p.42)

In Brody 1995c I argued against such mediating features. The major objection against NP- and V-features was that at LF the  $\phi$ -features of subject/object-verb agreement should only occur on the subject, they do not appear to have a consistent additional interpretive function on the verb. Additional copies of checking features would presumably be excluded by the principle FI. But under a checking theory where Agr has separate V- and NP-features,



there will still be two copies of the relevant features at LF after the mediating features have deleted. One of these will be on the subject and the other on the verbal head.

Such considerations led to a theory without splitting of Agr features: The approach, which I called "bare checking theory", takes seriously one of the original intuitions behind checking theory. It claims that "movement" or rather non-trivial chain formation is forced by bare output conditions at the LF interface. Chains are formed because of the way information is stored in the lexicon. The lexical items in a sentence duplicate certain features --the checking features-- these duplications must disappear by LF through checking. Checking is thus forced by the fact that the features involved are interpretively redundant and would violate the principle of full interpretation. (We may assume that checking involves marking the redundant features invisible, perhaps as a result of merger of features in some sense. Merger might make both checking features invisible creating a single visible LF unit.)

Given bare checking theory, checking must take place because multiple instances of what is in fact one feature are not tolerated at the interface. It is not necessary then to invoke non-interpretability of features to force a checking configuration. Bare checking theory is thus consistent with radical interpretability according to which all features have semantic content.

Dispensing with NP- and V-features leaves the further question of whether mediating features (now without the split into NP- and V- features) exist. The basic assumption of bare checking theory does not require the elimination of the mediating features in Agr, although it would not disallow this move either. This theory is not compatible with the splitting of the features of Agr into NP- and V-features since this would result in two copies of  $\phi$ -features at LF. But if  $\phi$ -features are present in three position, then bare checking theory will simply require all three positions to merge by LF through checking.

Nevertheless I will assume that mediating features do not exist. Apart from simplicity considerations, an argument against these in Brody 1995c was that the evidence for the third copy of  $\phi$ -features is missing even where it could be expected to occur. When the feature on Agr is weak we would expect it to show up at PF. It is only by stipulation (namely that unchecked weak features are invisible or delete in the PF component) that this incorrect prediction is neutralized. I assumed therefore that within bare checking theory subject verb agreement involves only a single checking relation: the  $\phi$ -features of the verb and the subject are checked directly against each other. (On the assumption that checking must involve spec-head configurations in functional phrases, this will entail the formation of NP- and V-chains.)

Chomsky 1995 also develops a theory that is not compatible with the existence of mediating features. He proposes to eliminate the Agr node altogether. Although he does not discuss the matter, eliminating Agr could give the result we seek: after subject verb checking there will be only a single set of  $\phi$ -features present at LF. Chomsky rejects the Agr node for reasons that are not dissimilar to those given in Brody 1995c for rejecting mediating features. Apart from simplicity (section 4.1 pp16-17), Chomsky notes that unlike other functional categories like C, T or D for example, Agr does not contain interpretable (either LF or PF) features (section 10). "We therefore have fairly direct evidence from interface relations about T, C and D but not AGR" (p.8).

Notice though, two differences. First the argument from PF invisibility does not distinguish between (semantically/conceptually) interpretable and non-interpretable features: it argues against triplications irrespective of this dimension. Chomsky's argument on the other hand is relevant only for features that have no interpretation at either interface.

Secondly, my argument questioned only the existence of mediating features while Chomsky argues against the Agr node itself. But the question of whether the Agr node exists is in fact composed of two issues: only one of which is the question of the mediating features.



Even if these do not exist, it might be that some node  $\alpha$  above T projects a phrase which hosts both a member of the subject chain (in its spec) and a member of the verbal head chain that enter into a checking relation here. One possibility is that T or perhaps all categories have the option of projecting recursively:  $\alpha$  would then correspond to the higher T node. Thus we could retain the spec and the head positions as appears to be empirically desirable, without assuming the existence of mediating  $\phi$ -features. The condition of interpretability (cf. section 1.2 above) could be satisfied jointly by the the two T nodes. (Similarly, recursively projecting V<sub>s</sub> may be a way of creating multi-layered VPs.)

Consider next English interrogatives.

- (33) I wonder (what) who +WH [(who) saw what]  
 (34) a. I wonder who +WH Bill saw (who)  
       b. Bill saw who  
 (35) Who did +WH Mary (did) see (who)  
 (36) Did +WH John see Mary

In (33) there are three instances of the wh-feature (two on the wh-phrases and one on the embedded C node). But there is only one question. So by full interpretation the wh-features must all merge. This is possible if both wh-phrases form a chain that has a member that forms a checking relation with the +wh head. Thus bare checking theory entails the existence of A'-chains that express "covert A'-movement" relations, in accordance with our earlier conclusions. As noted in section 1, a chain corresponding to an "overt movement" relation is the default case, it will occur if the relevant head (here the +wh C) is "strong" enough to license a head-external element in the checking domain in the morphological component; --in the case of an XP-chain word-externally as a specifier. (On the reasons for "extended structure preservation" that prevents word-internal phrases see Brody 1995b.) Given the generalization that SPELLOUT operates only on the highest copy in a chain, the lexical item will show up in spec-CP. In English this head licenses a single spec, hence additional wh-phrases checked by it must remain in situ forming feature-chains only.

(34b) will be possible only with a -wh C and an echo interpretation. If it had a +wh C then the two wh-features in the structure should merge through checking by full interpretation. This makes a chain necessary, linking the wh-phrase to the checking domain of the +wh C. Since this head licenses an overt specifier in morphology, the chain must be a full copy chain as in (34a).

In (35) the auxiliary did must have a wh-feature. We can analyze (35) in two ways. The choice between these depends on whether interpretable features on heads mediating spec-head relations between other elements can exist. If they do, then both C and T can have a wh-feature and T (together with the auxiliary) forms a chain because these must merge through checking. Additionally the wh-feature on who must also merge with this complex, hence an XP-chain is also formed. These will be chains involving copies of full categories since the relevant C licenses both a word-internal and a word-external element in its checking domain. Alternatively if mediating features are dispensed with, then the T with the wh-feature and the wh-phrase both form a chain linking them to CP because this is the the only way that they can establish a checking relation. (This latter alternative corresponds essentially to the account in Rizzi 1991, see also Brody 1990, 1995c.)

If yes-no questions contain an empty (wh-)operator in spec-CP then the analysis of (36) will not significantly differ from that of (35). If not, then (36) will be parallel to (34),



modulo the difference between an  $X^0$ -chain and an XP-chain.

Notice that we must apparently allow merger of a set of checking features without all of these occupying positions in the same checking domain. In (36) there are three instances of the +plural feature, but only one plurality: that of the DP. The +plural feature XP-chain of they apparently merges with the other two in two distinct spec's. Given the independently necessary assumption that checking features are properties of the chains rather of the members of the chain (cf. Brody 1995b), this should create no problems. At LF there is only a single +plural feature in (37), that of the chain.

(37) They(pl) seem(pl) clever(pl)

#### 4.2 The minimalist checking theory and the +/-Interpretable distinction

Let us next consider briefly the current version of checking theory in the minimalist framework (Chomsky 1995). He suggests that interpretable features like categorial features and  $\phi$ -features on nouns do not need to be checked, whereas noninterpretable features like for example Case or  $\phi$ -features on verbal heads or "strength" (presumably a feature) of an interpretable feature must be checked because this makes it possible to eliminate these, as required by full interpretation. In fact he assumes further that once checked, these features are not visible for the remaining syntactic computation either.) Thus checking relations and indirectly movement and chain relations are forced by noninterpretable features. These must be eliminated, and they can be eliminated only when already checked.

The principle of radical interpretability is incompatible however with existence of noninterpretable features: according to this principle all syntactic features must be potentially interpretable. As we have seen, under bare checking theory it is not necessary to make use of noninterpretable features: checking relations are forced by the duplication of interpretable features. Furthermore the checking theory based on this distinction seems to lead to less optimal analyses. Consider for example the analysis of interrogatives within this theory.

The wh-feature is clearly interpretable hence not in need of being checked. A +wh feature will be checked only if it is "strong" and then overtly (some of the problems with the notion of strength used here were discussed in section 2 above). Thus in English the wh-feature on C is strong and hence it can be checked either by (T+)did as in (36) or by a wh-phrase as in (34). Since a strong feature can be satisfied by a single element, the analysis raises the question of why (38a) is unacceptable. Here the strong wh-feature of C is satisfied by the hosted verbal element.

- (38) a. did John give which book to Mary  
b. +WH John gave which book to Mary

Chomsky suggests that (38a) "converges as gibberish". (Notice that this contrasts with (38b), which crashes since the strong wh-feature on C has not been checked.) But since Chomsky rejects covert A'-movement type syntactic relations, he needs to assume the existence of some interpretive mechanism to link in situ wh-phrases to their scope. It is then unclear why (38a) is gibberish, why it cannot be interpreted with the interpretive strategy generally used for in-situ wh-phrases. As we have seen, under bare checking theory, where all wh-features must merge the problem does not arise: which book in (38a) must form a chain linking it to the wh-feature of the auxiliary (and perhaps also of C). Further, the chain must be a full category



copy chain, that is one that corresponds to overt movement of the minimalist framework, since English C is strong, ie. it licenses a specifier in addition to an element in the word ( $X^0$ )-internal checking domain.

(Further elaboration, like for example multiple strong features are necessary to generate (35) within Chomsky's system of assumptions. Note that (35) like many other examples will violate also the MLC of the minimalist system: the wh-phrase moves to CP even though another element that could (and does) legitimately move there, namely the verbal complex, is nearer.)

The remaining prediction in the theory where noninterpretable features are crucially involved is the exclusion of movement from Case positions. On the assumption that Case is noninterpretable and that such features once checked are invisible for further computation, the derivation of (39) will crash. The Case feature of the DP John are checked and deleted in the embedded clause (together with the Case feature of the embedded T. Hence the noninterpretable Case feature on the matrix T cannot be checked and deleted when DP raises, and therefore the derivation will crash at LF.

- (39) a. \*John seems [t saw Mary]  
b. \*There seems [a man was clever]

If this is the only case where noninterpretable features are needed then the explanation is less appealing. Additionally easy-to-please and grammatical superraising suggest that Case is not relevant here:

- (40) John is easy Op to please t  
(41) John seems [t' Mary liked t]

If the analysis in Brody 1993b is correct then in (40) we have a chain [John, Op, t] that involves two Case positions. Similarly in the grammatical superraising cases like (41) the superraising chain [John, t', t] appears to involve two Case positions (Ura 1994). Notice that if John in (41) is taken not to check (accusative) Case in the position of t, then John in (39a) (and "a man" in (39b)) should similarly be able to avoid Case-checking in the lower clause. This would however eliminate the explanation of the ungrammaticality of (39): these DPs could check the Case feature of the matrix T and the derivation would converge. (Ura suggests that in the grammatical superraising constructions lack of Case assignment to the trace correlates with the possibility of filling the position with a pro element. This generalization would still incorrectly allow (39) in subject pro drop languages like Italian or Hungarian.)

I will not attempt an alternative explanation of (39) here beyond noting that the contrast between (39) and (40)/(41) suggest that the subject non-subject difference may be relevant.

## 5. Conclusion, Uniformity, Last Resort, Procrastinate

Although various issues remain, the results so far seem encouraging. Stipulative conditions of the minimalist framework on the Move/chain relation, like the cycle/c-command, and the MLC appear unnecessary or attributable to the interpretive components as perfect syntax leads us to expect. Checking relations and indirectly (non-trivial) chain formation is forced by bare



output conditions. Given bare checking theory this need not involve noninterpretable features that would violate radical interpretability. I have discussed in detail and rejected the remaining major stipulative condition on Move/chain, namely Uniformity, in Brody 1995b. I argued there that a well-designed theory should not make available devices that would make it possible to violate the condition, which requires that all chain members be of the same projection level. If projectional levels are not defined relationally, then the question of uniformity cannot even arise. Since chains are sets of copies the chain members necessarily share also the projection level property.

Let me comment finally briefly and incompletely on two more conditions: Last resort and Procrastinate. Last resort can be thought of as an interface condition on the assumption that all categories must be licensed by full interpretation. This licensing involves participation in either (a) a projectional relation (selectional/ modificational relation or categorial projection) or (b) in a checking relation. Since by the generalized projection principle projectional relations are relevant only for chain roots (cf. Brody 1995b), non-roots of non-trivial chains (i.e. "Move" of the minimalist framework) can be licensed only by a checking relation.

I assumed that in the default case chains are formed on full categorial copies, feature chains occur only when full copies are not licensed in the checking domain of some head. This preference for "overt movement" type relations is consistent with the Transparency principle of Brody 1995a and incompatible with Procrastinate which would require feature chains as the default case. There were essentially two main reasons for rejecting Procrastinate. First it has difficulties with a configuration where there is multiple overt movement of a given type of element to some head. For example in a language where all wh-phrases must "overtly move", a strong +wh head appears to be accidentally satisfied by exactly as many wh-phrases as there are in the sentence. This seems to be a serious problem that is avoided by a preference for "overt movement", that is for chains that involve full copies of categories. Secondly Procrastinate (in present terms preference for feature-movement) predicts that in the default case LF and PF will be maximally different, and wherever they are similar, this is only due to morphological accidents (presence of a strong feature). This again appears to be a rather strange consequence that a preference for full copies avoids: now categories at LF and at PF will have similar distribution in the default case, it is when they do not, that we have the morphological accident. As we have seen this arises due to the limited expressive power of this component: a head may not always be able to check elements external to it.

There is a particular prediction of Procrastinate that our account so far has nothing to say about. The ungrammaticality of the examples in (42) has been attributed to this principle. Procrastinate will predict this, if at the embedded subject position it prefers insertion (of the expletive *it*, *there*), over the raising the associate (*a man*). Instead of the ungrammatical (42), we will then derive the grammatical (43).

- (42) a. \*There seems a man to have arrived  
b. \*It was believed a man to have been here
- (43) a. There seems (there)to have arrived a man  
b. It was believed (it) to have been a man here

Let us consider the alternative of claiming that the embedded subject position simply does not exist. If so it cannot be occupied by the associate and (42) cannot arise. This entails immediately that English object shift must be overt: *him* in (44) cannot be in the embedded



subject position. (For independent arguments for the overtness of English object shift see eg. Johnson 1991, Koizumi 1993, Lasnik 1994.)

(44) John believed him to be clever.

The claim that the governed and Caseless subject positions of Government Binding theory do not exist, appears to entail also giving up the successive cyclicity of A-chains since intermediate A-chain links occupy such subject positions. But it is possible to construct direct evidence for A-chains being successive cyclic, --the intermediate traces in (45a) and (46a) act as interveners for the binding relations indicated. (The principle C violation in the b. examples shows that c-command holds between the relevant elements.)

(45)a. John seems to Mary<sub>x</sub> [ (John) to appear to her<sub>x</sub>//\*herself [ (John) to have met me ]]

b. \*John seems to her<sub>x</sub> [ (John) to appear to Mary<sub>x</sub> [ (John) to have met me ]]

(46)a. John seems to Mary<sub>x</sub> [(John) to be considered [(John)clever by her<sub>x</sub>//herself]]

b. \*John seems to her<sub>x</sub> [ (John) to be considered [ (John) clever by Mary<sub>x</sub>]]

If the full subject positions are not licensed, then the discussion so far leads to the hypothesis that intermediate traces in A-chains involve only feature copies. The feature copies on the checking head (presumably T) can act as interveners for the binding theory. If this idea can be elaborated into a tenable analysis, it would represent a case where a full copy is higher than a feature copy. This would be an impossible scenario in the minimalist theory: a "covert movement" type relation cannot be followed by an "overt movement" relation. (This is not to say of course that the configuration is not achievable through stipulation, say by deleting the intermediate trace but leaving its relevant features on the checking head.)

On the other hand given the assembly system of Brody 1995a,b, such a configuration would be expected to occur. Here copies (of features or of full categories and their content) can be freely made. (Full category copies are preferred as discussed in section 1 above.) Unless further constraints prevent this, feature copies and full copies can be inserted in the structure in any c-command order. In particular the original may be higher than several feature copies (which in turn may be higher than another full copy) as appears to be required for the analysis of successive cyclic A-movement, given the evidence in (45) and (46).



## References

- Brody, Michael. 1990. Some remarks on the focus field in Hungarian. In UCLWP vol.2. University College London.
- Brody, Michael. 1993a. Lexico-Logical Form. Ms., University College London
- Brody, Michael. 1993b. Theta theory and arguments. Linguistic Inquiry 24:1-23.
- Brody, Michael. 1994. Dependence and phrase structure. UCLWP vol.6. University College London
- Brody, Michael. 1995a. Lexico-Logical Form. A Radically Minimalist Theory. Cambridge, Mass.:MIT Press.
- Brody, Michael. 1995b. Phrase Structure and projection. Ms., University College London
- Brody, Michael. 1995c. Hungarian focus and bare checking theory. In Arbeitspapiere des Sonderforschungsbereichs 340, University of Tubingen
- Cinque, Guglielmo. 1991. Types of A'-dependencies. Cambridge, Mass.: MIT Press.
- Cinque, Guglielmo. 1995. Adverbs and the universal hierarchy of functional projections. Tromso GLOW abstract
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. MIT Occasional Papers in Linguistics, MIT, Cambridge, Mass.
- Chomsky, Noam. 1994. Bare Phrase Structure. ms., MIT, Cambridge, Mass.
- Chomsky, Noam. 1995. "Chapter 4". ms., MIT
- Chierchia, G. 1991. Functional wh and weak crossover. In The proceedings of the WCCFL 10, 75-90. Stanford Linguistics Association, Stanford.
- Hornstein, Norbert. 1994. LF: The grammar of logical form. From GB to minimalism. Ms., University of Maryland.
- Jaeggli, Osvaldo. 1981. Topics in Romance syntax. Foris, Dordrecht.
- Johnson, Kyle. 1991. Object positions. Natural Language and Linguistic Theory 9:577-636.
- Kawashima, R and Hisatsugu Kitahara. 1995. Strict cyclicity, linear ordering and derivational c-command. to appear in WCCFL 14.
- Kayne, Richard. 1981. ECP extensions. Linguistic Inquiry 12:93-133.
- Kayne, Richard. 1983. Connectedness. Linguistic Inquiry 14:223-249.
- Kayne, Richard. 1995. The Antisymmetry of Syntax. Cambridge, Mass.:MIT Press
- Koizumi, Masatoshi. 1993. Object Agreement Phrases and the Split VP Hypothesis In MITWPL vol.18. MIT, Cambridge, Mass.
- Lasnik, Howard. 1994. Antecedent contained deletion and/or pseudo-gapping. Talk, presented at MIT
- Lasnik, Howard and Mamoru Saito. 1992. Move  $\alpha$ . Cambridge, Mass.: MIT Press.
- Longobardi, Giuseppe. 1991. In defense of the correspondence hypothesis: island effects and parasitic constructions in logical form. In: Logical structure and linguistic structure. eds. C.-T. James Huang and Robert May, Dordrecht: Kluwer.
- Pesetsky, David. 1987. WH-in-situ: movement and unselective



- binding. In: The representation of (in)definiteness. 98-129.  
Eric reuland and A. ter Meulen eds., Cambridge, Mass.: MIT  
Press.
- Pesetsky, David. 1989. Language particular processes and the  
earliness principle. Ms., MIT.
- Rizzi, Luigi 1982. Issues in Italian syntax. Foris, Dordrecht.
- Rizzi, Luigi. 1990. Relativized Minimality. Cambridge, Mass.: MIT Press.
- Rizzi, Luigi. 1991. Residual verb second and the Wh-Criterion.  
Ms., Universite de Geneve.
- Ura, Hiroyuki. 1994. Varieties of raising and the feature-based phrase structure  
theory (chapter 1). Ms., MIT.
- Watanabe, Akira. 1991. Wh-in-situ, Subjacency and chain formation. Ms., MIT.
- Williams, Edwin. 1994. Thematic Structure in Syntax. Cambridge,  
Mass.: MIT Press.





**PREVIOUS TITLES IN THIS SERIES:**

- 1/1: M. BRODY, *Phrase Structure and Dependence*.
- 1/2: K. É. KISS, *Generic and Existential Bare Plurals and the Classification of Predicates*.
- 1/3: L. KÁLMÁN, *Quantification, Conditionals and Bipartite Meanings*.
- 1/4: Z. BÁNRÉTI, *Modularity in Sentence Parsing: Grammaticality Judgments in Broca's Aphasics*.
- 2/1: A. SZABOLCSI, *Strategies for Scope Taking*.
- 2/2: G. RÁDAI and L. KÁLMÁN, *Compositional Interpretation of Computer Command Languages*.
- 2/3: L. KÁLMÁN, *Strong Compositionality*.



