THEORETICAL LINGUISTICS PROGRAMME, BUDAPEST UNIVERSITY (ELTE)

PHRASE STRUCTURE AND DEPENDENCE

MICHAEL BRODY

RESEARCH INSTITUTE FOR LINGUISTICS, HUNGARIAN ACADEMY OF SCIENCES

WORKING PAPERS IN THE THEORY OF GRAMMAR, VOL. 1, NO. 1

RECEIVED: SEPTEMBER 1994

PHRASE STRUCTURE AND DEPENDENCE

MICHAEL BRODY

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

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

Syntactic structures have been analysed in terms of constituent hierarchies and also in terms of dependency relations. While these two traditions have sometimes been presented as competing with each other, there is no reason in principle why syntactic dependencies and constituent hierarchies should not both be part of the grammar, complementing each other. Indeed, within transformational grammar and more specifically within the Principles and Parameters theory, various dependencies are postulated alongside the basic constituent hierarchy analysis of sentences. The most salient of these are the various types of anaphoric dependencies, but other relations, like for example thematic, government or chain relations, have also from time to time been thought of as constituting dependencies. Within this framework it remains true however, that the basic structural organizing principles are couched in terms of constituent hierarchies, dependency relations are not taken to play a role here. In what follows I shall propose that dependency relations are more central to the grammar. I shall postulate a new module: dependency theory. The basic syntactic analysis of a sentence will not simply consist of a set of elements arranged in a constituent hierarchy. It will include also a dependency structure defined on these elements: all syntactic elements participate in both a hierarchical and a dependency structure.

In section I of this paper I shall show that dependency theory makes it possible to radically simplify the theory of phrase structure. In section II, I shall argue that the evidence that has standardly been taken to motivate binary rightward branching analyses is better treated in terms of dependency relations between constituents.

I. Phrase Structure and Dependency

I.1. Structural Dependencies

Let me start with sketching the dependency structures I have in mind. I shall propose that the concepts of specifier-head and head-complement relations on one hand and dependency relations on the other should be brought together: at least some specifier-head and head-complement relations are in fact dependencies. Suppose that phrasal categories can be in a structural dependency relation with some head X. It will follow from the principles of the theory (to be presented immediately below) that there is a single category on which the head X directly depends, the **spec** of X, and a single category which directly depends on the head, the **comp** of X. (Notice that spec and comp are not defined hierarchically here, but in terms of dependencies.) I shall call the spec-head and the head-comp dependencies **direct structural** dependencies. Direct structural dependencies are then relations between a head and a phrase.

Spec-head dependencies can be taken to be instantiated by checking

relations, --taking the head or its checking features to depend on the spec or on the relevant features of the spec. A typical comp-head dependency is object theta assignment/selection.

Turning now to the principles of dependency theory I assume first that direct structural dependencies interact transitively creating indirect structural dependencies. I assume also an adapted version of Kayne's (1993) Linear Correspondence Axiom. I shall take the principles of Precedence and Totality to regulate the relationship between the linear ordering of terminals and the **dependency** relations holding between nodes dominating them. I give a somewhat informal statement in (1) and (2):

(1) **Precedence**
if x structurally depends on y then the terminals dominated by y precede the terminals dominated by x

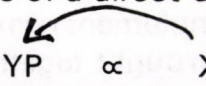
(2) **Totality**
all terminals must be ordered by the Precedence condition on structural dependencies

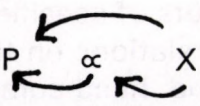
Williams (1992) has independently argued for a precedence condition on anaphoric dependencies. Precedence in (1) can also be thought of as an application of his condition to structural dependencies.

In addition I assume that no redundant direct structural (spec-head or head-comp) dependencies are allowed: there can be no direct dependencies beyond those that fully determine the precedence relations between terminals.

(3) **Non-redundancy**
there are no direct structural dependencies that are not required to satisfy Totality

The principles of dependency theory entail that direct structural dependencies always link adjacent categories, they cannot skip over them. This is not difficult to see. Consider the case where there is some element, α, intervening between the two members of a direct dependency, as in (4):

(4) a. 

b. 

Such a configuration violates either Totality or Non-redundancy. Totality requires that there be some dependencies that force α to precede the terminals dominated by X and follow the terminals dominated by YP in (4). Suppose that such structural dependencies are present as in (4b). Then X indirectly depends on YP by virtue of these and the fact that the terminal dominated by X follows the terminals dominated by YP will be ensured. Thus the structural dependency of X on YP is unnecessary to order the terminals dominated by them and therefore it will violate Non-redundacy.

So spec-head and head-comp dependencies presuppose adjacency between

the linked elements. This has two immediate consequences. First a head can have only a single spec and a single comp and second the spec and the comp must be on different sides of the head. Both consequences follow since only one category can be adjacent to a head on each side.

There is a difference standardly made between specifier-head and head-complement relations that I shall assume and exploit. Head-complement relations are lexically determined, --these are taken to always hold between lexically related sister constituents. I take head-comp relations to be a special case of head-complement relations covered by the same generalization. Spec-head relations/dependencies on the other hand are much freer. They can hold between lexically, and (as I shall assume following essentially Pesetsky 1992) even semantically, unrelated categories. I shall therefore take spec-head dependencies to be in principle freely assignable, --although subject to certain restrictions to be discussed below. The axioms of dependency theory will ensure that this does not result in overgeneration.

Notice now that if spec and comp were understood in both hierarchical and dependency terms then dependency theory would force hierarchical structures to be strictly binary branching: there can be only one spec and one comp associated with any head. But since I understand spec and comp of a head X as categories in a direct structural dependency relation with X, these considerations do not exclude structures in which a head has more than one specifiers or complements. As long as the additional maximal projections do not intervene between the spec or comp and the head, there can be dependency structures that do not violate the axioms. For example in (5a) the head X has its spec and comp dependency (to QP and from RP respectively) and in addition it has another complement:

(5)a. QP X RP $[_{YP}$ Y ZP]

As just noted, spec-head dependencies are assigned freely, hence unless ruled out by the dependency axioms there will be spec-head dependencies that do not correspond to the configurational spec-head relations of standard structural analyses. Such a spec-head dependency from the head of YP to RP, the comp of X can provide the required ordering for the terminals dominated by YP in (5a).

We can define the **complement** of X as any sister category that (+/- properly) contains some element (directly or indirectly) structurally dependent on X. In a parallel way any sister category of X' that (+/- properly) contains some element on which X depends will be a **specifier** of X. The following four points need to be mentioned in connection with these definitions. First I shall argue below for the elimination of the X'-level, and accordingly I should substitute X for X' in the definition of the specifier. Secondly note that a spec is not necessarily a specifier in the present theory: specifiers but not spec's are defined as sister nodes. Thus for example RP in (5a) is the spec of Y, since Y directly structurally depends on it, but RP is not the specifier of Y: it is outside YP. Thirdly, I assume that in a structure like (5a) if RP is a PP, the head P of this PP can always reanalyse with the head X forming a complex head whose comp is the comp of P. Thus if the head of the

sister phrase (corresponding to YP in (5a)) following this PP structurally depends on the comp of P, this second sister will still satisfy the definition of complement of X: it indirectly depends on the reanalysed X + P. Fourthly note also that the system allows not only for multiple complements but also for multiple specifiers. For example in a structure like (5b), where YP is the spec of X and both QP and YP are specifiers of X, the axioms of dependency theory can be satisfied. Similarly to the multiple complement case of (5a), a spec-head dependency between QP and Y can order the terminals in the two specifiers.

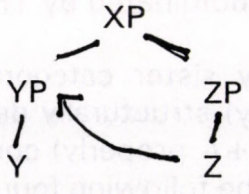
(5)b. QP $[_{YP}$ Y ZP] X

Consider next the choice between spec-head-comp and comp-head-spec word orders. The assumption that the dependent element must follow the one it depends on is a natural one given general processing considerations. Thus if spec depends on the head and the head on comp then Precedence entails a comp-head-spec word order, while if the comp depends on the head and the head on the spec then Precedence forces the spec-head-comp order. Spec-head relations give no straightforward clue as to which is the correct choice, but comp clearly is a function of the head rather than the other way around. Hence comp depends on, and therefore by Precedence follows, the head. Since, as we have seen, the spec must be on the opposite side of the head the word order must be spec-head-comp. It then follows also by Precedence that the head must depend on the spec. If complements and specifiers are also defined in dependency terms as just suggested, then all specifiers will precede and complements follow the head, --again as a consequence of Precedence.

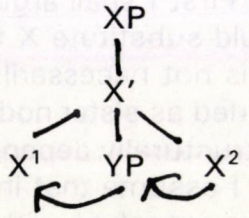
I.2. A Minimal Theory of Phrase Structure

Let us next turn to some of those predictions of Kayne's LCA for phrase structure that the proposed dependency theory does not reproduce. Kayne's LCA is violated by XP's that have no X^0 head as in (6a) and ones that have more than one such head as for example in (6b). ((6b) can also be thought of as a structure where the complement of a head is not a phrase but another head.)

(6) a.



b.



Dependency theory on the other hand has no such consequence. The dependency

structure in (6b), where YP to X^1 is a head-comp and X^2 to YP is a spec-head dependency, does not violate the principles of dependency theory. As for (6a), if ZP has no specifiers then a spec-head dependency from Z to YP will order the terminals appropriately. Nothing forces the simultaneous presence of a dependency between Y and ZP that would result in a violation. It seems to me however that the fact that dependency theory does not entail that a phrase must immediately dominate a unique head is in fact an advantage, since it eliminates a potential redundancy.

Since phrases and their heads share properties, the assumption that phrases are in some sense projected by their heads seems ineliminable. The fact that a phrase must have a unique head will follow if we assume that phrases can only arise through projection and that furthermore all heads must project phrases. Since the latter statement is false for word-internal heads, a minimal statement of this would have to be along the lines of (7), which I shall call the Principle of Phrasal Projection.

(7) **Principle of Phrasal Projection (PPP)**

a. Every non word-internal head must project a distinct phrase

b. Every phrase is projected by a (non word-internal) head

The PPP expresses the idea that syntactic categorial structure is projected from the lexicon. This involves two assumptions. Clause (a) of the PPP states that a precondition for a lexical element to enter the syntactic structure is for it to project some nonlexical category, ie. a phrase. Clause (b) 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 now from clause (b) of the PPP, that is from the fact that all phrases are projected by their heads. That a phrase must not have more than one head will follow from clause (a).

A basic property of projection is that it is local: a phrasal node projected by a head is always near to this head. Assume that the relevant notion of locality is immediate domination: the phrase projected by a head must immediately dominate this head. Then a phrase HP will immediately dominate a unique head H, --its own. Any other head must be immediately dominated by the phrase that this other head projected, necessarily distinct from HP as we have seen.

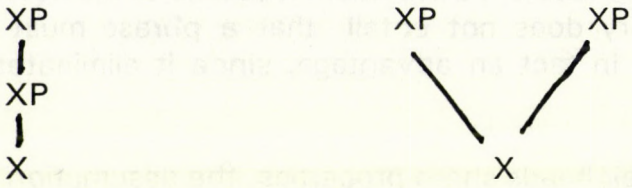
Consider next the converse statement: that a head projects a unique phrase. Optimally this should also hold, since it would radically simplify the theory of phrase structure. Suppose that it does. Let us first ask why this should be so. We have already seen that a head must project a phrase, so we only need an account of why it cannot project more than one. Uniqueness of the projected phrase will immediately follow if the locality relation between the head and the projected phrase is understood strictly, say as immediate domination defined as in (8) (where x,y,z are categories and domination is irreflexive):

(8) x immediately dominates y iff x dominates y and for all z distinct from x

if z dominates y then z dominates x

Locality as in (8) rules out multiple projection by one head not only in the standard configuration of (9a) but also in the "inversely branching" (9b).

(9) a.



It thus follows also that a head can project only a single phrase, since only a single phrase can immediately dominate the head that projected it.

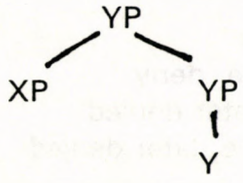
I thus assume that in the context of the theory that assumes presyntactic projection of phrases, the biunique relation between phrases and non word internal heads projecting them follows from the PPP and the locality requirement on projection. Optimally, the PPP and the locality requirement should exhaust the theory of phrase structure. Suppose that they do. Then the theory of phrase structure requires that every XP consists of a unique head X and an arbitrary number of other phrases. As we have seen, dependency theory ensures that only one of the included phrases can serve as a spec and only one as comp, the designated first complement. The spec and the specifiers will precede and the comp and the other complements will follow the head for the reasons given.

Current theories of phrase structure diverge from this simple picture which only contains the configuration where an XP dominates a head and a number of other XP's 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. These two additions can be reduced to one if, as proposed by Kayne (1993), the intermediate X'-level is treated as the lower segment of adjunction. But the ad hoc segment-category distinction and the attendant complications in the definition of c-command necessary in this theory make a stipulative configuration out of adjunction. Given such complications, adjunction is clearly not a notion "drawn from the domain of (virtual) conceptual necessity" to use Chomsky's (1993) terms.

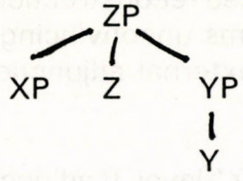
Within the present framework word-external adjunction is both impossible and unnecessary. The fact that adjunction is not a possible option word-externally, follows immediately from the PPP. Since no segment-category distinction is postulated, adjunction would by definition create a new category without a head projecting it, violating the PPP. There are at least two alternative structures for phrases standardly treated as adjoined. First since multiple specifiers are allowed, an adjoined category can be analysed as an additional specifier. This would be a natural approach for example to multiple wh-elements in phrases headed by a +WH category. (Notice that the "head of COMP" (Lasnik and Saito 1984) ie. the substituted wh-phrase in the specifer of the +WH C under standard treatments can

still be distinguished from the other "adjoined" wh-phrases: this element is the spec of the +WH head under the dependency module, while the other wh-categories are only specifiers of this head but they do not serve as its spec.) Alternatively an adjoined category can also be taken to be the specifier of some higher head. Under this option, instead of left-adjunction of XP to YP as in (10a), we can have the configuration in (10b) with the higher head Z which is either invisible for selection (selectional requirements are satisfied by the lower head Y) or it has the ability to satisfy the same selectional requirements as the lower head Y. (Apparently fronted wh-phrases in some multiple wh-fronting languages instantiate also this possibility, --cf. Rudin 1988).

(10) a.



b.



Thus we can dispense with phrasal adjunction in general, --a welcome result for the reasons noted. See also Sportiche (1994) who reaches the conclusion independently, that adjunction does not exist in (presumably word-external) syntax.

Chomsky 1994 develops a theory that radically restricts word-external adjunction to 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 this configuration. LF adjunction of the associate to its expletive chain-mate is a problematic and probably unnecessary operation cf. eg. Brody 1993. The necessity of adjoined intermediate traces in non-uniform chains is equally moot (cf. eg. Manzini 1992). As for scrambling, Chomsky suggests that LF reconstruction will provide an account of the contrast he finds between (11a) and (11b). Here the expectation is that (11a), 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 (11b), this example will not be similarly excluded:

(11) a. Pictures of John$_x$'s brother, he$_x$ never expected that I would buy
b. Which pictures of John$_x$'s brother did he$_x$ expect that I would buy
c. Near John$_x$, he$_x$ saw a snake

Note first that if topicalization is treated as adjunction then in the minimalist

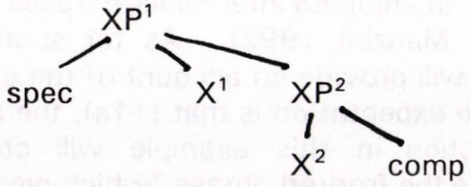
framework it is quite unclear how the PF presence of the topicalized element in the fronted position gets triggered. (For a substitution analysis of this construction cf. Watanabe 1992.) Secondly, it is not clear if there really is a contrast between (11a) and (11b), especially in view of well known examples like (11c), that appear to allow coreference. Thirdly forced reconstruction of adjuncts would loose the explanation of the contrasts like those between (12a) and (12b) or (12c) and (12d), 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 to appear, for different ways of instantiating this idea). If topicalization is adjunction and adjuncts are reconstructed then (12a,b) should contrast with (12c,d) rather than (12a) and (12c) with (12b) and (12d):

(12) a. Which claim that John$_x$ made did he$_x$ deny
b.?*Whose claim that John$_x$ was asleep did he$_x$ deny
c.The claim that John$_x$ made about this, he$_x$ later denied
d.?*The claim that John$_x$ was asleep all day he$_x$ later denied

Thus the evidence for an account involving forced reconstruction of elements adjoined to semantically nonvacuous categories seems unconvincing. It therefore cannot support the more general claim that word-external adjunction --although under restricted circumstances-- exists in syntax.

Consider next the question of the intermediate X'-level. If adjunction does not exist, then clearly this category type cannot be treated as a segment of adjunction. Suppose that there is no intermediate X' level. Then the question arises, how spec(ifiers) and comp(lements) can be distinguished. We could do this 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 (13). The empty head X^1 and the lexical head X^2 are then taken to form a unit, -- a head-chain. (The tree in (13) is only partly Larsonian, since although it involves an empty shell, it is not binary branching.)

(13)



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.

I do not think that this solution is the correct one however. Larson's empty shell approach to the evidence that is generally taken to motivate binary branching structures is incompatible with the generalization that categorial projection and the

selectional properties of a head must be satisfied in the root position of its chain (Brody to appear). This problem would carry over to the analysis of the phrase in (13). 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 (13) 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.

The generalization that Larsonian empty shell analyses of multiple complementation violate was captured in the pre-minimalist framework by the concept of deep structure. 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 generalization it expressed (Brody 1993, to appear, see also Chomsky 1993, 1994 for relevant discussion). This generalization is a major and a pervasive one. Not only categorial projection and thematic selectional requirements, but syntactic and semantic selection in general hold invariably in the root positions of chains. Thus for example a verb V raised to some higher functional projection, say C, never forces the spec and complements of this head to satisfy the selectional requirements of V. I argued in earlier work that an appropriately formulated projection principle is compatible with a minimalist framework and is in fact necessary to ensure the generalization that selection and categorial projection holds invariably in the root positions of chains. Thinking of the relevant features as "projectional" ie. as being projected from the lexicon, I proposed that the projection principle requires that projectional features must hold in and be satisfied by the root positions of chains:

(14) **Generalized Projection Principle (GPP)**
Projectional requirements can involve only the root positions of chains (ie. they can hold in, and be satisfied by root positions only)

Thus I take the projection principle to require not only that the relevant features hold only in root positions but also that they must be satisfied by root positions. Given this further natural generalization the projection principle entails that only the root positions of (XP-)chains can be thematic (ie. that movement cannot land in a theta position) --the Main Thematic Condition.

The projectional features include centrally the categorial features of a head. Phrasal nodes are projected by heads, so categorial features are under the jurisdiction of the projection principle. As expected, these also take effect invariably in the root position of the relevant head chain: a verb forces the presence of a phrasal node VP in the root position of its chain, but not in any higher position. A V raised to C for example will not turn the CP into a VP. On the other hand a categorial feature can apparently be assigned to a phrase in non-root positions in XP-chains. I assume that this is because categorial features, in contrast to selectional features, are assigned to categories and not to positions. Both types of features are assigned though through the position occupied by the relevant head, hence by the GPP only in root positions of chains. The generalization expressed by the GPP in (14) thus captures the behaviour of categorial and various selectional

features and in addition it entails also the Main Thematic Condition. Since Larsonian empty shells violate this pervasive generalization, the attempt to find an alternative approach seems well-motivated.

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 head whose features are shared between a number of head positions (This consequence of Larson's approach is noted in Brody to appear.) Multiple complement verbs under a Larsonian analysis would all require such a decomposition treatment. Consider applying this treatment to the present problem of eliminating the intermediate X'-level in terms of a structure like (13). If X is decomposed into X^1 and X^2 and categories standardly taken as sisters of X' and sisters of X are distinguished as sisters of X^1 and sisters of X^2 then also simple transitive and intransitive heads must decompose into two heads. The verb **see** would have to be composed of an agent selecting segment and a non-agentive SEE, something like the passive **was seen**. While this is logically possible, there appears to be little independent evidence for proceeding along these lines. Furthermore 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**. This seems to reduce the approach to vacuity.

But there is no immediate need to explore this avenue further. As far as the intermediate X'-level is concerned, we in fact already have the alternative solution. Given dependency theory, the asymmetry between spec and comp is ensured without a difference in their c-command relations to each other. Thus according to the analysis that results from the interaction of dependency theory and the minimal theory of phrase structure based on the PPP, the basic structure of an XP will be (15):

(15)

In (15) YP, the comp asymmetrically depends on the head X, and X asymmetrically depends on ZP, the spec. Since, more generally, complements and specifiers were also defined in dependency terms, it is unnecessary to express this asymmetry also in the hierarchical structure. In sum the minimal, "virtually conceptually necessary" theory of constituency can be assumed which requires nothing beyond that heads project phrases and phrases be projected by heads, --and which entails that a phrase contains its head and optionally an arbitrary number of other phrases.

I.3. Some Further Consequences of the PPP and the GPP

Chomsky (1994) rejects the assumption that "certain features (categorial features) project from a terminal element to form a head, then on to form higher categories with different bar levels". There are at least two assumptions involved here: (a) categorial features project to label higher categories and (b) higher categories have different bar levels. But he assumes at the same time that the operation of Merge: "project[s] one of the objects to which it applies, its head becoming the label of the complex formed". Thus assumption (a) that categorial features project is in fact not rejected, categorial labeling is accomplished by Merge, which in addition carries out the task of joining (projected and unprojected) elements. It is assumption (b) that projection creates different bar levels that is rejected. Following Muysken (1982), Chomsky assumes that bar level status (minimal, maximal, both or neither) is contextually determined: a maximal projection is one that does not project further, a minimal projection is the lexical element itself, an intermediate projection is one that is neither maximal nor minimal.

Within the minimal theory of phrase structure proposed here, there are no intermediate projection levels, and thus a category is either a head or a phrase. Under a set-theoretical formalization these would correspond to a non-set element and a set respectively. Thus there remains no 'bar level status' and therefore no question as to how this is to be determined, --an unprojected element is a head and a projected one is a phrase. Note also that since branching is taken to be n-ary there is no need to restrict (directly or indirectly) the principles building syntactic structures to binary operations. Given the simplest assumption about branching, namely that it is in principle unrestricted, there will be both unary and more than binary branching structures.

The PPP allows a further simplification. As we have seen clause (b) of the PPP ensures that every phrase must immediately dominate a head that projects it and clause (a) entails in the context of presyntactic projection that this head is unique. This means that there is no need to label phrases at all, phrasal labels can be thought of as only informal notation. $\propto$P becomes shorthand for a phrase headed by $\propto$. The PPP thus eliminates the indeterminacy of labeling present for example in Chomsky's (1994) system, where the operation of Merge applied to two categories has to specify which of the two projected. The phrase structure becomes genuinely "bare": phrases correspond to sets whose elements are either other phrases/sets or non-set elements/heads. The additional and quite unnatural complication of allowing some of the elements in these sets to be labels can be dispensed with.

Recall that word-internal heads do not need to project phrases, hence such heads must be exempted from this requirement. Clause (a) of the PPP was formulated accordingly. Clause (b) was stated symmetrically: all phrases must be projected by **word-external** heads. This has the important effect of excluding phrases adjoined to heads: such phrases would necessarily be projected by word-internal heads. (I suggested in an earlier version of this paper that phrasal adjunction is ruled out by a PF-requirement: words cannot contain phrasal boundaries at PF, --cf. also Chomsky 1994 for essentially the same proposal. But Gugliemo Cinque points out that examples like "his out of this world attitude",

where the stress pattern indicates a word-internal phrase, may counterexemplify this analysis. Such examples are in fact strongly suggestive that a PF condition is not the relevant one, since they seem to result from a marginal or ungrammatical construction being saved by the device of creating a PF phrasal word. The condition prohibiting such configurations would therefore appear to be strictly syntactic and not phonological in nature.)

Since the PPP is not sensitive to whether a head is in the root position of its chain, it provides a straightforward answer also to why "moved" heads (ie. heads in non-root positions) cannot substitute into or adjoin to maximal projections. This is because the resulting configurations equally violate the PPP: such heads are not dominated by a phrase projected by them.

As we have seen, the GPP ensures that a head in a non chain-root position cannot project, and thus in particular it cannot project a phrase. Suppose that such a "moved" head H adjoined or substituted to some phrase avoids exclusion by the PPP by projecting a phrasal node HP. There are two options to exclude: the projected phrase HP may be either internal to the phrase to which H substituted/ adjoined or it may force the category label of the target phrase or segment to be HP. Both options are of course excluded by the GPP: projection is restricted to root positions. Thus the GPP and the PPP 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 the PPP and the GPP. They only create a contradiction for word-external heads in non chain root positions. The PPP requires all word-external heads to project a phrase and the GPP restricts all projection to root positions of chains. Hence word-external heads that are not in root positions can neither project nor not project: they cannot exist.

Let us turn finally to the question of the explanation of the two non-dependency principles involved in this account, the PPP and the GPP. As for the PPP there does not seem to be much to explain. This principle states that all and only word-external heads have a phrase projecting property and that there is no lexicon-independent syntactic category: phrases must be projected by heads. These seem to be natural and minimal assumptions.

In Brody (to appear) an explanation of the GPP is provided along the following lines. The question is why checking of projectional features has to hold and (at least in the case of selectional features also be satisfied in) the most deeply embedded position in the chain. Take a configuration in which this is not the case, where a head is in a chain in which it projects (categorially and/or selects) from a non-root position or where an XP is selected in a non-root position of its chain. Suppose that a head can only project and an XP can only be selected in a single position in a chain. (This may be a simplification, cf. Brody to appear). Suppose however also that all positions in a chain must be projectionally identified. An assignee position is so identified if it has the appropriate feature while the assigner position is identified if it has some feature indicating that proper assignment has taken place. Suppose finally that feature percolation in chains can only take place bottom to top, it is strictly upward directional. It follows that the projectional

feature must be assigned to the most deeply embedded position in the assignee chain, otherwise lower positions in this chain will not be projectionally identified. Similarly the projectional feature must be assigned from the most deeply embedded position of the assigner chain, otherwise the feature indicating the satisfaction of the projectional requirement cannot percolate to all members of the assigner chain.

The requirement that feature percolation in chains is strictly upward is in effect the representational equivalent of the principle excluding lowering. In a derivational framework 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 projectional features or through lowering from this position. Downward spreading of the projectional features in a representational theory appears to correspond to a lowering in a derivational system. This needs to be excluded. But since projectional feature checking like all feature assignment by heads can only involve a single position, the GPP now reduces to the principle that all positions in a chain need to be projectionally identified.

(Notice that while the GPP follows from simple assumptions once the equivalent of lowering is excluded in a representational theory, the same is not true in a derivational system. Excluding lowering rules would not help to explain why raising into a projectional position is impossible. See the appendix for a discussion of Chomsky's (1994) alternative explanation of the effects of the GPP.)

II. Dependency and Binary Branching

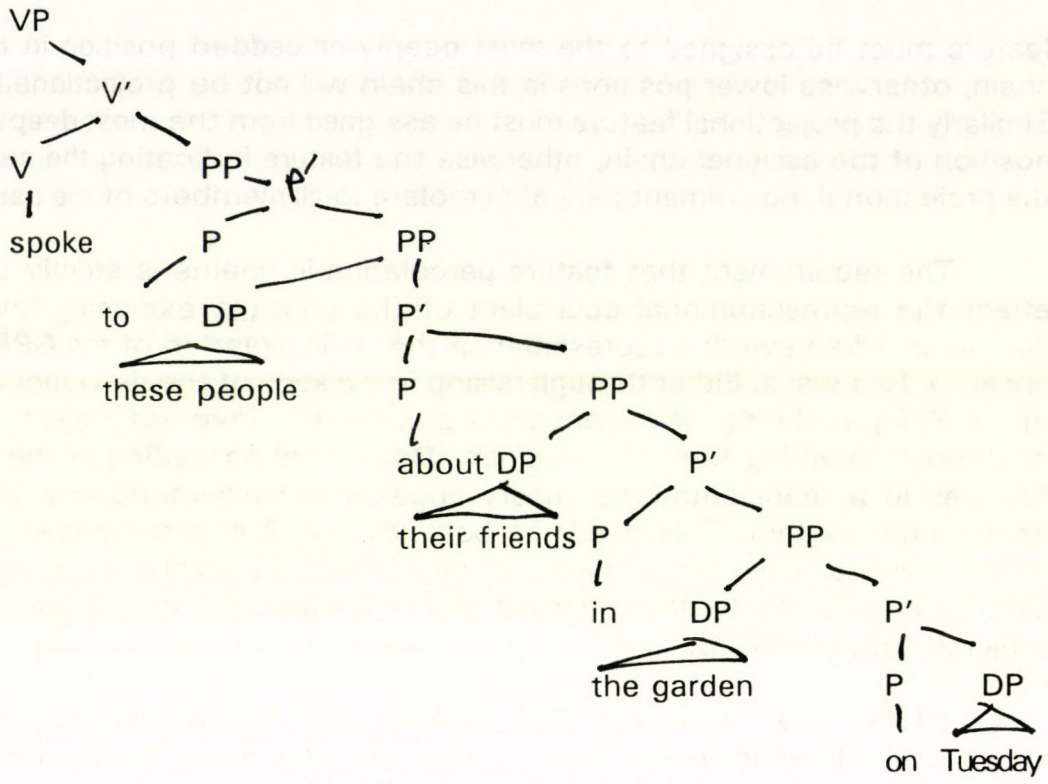
II.1. Dual Derivations and Dependency

Asymmetric behaviour of anaphoric relations in the VP can be analysed in terms of precedence in addition to c-command. For example the paradigm in (16) through (19) (from Pesetsky 1992) can be accounted for if we assume that the antecedent, in addition to the c-command and locality requirements, must in these cases also precede the anaphor.

(16) Sue spoke to these people$_x$ about each other$_x$'s friends in Bill's house
(17) a. John spoke to Mary about these people$_x$ in each other$_x$'s houses
b. Mary danced with these people$_x$ in each other$_x$'s hometowns
(18) *Sue spoke to each other$_x$'s friends about these people$_x$
(19) a. *Sue spoke to Mary about each other$_x$'s flaws in these houses
b. *Mary danced in each other$_x$'s cities with these mayors$_x$.

The curious conjunction of precede and c-command can be eliminated in terms of c-command defined on binary rightward branching structures. (eg. Larson 1988, Pesetsky 1992, Kayne 1993) Given the evidence discussed above against the Larsonian empty shell approach from the projection principle, I shall adopt as a basis for discussion the analysis in Pesetsky 1992. Given binary branching trees like those in (20), the antecedent will c-command the anaphor in all and only the grammatical examples in (16) through (19): precedence is eliminated.

(20)



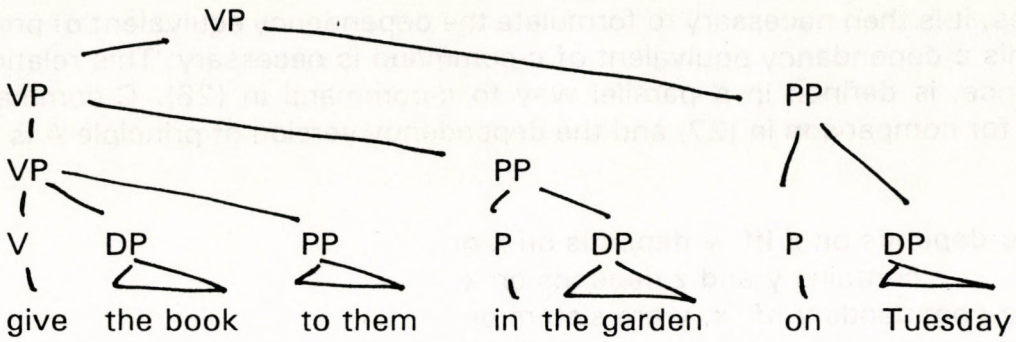
As Pesetsky emphasizes, although such structures appear to be motivated in various ways, they do not provide the appropriate constituent structure for movement/chain relations. As (21) shows, on many of the constituents in (20) no non-trivial chain can be formed. The members of the chains in (22) and (23) are not constituents of the binary branching tree, but only of the more traditional representation in (24), --the examples are again Pesetsky's, with minor modifications:

(21) a. *[To John about himself] Mary spoke __
b. *I wonder [to whom on Tuesday] Mary spoke __

(22) a. [To which adults] Sue showed the kids __ on each other's birthdays?
b. [To none of the officials] did Sue send money __on each other's birthdays
c. [On which table] did John put the book __ during its construction?
d. [To the children] were given books __ on each other's birthdays.

(23) a. ... and [give the book to them in the garden] he did [__ on each other's birthdays]
b. ... and [give the book to them] he did [__ in each other's garden]

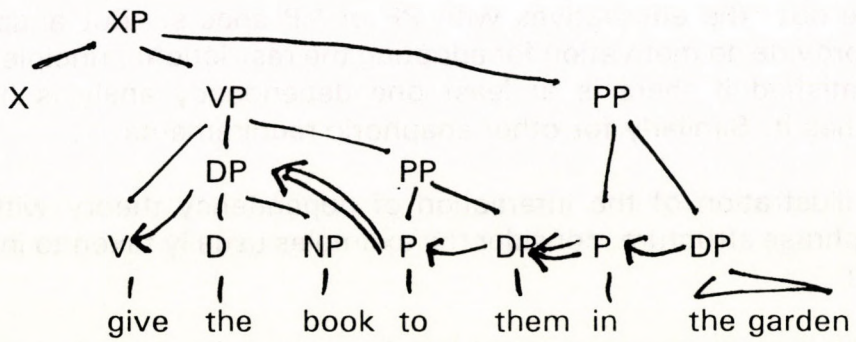
(24)



Thus there appears to be systematic evidence for both types of phrasal organization. Accordingly, Pesetsky postulates that sentences have two parallel derivations, one involving binary branching and the other more traditional trees, his "cascade" and "layered" representations. He suggests that selectional and other related restrictions hold in both derivations and postulates a set of correspondence principles to relate them. But this means that the approaches in terms of binary branching structures to the problem of anaphora are not fully successful either since they create another unexplained duplication.

Dependency theory captures the operative relations of binary branching analyses and thus it makes such analyses dispensable. In discussing the relation between dependency theory and the theory of phrase structure I have been concentrating on structures where a phrase contains only two maximal projections, a spec and a comp. Let us now consider more closely the situation where an XP immediately dominates more than one complement: VPs with multiple complements and non-subcategorized elements.

(25)



In order to ensure the correct constituent structure, I analyse the non subcategorized PP in (25) as a sister of a higher head. (Recall that I assume the nonexistence of phrasal adjunction.) Single line arrows indicate the lexically determined head-comp dependencies. These relations leave the two PP's "to them" and "in the garden" unordered with respect to each other and the rest of the structure. Let us assume that structural dependency relations cannot cross argument (CP and DP) category boundaries. One of the ways in which (25) can satisfy the axioms of dependency theory is if the two spec-head dependencies (indicated by double line arrows) are present in the structure. This will ensure that the structure expresses both the constituency facts and the asymmetry shown by the behaviour of anaphoric elements.

Given the assumption that anaphoric relations are defined over dependency structures, it is then necessary to formulate the dependency equivalent of principle A. For this a dependency equivalent of c-command is necessary. This relation, c-dependence, is defined in a parallel way to c-command in (26). C-command is restated for comparison in (27) and the dependency version of principle A is given in (28) :

(26) y c-depends on x iff y depends on x or
z contains y and z depends on x

(27) x c-commands y iff x, y are sisters or
z contains y and x , z are sisters

(28) **Principle A**
An anaphor must c-depend on an antecedent in its local domain

Since all dependencies are right to left by the Precedence requirement, it is easy to see that the principle A in (28) will be violated by the ungrammatical examples in (18) and (19). In contrast, those in (16) and (17) will satisfy (28), since in each of these the anaphor is in a DP that indirectly depends on the antecedent of this anaphor. It is clear then that the analysis in (25) can account for both the constituency and the anaphoric relations exhibited by multiple complement VPs.

Notice that there are alternative dependency structures for (25). The spec-head dependency from the preposition **in** could lead not to the DP **them** but to the PP or VP dominating this DP. In sentences like (16) and (17) similar dependency structures would violate principle A: under such analyses the anaphor would not c-depend on its antecedent. If we assumed that only argument categories (DPs and CPs) can serve as non-derived (chain root) spec nodes (Kayne 1984, Pesetsky 1992) then only the dependency analysis indicated in (25) would be legitimate, this principle would rule out the alternatives with PP or VP spec's. But anaphoric relations appear to provide no motivation for adopting the restriction: principle A for example can be satisfied if there is at least one dependency analysis of the structure that satisfies it. Similarly for other anaphoric requirements.

For another illustration of the interaction of dependency theory with the multiple branching phrase structure, consider the examples usually taken to involve Heavy Shift in (29):

(29) John offended t [by not recognizing pg] [his favourite uncle from India]
(30) We gave t to them on Friday [copies of the reports on each other]

This operation exhibits both raising and lowering effects. The fact that it appears to license parasitic gaps as in (29) shows that the shifted phrase is higher than its trace. On the other hand the fact that an anaphoric elements is licensed inside the shifted phrase whose antecedent is not higher than the trace of the shifted phrase suggests lowering.

Williams (1990, 1992) has denied that Heavy Shift has properties of A'-chains, arguing against the claim that it licenses parasitic gaps. He takes examples

like (29) to be derived by the operation of Right node raising. His evidence is that constructions with parasitic gaps like (29) allow stranded prepositions, as in (31). As he points out the option of stranding prepositions is a property characteristic of Right node raising but not of Heavy Shift:

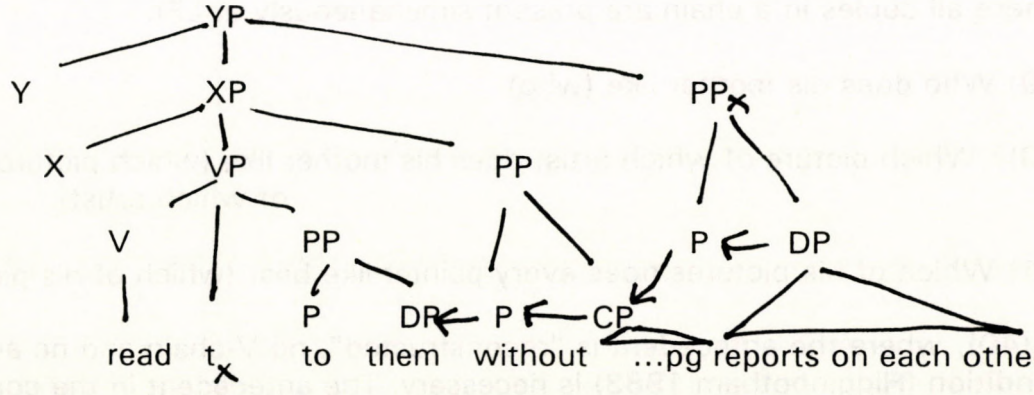
(31) John yelled at t, before punishing pg, all of those campers
(32) John talked to, and I yelled at, Peter
(33) *I talked to, about Mary, his favourite uncle

Although this data might suggest a way of avoiding the conclusion that Heavy Shift can license parasitic gaps, it clearly does not force this conclusion. This evidence is equally compatible with the claim that parasitic gaps are licensed both by Heavy Shift and by Right node raising, although prepositions are strandable only by this latter operation. Examples like (34) show that the conclusion that there is some operation (presumably Heavy Shift) that has both raising and lowering properties cannot be avoided in the way Williams suggests. These examples exhibit both types of properties (leftward parasitic gap licensing and rightward anaphor binding) simultaneously:

(34) a. I read to them, without carefully checking pg, several reports on each other
b. We read to each boy, after discussing pg in private, a report on his activities.

If anaphora is a dependency relation and the licencing of (syntactic) variables --whether ordinary traces or parasitic gaps-- is not, then the modular analysis that postulates a dependency structure again predicts exactly the observed facts. As the structure of (34a) in (35) shows, the trace and the parasitic gaps are c-commanded by the shifted phrase but the anaphoric element inside the shifted phrase can c-depend on its antecedent (I assume with Pesetsky that the shifted phrase is a PP):

(35)



Thus the modular interaction of dependency theory with the minimal theory of phrase structure can reconstruct Pesetsky's dual derivation analysis, without the duplicate selectional restrictions and correspondence principles of his analysis. But perhaps even more importantly, the interaction of dependency and phrase structure theory is naturally taken to provide a classification of syntactic relations. Given the Precedence principle, we expect precedence or binary branching effects between

two elements just in case that there is a dependency relation between them.

II.2. Dependency and Precedence

Typical such cases involve anaphoric dependencies: anaphors under principle A, polarity item licensing or pronominal bound anaphora. Consider pronominal bound anaphora. Williams (1992, 1994) proposed an account of WCO and other related relations in terms of a restriction that anaphoric dependencies are governed by a precedence requirement, --an approach partly similar to Chomsky's "leftness" condition. Thus in the case of the WCO configuration in (37), the pronoun cannot be dependent on its putative antecedent **everyone** because this antecedent does not precede it:

(37) His mother likes everyone

Williams account of WCO will follow from the Precedence requirement proposed here if it is generalized in the obvious way: the restriction of Precedence to structural dependencies needs to be removed so that it constrains all syntactic dependencies. (Notice that to preserve earlier results, Totality and Non-redundancy must not be similarly generalized, they must remain conditions on structural dependencies.)

(38) **Precedence (Generalized)**
if x depends on y then the terminals dominated by y precede the terminals dominated by x

Given a copy theory of chains, the more complex, reconstruction cases of WCO will also fall out on the assumption that the bound pronoun must depend on an antecedent in A-position. (See Brody to appear, for arguments for a theory where all copies in a chain are present simultaneously at LF).

(39) Who does his mother like (who)

(40)? Which picture of which artist does his mother like (which picture of which artist)

(41) Which of his pictures does every painter like best (which of his pictures)

In (40), where the antecedent is "reconstructed", no V-chain and no accessibility condition (Higginbotham 1983) is necessary. The antecedent in the copy position is to the right of the pronoun. (41), where the pronominal is "reconstructed" is correctly allowed on the assumption that an anaphoric element in a chain needs to satisfy its dependency conditions in only one of its positions in the chain (Barrs 1986, Chomsky 1993), --here the copy position.

Williams generalizes his treatment of WCO as involving a dependency governed by precedence to other anaphoric relations including the relation between

a pronoun and its nonquantificational antecedent. Since the generalized Precedence principle of the dependency theory proposed here is identical to Williams' assumption about the restriction governing these cases it should be clear that the present account can be similarly generalized.

Consider in contrast principle C of the binding theory. This is not a restriction on antecedence/dependence but rather a disjointness requirement. For example in (42) coreference between **he** and the object **John** is prohibited whether or not there is an alternative antecedent for the pronoun.

(42) *(John$_x$'s mother said that)he$_x$ thought I liked John$_x$.

Therefore the present theory predicts that principle C must be stated in terms of c-command and that this principle will show no binary branching or precedence effects. Thus the fact that it does not require disjointness in the well known cases of (43a,b) is as expected if the adjunct and the extraposed clause are in a high enough position in which the pronoun does not c-command the name that the clause includes:

(43) a. We sent him$_x$ there in order to please John$_x$'s mother
b. Someone had phoned her$_x$ who Mary$_x$ met at the party

Again, as is well known, there is independent evidence from constituency tests that there exists a VP in (43) that does not include the sentence final clauses, hence also for the claim that the pronoun indeed fails to c-command the elements contained in these clauses. Final clauses like the ones in (43) can be left stranded by VP-deletion as (44) and (45) exemplifies, and sentential adjuncts can also be stranded by VP-fronting, as in (46):

(44) Although none of the MEN did who were visiting from NEW YORK, several of the WOMEN went to the concert who were visiting from BOSTON
(Culicover and Rochemont 1990)

(45) a. MARY sent him there in order to PLEASE John's mother and KLARA did in order to UPSET her.
b. Although MARY did in order to UPSET John's mother, KLARA sent him there in order to PLEASE her.

(46) ...and send him there Klara did, in order to please John's mother

Sentential adjuncts like the **in order to** clause in (43a) behave like the locative and temporal adjuncts and the Heavy Shifted elements considered earlier: under an anaphoric dependency like WCO they act as if they were embedded more deeply in the tree:

(47) Who$_x$ did you criticize t$_x$ in order to please him$_x$

As usual with configurations where an element on the right is in a higher position examples can be constructed that exhibit both configurational and dependency relations simultaneously, creating an apparent contradiction. (48) for example

violates neither WCO (a dependency effect) nor principle C:

(48) a. Who$_x$ did you talk to t$_x$ about her$_y$ before Mary$_y$ told him$_x$ not to listen
b. I sent each boy$_x$ to her$_y$ in order to make Mary$_y$ meet him$_x$

It has been suggested that principle C can be stated on binary branching trees. But judgements on examples like (49) are quite equivocal:

(49) a. Sue spoke to him$_x$ about Bill$_x$'s mother
b. Sue spoke to Mary about him$_x$ in Bill$_x$'s house

The slight unacceptability of these could be due to the infelicitious repetition of an R-expression presupposed in the context. The difference between a violation of this weaker prohibition against repeating an R-expression and genuine principle C effects is illustrated by (50) and (51). (50), which violates the weaker, presumably pragmatic, prohibition contrasts with the more strongly ungrammatical (51) that violates principle C:

(50) ?Bill's mother spoke to Bill
(51) *Bill spoke to Bill's mother
(52) a. Sue spoke to Bill$_x$ about Bill$_x$'s mother
b. Sue spoke to Mary about Bill$_x$ in Bill$_x$'s house

The examples in (52) appear to be on a par with (50) rather than (51), reinforcing the conclusion that principle C is not sensitive to precedence in the same way as principle A or WCO is. Thus it is not a principle stateable on binary branching trees. As noted, the present approach predicts this result: since principle C is not a dependency principle it will not constrain dependencies and should therefore be sensitive to standard c-command.

Notice that the present approach entails also that wherever there is subject object asymmetry with respect to principle C, the subject must be outside the (lowest) VP that contains the object: otherwise there can be no asymmetry effects.

Haider (1993, 1994) who proposes a version of the binary branching hypothesis notes problems with stating principle C on such trees. He then argues that principle C is not a structural principle at all, he claims that it could not be stated on non-binary branching trees either. He questions the account of (43b) that assumes that the relative clause extraposed from subject is in a high enough position in the tree where the object does not c-command it. Many of his arguments do not distinguish between various types of extraposition constructions. I shall look briefly below at some of those that are relevant to the question of the position of the extraposed clause in extraposition from DP constructions. His main argument is that extraposed relatives always precede an extraposed argument clause and object pronouns must be disjoint from any R-expression in the argument clause:

(53) Someone has told her$_x$ [who Mary$_x$ met] [that Mary$_x$ will inherit the castle]

If the pronoun in (53) must c-command the extraposed argument clause then there is no way to avoid it c-commanding the extraposed relative. Since no disjointness effect obtains with the R-expression in the relative this appears to question the possibility of a structural account of principle C.

There is strong evidence however that extraposed object argument clauses originate under the VP, as originally proposed by Stowell 1982 (see also Postal 1986, Pesetsky 1992). If we take this extraposition to be A'-movement, then principle C should hold for the VP-internal position, just like in the case of (54) it holds for the source position:

(54) *[Which claim that John$_x$ was asleep] do you think he$_x$ denied t

Hence (53) in fact creates no problems. It's structure is like (55), where the pronoun c-commands the R-expression in the appropriate position in the chain of the extraposed argument clause, but fails to c-command the R-expression in the extraposed relative whose structural position is higher:

(55) Someone has told [$_{VP}$ her$_x$ [that Mary$_{*x}$ will...]][who Mary$_x$ met] [that Mary will inherit the castle]

(Martina Wiltschko points out that when the R-expression is in a relative clause inside the extraposed object argument clause, the example improves. If correct, this provides further evidence for the analysis suggested here. Since non-selected arguments (adjuncts) like the relative clause reconstruct only optionally (Lebaux 1989), the contrast between this case and (53) would be predicted on the reconstruction analysis proposed -cf. Brody to appear.)

Another set of arguments against adjoining the extraposed relative clause higher than the VP and thus against the possibility of a structural account of principle C predicated on non-binary branching trees involves binding facts.

(56) I would not tell everyone$_x$ all the details at once [that he$_x$ might be interested in]

Haider argues that in (56) the trace of the quantifier must c-command the pronoun, otherwise the structure would give rise to a WCO violation. It would follow from this that the extraposed clause must be not higher in the tree than the complements of the verb, making a structural explanation of the fact that an R-expression in an extraposed relative can be object bound impossible. But as we have seen, WCO is a dependency principle, hence it provides no evidence for the c-command relations involving the extraposed clause.

Arguing against Culicover and Rochemont's IP-adjunction analysis, Haider shows that restrictive relatives cannot take antecedents split between elements in two co-ordinate clauses:

(57) *I'll interview everyone and tape every man here who know each other

very well

He takes this to be evidence for the strictly right branching analysis. If the clause is embedded in the VP of the second conjunct then it cannot be construed with an element in the first conjunct which does not c-command it. But the example indicates only that the restrictive relative is not IP-adjoined, since the ungrammaticality of (57) follows as long as the clause is VP-internal. It does not matter for this explanation how deeply the clause is embedded in the VP.

I conclude that Haider's cases raise no difficulty for the structural account of principle C. This can be maintained, as long as trees are not strictly binary branching. Principle C thus continues to provide additional evidence for the necessity of a non-binary branching phrase structure analysis.

III. Summary and Conclusion

Two types of relations have recently been argued to have ordering consequences for syntactic categories. Kayne (1993) argued that asymmetric c-command relation correlates strictly with the order of terminals, while Williams (1992) has revived the idea that anaphoric dependency entails precedence. An important argument in favour of the dependency theory proposed here that it reduces these two separate ordering principles to one.

I argued in the first part of this paper that given a dependency module the theory of phrase structure can be radically simplified. Since the asymmetry of spec and comp and also specifiers and complements is provided by this module, this asymmetry does not need to be expressed by the phrase structure. I proposed that the theory of phrase structure consists essentially of the the assumption that heads project under a condition of locality (immediate domination). This entailed biuniqueness of the relation between heads and phrases projected by them. I proposed that there is no adjunction to phrases and no intermediate projection levels. Labeling of phrases becomes unnecessary. The GPP and the PPP ensure that phrases cannot adjoin to heads and that non chain root heads are always head-internal.

In the second part of the paper I showed that dependency theory can do the work of binary branching phrase structure analyses. There is systematic evidence against binary branching analyses from constituency tests as Pesetsky stressed, and as we have seen above also from the disjointness requirement of principle C. Pesetsky's dual system achieves descriptive adequacy here at the cost of accounting for the data by a complex two-faced system. Dependency theory provides a modular alternative: it offers a solution where the evidence follows from the interaction of two conceptually distinct and simple subtheories.

Dependency theory appears preferable to the dual derivation analysis for a number of more specific reasons. First since dependency theory is an annotation it avoids duplicate selectional and other contextual statements and correspondence

principles of the dual derivation system. Secondly, dependency theory is more parsimonious also in that it postulates only the operative relations of the binary branching structures: the spec-head and head-comp relations. Thirdly, since dependency theory postulates an asymmetry between the spec and the comp of a head, it has independent motivation in that it makes it possible to reduce the theory of phrase structure to the minimal projectional requirement. Lastly, dependency theory is more restrictive in that it provides a classification of syntactic relations: it predicts that binary branching/precedence effects will only be exhibited by various types of dependencies.

Let me finally turn to an additional argument from co-ordination for a dependency module. Constituents in Pesetsky's cascade structures correspond in dependency terms to a unit that contains a category (the spec or the head of the cascade constituent) together with all its (direct and indirect) dependents. Thus co-ordination of cascade constituents generally correpond to co-ordination of some category together with all its direct and indirect dependents. As Pesetsky notes there are some recalcitrant cases of co-ordination however where the co-ordinated elements do not form constituents under anybody's analysis:

(58) Mary will give [some books to John] and [some records to Bill] in the garden on Tuesday

Dependency structures provide the appropriate units for this type of co-ordination, that neither standard nor binary branching trees offer. Given the dependency analysis, we can take co-ordination in general to operate on a category C together with all categories that structurally depend on C. These sets of dependent categories will correspond to constituents under the binary branching analysis. To allow cases like (58) we can assume in addition that (spec-head) dependencies thatare semantically empty can behave as if they were not present, that is that they can be optionally ignored in the determination of the structural dependents of a category. Thus in (58) the spec-head dependency between **Bill** and **in** can be ignored and if it is then the phrase "some records" will have only **to** and **Bill** as its structural dependents. Hence "some records to Bill" can be a conjunct of co-ordination.

Appendix

Chomsky (1994) provides a different explanation of the effects of the GPP. In this appendix I shall comment on his approach. Chomsky does not discuss the problem of why the selectional (thematic etc.) requirements of heads have to be satisfied in the root position of their chain. This problem can be reduced to the question of why categorial projection is always in the root position on the assumption that only projecting heads can select. Let us make this assumption. Then one can distinguish two types of GPP effects. One effect of the GPP is that (i) categorial projection is always in the root position and another consequence as we have seen is that (ii) selectional (thematic) properties are always satisfied by root positions.
It is possible to think of (i) as creating four separate questions, (ii) adding one more:

(59) 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?

(60) Why is movement to a selected (theta) position prohibited?

Chomsky provides the following answers to these questions.

(59a,b), --heads in non root position cannot project: This is because the HMC would force such a raised head α to substitute into or adjoin to the αP, the phrase α itself projected. This is prohibited by the fact that such "self attachment" would create an ambiguity: it is not clear in such structures if the category/segment that dominates α in the non root position inherits its label (αP) from α in the non root position or from the αP that α projected in its root position. (In the "official" bare phrase structure notation that Chomsky proposes the resulting category would be $\{\alpha, \{\alpha, K\}\}$ for substitution, and $\{<\alpha, \alpha>, \{\alpha, K\}\}$ for adjunction.).

(59c), --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 α only if morphological properties of α itself would not be satisfied in the derivation". In a configuration like (61) if XP* raises to spec Y and then projects XP+, then XP* ceases to be a maximal projection, given the relational definition of projectional status.

(61) * $[_{XP+}$ XP_x* $[_{Y'}$ t_x]]

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

There is an additional reason for the ungrammaticality of (61), namely the Uniformity Condition, according to which a "chain is uniform with regard to phrase structure status", "where the "phrase structure status" of an element is its (relational) property of maximal, minimal or neither". This also rules out (61) since here the trace of XP* is maximal (by hypothesis) but XP* is not.

(59b,d), --adjoined heads cannot project and non-heads (phrases) cannot project further:

An adjoined element that projects would create the following configuration:

(62) *$[_{\alpha}$ α_x $[_K$ t_x]

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 (62). Whichever of the two categories, $\propto$ or the two-segment element [$\propto$, $\propto$] 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 [$\propto$, $\propto$] to be the head of the chain is ruled out additionally by the UC. (This seems to me to work only where $\propto$ is non-maximal, if $\propto$ is maximal then the UC is not violated.)

(60), -- the MTC effect of the GPP:

Chomsky attributes this also to the principle of Greed, the DP **John** in (63) 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.

(63) a. John [$_{VP}$ t' [HIT t]]
b. John [$_{VP}$ t' [BELIEVE [t to VP]]]

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".

Let me now enumerate some problematic aspects of Chomsky's account of GPP effects:

(a) It is not clear why the ambiguity of "self attachment" cases of adjoined and substituted projecting head should create a violation. Ambiguity of structures does not generally lead to ungrammaticality.

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 $\propto$ 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. $\propto$ 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.

(b) The raised XP in (61) 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'.

(c) Under Chomsky's bare PS system a category can be both a minimal and a maximal projection (eg. **it** or **John** in "John saw it"). So a category does not obey the spirit of the UC. Although this creates no serious technical problem, it is unclear conceptually why a chain, a set of categories, should then be constrained by this

condition.

(d) The UC would exclude word-internal head adjunction, so this configuration must be exempted: "at LF X^0 is submitted to independent word interpretation processes, WI" where WI ignores principles like UC. "WI is something like the LF analogue of Morphology..." Little independent evidence exists for the otherwise interesting concept of LF analogue to Morphology.

(e) 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, whichever category projects. 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 also above in section I.3.). But if WI can neutralize the problem when a minimal projection $\propto$ adjoins to X and X projects, WI will also neutralize the problem if $\propto$ 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 $\propto$ and [$\propto$, $\propto$] could not jointly serve as the antecedent of the trace.

(f) The explanation of the MTC based on Greed does not seem 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 ON, which is like "on" except that it does not assign Case. This should allow a structure like (64).

(64) a. I gave John Sunday ON t (cf. I gave John a book on Sunday)

The MTC and the GPP which entails it, predict that such structures are ungrammatical and that therefore prepositions like ON 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.

(g) 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 (65):

(65) *John reads often books

(66) [$_{VP}$ John [$_{V'}$ v [$_{VP/2}$ often [$_{V'}$ reads books]]]]

He suggests that (65) has the structure in (66), 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 (67), that contains the verb HIT that assigns no accusative but is otherwise like **hit**, --the case that violates the MTC:

(67) John [$_{VP}$ t [$_{VP/2}$ 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 (65), 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.

In sum there are a number of 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. (Chomsky does not appear to question the claim that the generalization of the GPP is exceptionless. Boskovic 1993 points out cases apparently problematic for the MTC. These could not be used to support Chomsky's analysis however, since they would create further problems for it. See however Brody to appear, for an alternative analysis of Boscovic's cases that maintains the MTC and thus the GPP in full generality.)

Barrs, Andrew. 1986. Chains and anaphoric dependencies. Doctoral dissertation, MIT, Cambridge, Mass.
Boskovic, Zeljko. 1993. D-structure, Theta Criterion and Movement into Theta Positions. Ms., University of Connecticut and Haskins Laboratories.
Brody, Michael. 1993. Theta theory and arguments. Linguistic Inquiry 24:1-23.
Brody, Michael. to appear. Lexico-Logical Form. Cambridge, Mass.:MIT Press.
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.
Culicover, Peter W. and Michael S Rochemont. 1990. Extraposition and the Complement Principle. Linguistic Inquiry 21:23-47.
Haider, Hubert. 1993. Detached Clauses --the Later the Deeper Ms., University of Stuttgart
Haider, Hubert. 1994. On the Universal Directionality of Branching. Paper presented at the Vienna GLOW meeting
Higginbotham, James. 1983. Logical Form, binding and nominals. Linguistic Inquiry 14:395-420.
Kayne, Richard. 1993. Connectedness. Ms.,CUNY.
Larson, Richard. 1988. On the double object construction. Linguistic Inquiry 19:335-391.
Lasnik, Howard and Mamoru Saito. 1992. Move-@. Cambridge, Mass.: MIT Press.
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. 1992. Zero Syntax. Ms., MIT.
Rudin, Catherine. 1988. On Multiple Questions and Multiple Wh Fronting. Natural Language and Linguistic Theory 6:445-501
Sportiche, Dominique. 1994. Adjuncts and Adjunction. Vienna GLOW abstract
Stowell, Timothy. 1982. Origins of Phrase Structure. Doctoral Dissertation. MIT, Cambridge, Mass.
Watanabe, Akira. 1992. Larsonian CP Recursion, Factive Complements and Selection. Ms., MIT, Cambridge, Mass.
Williams, Edwin. 1990. The ATB Theory of Parasitic Gaps. The Linguistic Review 6:265-279
Williams, Edwin. 1992. Thematic Structure in Syntax. Ms., Princeton University.
Williams, Edwin. 1994. Precedence and Presupposition in Antecedence. Paper presented at the Vienna GLOW meeting