Generalized Transformations and the Theory of Grammar[*]

Robert Frank
Department of Linguistics
University of Delaware
46 E. Delaware Avenue
Newark, DE 19711
rfrank@cis.udel.edu

Anthony Kroch
Department of Linguistics
University of Pennsylvania
619 Williams Hall
Philadelphia, PA 19104
kroch@linc.cis.upenn.edu

March 27, 1995

Consider the model of grammar advocated in Chomsky (1955, 1957): a non-recursive set of phrase structure rules generates kernel sentences. To these kernel sentence structures singulary transformations are applied, such as affix hopping and passive, and the resultant derived kernel sentences are combined using generalized transformations. Such generalized transformations build multi-clausal syntactic structures from simpler ones and are exploited to handle the phenomena of sentential complementation and coordination. The generalized transformation for sentential subjects given in Chomsky (1957), for example, looks as follows:

(1)   Structural Analysis of S_1: $NP - VP$
    of S_2: $X - NP - Y$ (X or Y may be null)
    Structural Change: $(X_1 - X_2 \; ; X_3 - X_4 - X_5) \rightarrow X_3 - to + X_2 - X_5$

This generalized transformation selects a sentence S_1, deletes its subject NP, prefixes it by the lexical item *to* and inserts the result into an NP position in another clause. Thus, given the kernel sentences in (2)a, this generalized transformation produces the sentence in (2)b.

(2)   a. S_1: John – prove the theorem
        S_2: $\emptyset$ – it – is difficult
    b. $\emptyset$ – to prove the theorem – is difficult

In Chomsky (1965), it was observed that the function of generalized transformations, that of building arbitrarily large pieces of phrase structure, can be taken over by the base component if phrase structure rules are not prevented from applying recursively. This seems appropriate since this restriction on the form of the base was an unnatural one to begin with; and as a result of this revision in the model, the need for generalized transformations in the grammar is obviated. Moreover, as Chomsky points out, a model of grammar which allows generalized transformations is less restrictive than one which does not. Since it appears that the power lent by generalized transformations to a grammar is never fully exploited, Chomsky proposes that they be eliminated from the theory of grammar on the ground that the theory of grammar should allow only grammars of possible natural languages.

There are, however, reasons why the elimination of generalized transformations is not entirely beneficial. At the empirical level, Chomsky (1993) points out that under standard assumptions

governing which elements may be generated at the level of D-structure, the phenomenon of *tough*-movement necessitates recourse to generalized transformation-like devices: since the subjects of predicates such as *tough* or *easy* do not receive theta roles, but are rather licensed via predication, they cannot be inserted by the base component at D-structure, as D-structure is supposed to be "pure representation of GF(θ)". Chomsky notes that the solution advocated in Chomsky (1981), that lexical insertion may take place after D-structure, is inadequate since these subjects can be arbitrarily complex, perhaps even containing another instance of *tough*-movement, as seen in (3).

(3)   That the economy was difficult to control is easy to see.

In order to derive such examples, we need to allow for parallel derivations of the subject clause and the remainder of the sentence, proceeding from distinct D-structures which are integrated prior to S-structure. This integration no doubt calls for a device similar to generalized transformations.

There is also a conceptual argument for retaining in the theory of grammar generalized transformation-like operations which compose phrase structure from smaller pieces. When phrase structure is built compositionally, the initial chunks constitute a natural domain over which grammatical processes might be defined. Consequently, the existence of locality conditions on such processes could be derived from the very nature of the grammatical derivation. Indeed, in Chomsky's (1955, 1957) generalized transformation-based theory, the kernel sentences formed the domain over which obligatory transformations applied. In standard versions of GB theory, though, in which the initial level of grammatical representation is a single, potentially unbounded, phrase marker, locality results only from explicit stipulation.

Given these reasons for permitting the use of generalized transformations, let us reconsider the force of the arguments against them. First, there is the question of the recursive application of the base. At least since the work of Stowell (1981), it has been widely accepted that the grammar does not contain an explicit set of phrase structure rules in the sense of *Aspects*. Instead, the set of allowable structural configurations is deduced from more general constraints on structural well-formedness, and the residue of phrase structure rules that does remain, i.e., X-bar theory, is viewed as a set of well-formedness conditions rather than as rewrite rules. This being the case, the unnaturalness of the stipulation that phrase structure rules may not apply recursively disappears: there is no set of phrase structure rules which is applying at all, recursively or otherwise.

The second objection against generalized transformations deals with their excessive descriptive power. One aspect of this lack of restrictiveness shows up as the systematic absence of certain cases of extrinsic ordering among transformations. As Fillmore (1963) noted in the context of the *LSLT* theory, there seem to be no cases of ordering among generalized transformations, or cases of a singular transformation which must apply to a matrix sentence prior to the application of a certain generalized embedding transformation.[1] Generalized transformations of the sort we are considering are unrestrictive also in their expressiveness. They can be written to rearrange pieces of two independent phrase markers in complex ways. However, such power was not exploited; typically generalized transformations were used only to substitute one structure into another. Further, generalized transformations are able to add and delete specific lexical material. In the case in (1), we see addition of the infinitival marker *to*, and in other cases conjunctions such as *and* and *but* were added by a generalized transformation. Under contemporary assumptions concerning the constrained nature of grammatical principles and operations, however, these sorts of operation will surely be disallowed. Extrinsic ordering among operations have long since been banished from syntactic theory. And just as construction-specific transformations of the *Aspects* model have

---

[1] Fillmore's own response to these observations is to construct a theory in which these absences follow from the very nature of the derivations. See Fillmore (1963) for further details, and Bach (1977) for illuminating discussion.

given way to the more general operation of move-α, so too can we expect complex language and construction-specific generalized transformations to be replaced by universally applicable operations. This removes the possibility of the arbitrary rearrangement of tree structures and addition of lexical material; and under this view of generalized transformations, there remains no conceptual reason to exclude them from the grammar.

Such a line has indeed been taken up in the "minimalist framework" of Chomsky (1993). Chomsky proposes to eliminate the D-structure level of representation and instead to derivationally compose phrase structure out of individual X-bar projections. These projections are combined using two universally applicable generalized transformations: GT and adjunction. Unfortunately, Chomsky's particular system of generalized transformations has some unnatural properties, and suffers from certain empirical problems. First, in this system one is forced to stipulate a distinction between certain applications of the generalized transformations GT and applications of adjunction. In particular, all applications of GT which occur prior to SPELL-OUT, and more generally all applications of substitution, must "extend" their target, while instances of adjunction need not. Further, the extension constraint (coupled with a shortest move requirement) uniformly rules out instances of long-movement, both grammatical and ungrammatical.

In this paper, we continue our exploration, now of some year's duration, of a particular "compositional theory of phrase structure", one related to Chomsky's recent work as he himself notes, but making use of different operations from those he employs (cf. Kroch and Joshi 1985, 1987, Kroch and Santorini 1991, Frank 1992, Frank and Kroch 1994, *inter alia*). In particular, we exploit the combinatorial operations of Tree Adjoining Grammar (TAG), namely *adjoining* and *substitution*. This system avoids the unnecessary distinction between Chomsky's two modes of phrase structure composition, GT and adjunction, as to which constraints apply to which mode of phrase structure composition. In fact, the use of the TAG operations enables us to eliminate completely the requirement that certain applications of generalized transformations extend their target. Instead, the empirical consequences of this stipulation follow from the nature of the TAG adjoining operation. We will see how the operation of adjoining provides a new perspective on the interaction of phrase structure composition and movement transformations: dependencies which have traditionally been taken to involve iterated applications of move-α over unbounded domains will be shown to necessitate only local transformational movement, coupled with the machinery of phrase structure composition. Additionally, our proposal permits us to resolve the tension between two conceptions of economy of derivation, shortest links or fewest operations, which arises in cases of successive cyclic movement. We do this by giving content to the operation Form-Chain as a primitive operation which maps elementary phrase structure objects onto elementary objects. Next we show that the TAG interpretation of generalized transformations allows us to express dependencies created both by successive cyclic and by long movement. Finally, we demonstrate that in spite of the absence of unbounded movement and intermediate traces in our derivational model, we see that it nonetheless provides an understanding of certain connectivity effects and yields an elegant account of the scope possibilities that arise out of wh-movement.

1 The Minimalist Framework and Generalized Transformations

Before proceeding, it will be useful to review the proposal for phrase structure composition using generalized transformations developed in Chomsky (1993). Chomsky proposes to eliminate a D-structure level of representation, a level which provides a pure representation of thematic relations, $GF(\theta)$. Instead, during the course of a derivation, an LF phrase marker is constructed out of individual phrasal projections. In order to provide a PF representation, the operation SPELL-OUT

applies to the constructed phrase marker at some point in the derivation. This branching point is not intended to constitute a level of representation, and in particular cannot be the locus of application for what have been thus far assumed to be S-structure conditions on well-formedness.

In a derivation in this model, the lexicon interfaces with what Chomsky calls the "computational system" of syntax by providing (partial) projections which satisfy X-bar theory. Thus, a verb might be drawn from the lexicon as a V^0, V' or VP projection. Such partial projections are combined using one of two generalized transformation-like operations: *GT* and *adjunction*. The first of these operations, GT, is a binary substitution operation. Given two pieces of phrase structure σ and τ, GT targets some node n within τ. It then creates a further projection of n and attaches σ as the daughter of this projection so that the root of σ and n are sisters. As an example, suppose that we have drawn the verb *hit* from the lexicon as the V^0 projection shown in (4)a. Suppose also that we have some other previously constructed structure, say the DP shown in (4)b. Now, GT can apply by targeting the V^0 node of the projection of *hit* with the DP, creating a V' level of projection and attaching the DP as sister to V^0 and daughter to V'. The result is shown in (4)c.

(4) a. [$_V$ hit]
     b. [$_{DP}$ the ball]
     c. [$_{V'}$ [$_V$ hit] [$_{DP}$ the ball]]

Note that since GT always creates a further projection of the targeted node, this targeted node must be either an X^0 or X' level of projection, assuming a version of X-bar theory which allows only two bar levels. Thus, any structure σ which is attached through GT will be placed either in a specifier or complement position, rendering all applications of GT equivalent to substitution.

The second operation, adjunction, behaves similarly to GT. However, the parent node which is created upon attaching the structure σ to the targeted node is not a further projection of the targeted node, but is rather an additional segment of the category of that node.[2] To take an example, suppose that we have already derived the VP structure in (5)a. We might now draw the adverb *quickly* from the lexicon as a fully projected AdvP, as in (5)b, and target the VP node with this AdvP using adjunction. The resulting structure is shown in (5)c.

(5) a. [$_{VP}$ read a book]
     b. [$_{ADVP}$ quickly]
     c. [$_{VP}$ [$_{VP}$ read a book] [$_{ADVP}$ quickly]]

Derivations will proceed in general by drawing elements from the lexicon, combining them together using GT and adjunction to build complex phrase markers, and combining these complex phrase markers together to build still more complex phrase markers. Intermingled among these applications of GT and adjunction, Chomsky also allows the application of move-α. In this framework, move-α can be seen as the singular version of GT. That is, given a single piece of phrase structure τ, a node n within τ can be targeted with some piece of structure drawn from within τ itself. Consider, for instance, a derivation in which we have already derived the C' phrase marker shown in (6)a. We can now apply move-α by targeting the C' root node by the DP *what*, to produce the structure in (6)b.[3]

---

[2]Chomsky (1994) abandons inherent distinctions between X' and XP nodes, but rather derives the notion of maximal projection relationally: a maximal projection is a node which is not itself further projected. Nonetheless, this system of "bare phrase structure" preserves the formal difference between segments and categories so as to distinguish the manner of projection during adjunction and substitution. Consequently, the dichotomy of the two operations of GT and adjunction must be preserved.

[3]We can now observe that the operation of move-α is identical in character to the generalized transformations GT and adjunction in Chomsky's system, in the sense that they all apply together as derivational operations. One might wonder whether it is appropriate to collapse these two types of operation, especially in light of Fillmore's

(6)   a. $[_{C'}$ $[_{IP}$ Sally heard $[_{DP}$ what]]]
      b. $[_{CP}$ $[_{DP}$ what] $[_{C'}$ $[_{IP}$ Sally heard t_i]]]

This model of grammar has a number of interesting properties. First, the lack of a level of representation which is a pure expression of GF(θ) leads to the result that thematic relations are fully represented only at LF. Thus, LF must also be the level at which the analog of the theta criterion, presumably now subsumable under the Principle of Full Interpretation, is checked. Moreover, the absence of any syntactic level of representation other than LF obviates the need for the Projection Principle: there are no relations which need to be preserved across levels of representation.

2   Cyclicity and the Extension Requirement

Consider now examples like (7).

(7)   * Hugh$_i$ seemed [it is certain [t_i to like pizza]]

In Chomsky's system, cases like this can be derived as follows: using only iterated applications of the operation GT, the following structural representation is built:

(8) $[_{I'}$ seemed $[_{I'}$ is certain [Hugh to like pizza]]]

At this point in the derivation, we move the embedded subject *Hugh* from its [SPEC,IP] position directly to the matrix [SPEC,IP] position yielding:

(9) $[_{IP}$ Hugh$_i$ $[_{I'}$ seemed $[_{I'}$ is certain [t_i to like pizza]]]

Chomsky imposes a "shortest move" requirement on the application of move-α, which he in turn aims to derive from principles of economy. This requirement, which demands that the landing site of movement be the closest possible position in which movement could have stopped, imposes certain types of locality on grammatical movement.[4] Note that the movement in (9) does not violate this requirement, precisely because there is no [SPEC,IP] position intervening between the position of the trace and the landing site present at this stage in the derivation. Now, we invoke one more application of GT: targeting the intermediate I', we attach a DP projection of the expletive *it* as the specifier of this clause. The result is example (7).

Since this derivation has produced an ungrammatical sentence, an instance of "super raising", something has evidently gone wrong. Chomsky suggests that the problem with this derivation is a violation of cyclicity. We have performed an operation on the matrix clause, movement of the DP subject to its [SPEC,IP] position, and then returned to performing an operation on the clause which it embeds, attaching the expletive as its subject. To rule out this unwanted derivation, Chomsky formulates what seems to be a minimal complication to his system: applications of substitution, including GT and move-α, must target the highest node in a phrase marker that is being constructed. In other words, application of substitution must extend the targeted phrase marker upwards. We will henceforth refer to this requirement as the *extension requirement*. With the extension requirement in place, it is straightforward to observe that the derivation given above for (7) is impossible. Insertion of the expletive *it* is blocked since it takes place internal to the

---

(1963) insight concerning the dichotomy in ordering among generalized and singular transformations. The proposal we make in this paper imposes a sharp distinction between the two, allowing only generalized transformations as derivational operations. In section 5 we see how this difference is exploited in the explanation of locality effects.

[4]The necessity of such a locality requirement in Chomsky's system derives, we claim, from the fact that the fundamental objects which enter into the derivation, i.e., the partial projections of lexical elements, are too small to effectively localize movement. See section 5 below.

previously constructed phrase marker and would not extend the target. If we now try to insert the expletive *it* prior to performing the raising, so as to satisfy the extension requirement, we violate the shortest move requirement instead.

While the extension requirement is adequate for the case at hand, it brings with it a number of conceptual oddities. First of all, it introduces an otherwise unmotivated distinction between the two generalized transformations, since it must be taken to apply only to applications of substitution and not to applications of adjunction. Indeed, the system that Chomsky outlines crucially exploits the idea that adjunction but not substitution can take place acyclically, to deal with certain data concerning reconstruction. Even if another solution is found in this case, applications of head movement by adjunction will necessarily violate the extension requirement. If a head H is to adjoin to a higher head J, then under usual assumptions concerning head movement, H must be the head of the complement of J. In order for J to have a structurally represented complement, though, it must project to at least a single bar level. But, now if H adjoins to J, it does not extend the phrase marker since this would require targeting the maximal projection of J rather than J itself.

Apart from this distinction between applications of substitution and of adjunction, the extension requirement also distinguishes among subclasses of substitutions. In particular, only substitutions which take place prior to SPELL-OUT are subject to the extension requirement. While one might view this as an accidental property of Chomsky's system, assumed only for the purposes of "concreteness", the distinction actually proves crucial. If all instances of structural case are checked via spec–head agreement with some AGR projection, then any language which does not require that its structural case be checked by SPELL-OUT will necessarily use non-extending applications of GT. Consider the checking of accusative case in English. It is uncontroversial to assume that English main verbs do not raise as high as T overtly. Since DP objects follow verbs in English, this tells us that an accusative case marked object cannot have moved to [SPEC,AGRP$_O$] prior to SPELL-OUT. Thus, the representation of a transitive sentence at the point of SPELL-OUT will look as follows (possible movement of V to AGR$_O$ aside):

(10)   [$_{\text{AGRP}_S}$ Lucy AGR$_S$ [$_{\text{TP}}$ T [$_{\text{AGR}_O}{}'$ AGR$_O$ [$_{\text{VP}}$ adores lettuce]]]]

Now, when the DP moves to [SPEC,AGRP$_O$] after SPELL-OUT it will target a node AGR$_O{}'$ which is not the highest in the phrase marker, and therefore will violate the extension requirement.
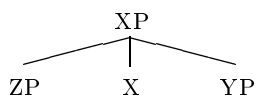
It would of course be desirable if we could eliminate such prima facie odd stipulations as constraints on derivations. Indeed, given the possibility of deriving locality constraints directly from a compositional model of phrase structure mentioned above, it would seem desirable to eliminate entirely constraints on locality such as the shortest move and extension requirements so much as possible. In the next section, we will consider a composition system of phrase structure which does just this.

3   An Alternative Theory of Generalized Transformations: Tree Adjoining Grammar

The TAG formalism (Joshi, Levy and Takahashi 1975, Joshi 1985, Kroch and Joshi 1985) was developed some twenty years ago as a mathematically restrictive formulation of phrase structure composition, inspired in part by Chomsky's earlier work on generalized transformations. As such, it provides an alternative to Chomsky's (1992) proposals. In TAG, structural representations are built out of pieces of phrase structure, called *elementary trees*, which are taken as atomic by the formal system. These trees can be combined using one of two operations. The first of these,

substitution, plays a role somewhat similar to that of Chomsky's GT. Given the structure in (11)a, we can substitute the structure in (11)b at the YP node along the frontier yielding the structure in (11)c.

(11) a.

$$
\begin{array}{ccc}
& \text{XP} & \\
& | & \\
\text{ZP} & \text{X} & \text{YP}
\end{array}
$$

   b.

$$
\begin{array}{ccc}
& \text{YP} & \\
& | & \\
\text{ZP} & \text{Y} & \text{WP}
\end{array}
$$

   c.

$$
\begin{array}{ccc}
& \text{XP} & \\
& | & \\
\text{ZP} & \text{X} & \text{YP} \\
& & | \\
& \text{ZP} \quad \text{Y} \quad \text{WP}
\end{array}
$$
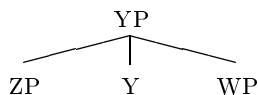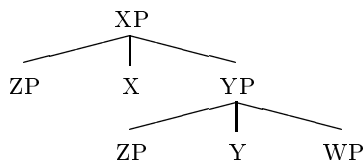
Like GT, the substitution operation can be used to insert XPs into the argument positions of syntactic predicates. Consider the tree in (12)a, whose phrase structural assumptions are similar to those of Stowell (1981) and Chomsky (1986). If we treat this as an elementary tree, we can use substitution to insert the DP elementary tree in (12)b into the object position of the verb in (12)a, yielding the structure in (12)c.

(12) a.

```
            CP
           /  \
          C    IP
              /  \
            DP    I'
                 /  \
                I    VP
                    /  \
                   V    DP
                   |
                 read
```

   b.

```
         DP
        /  \
       D    NP
       |    |
      the  book
```

c.

```
                        CP
                      /    \
                    C       IP
                          /    \
                        DP      I'
                              /    \
                            I       VP
                                  /    \
                                V        DP
                                |      /    \
                              read    D      NP
                                      |      |
                                     the    book
```
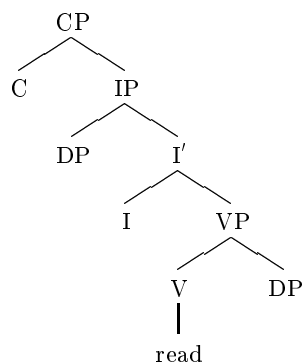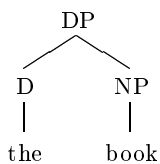
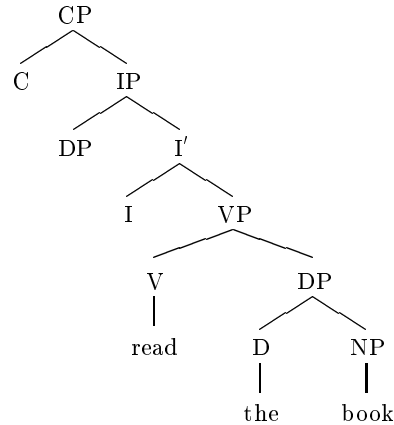Note that this operation is distinct from GT in that the site of substitution, in the above example DP, must be present in the elementary tree into which the substitution occurs. In contrast, when GT applies, the site which is targeted is augmented with an additional level of projection to accommodate the attachment of the new structure. This operation, then, violates the formulation of cyclicity imposed by Chomsky's extension requirement since it adds structure to an internal node.
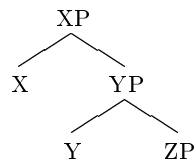
The second TAG operation, that of *adjoining*, has no real analog in Chomsky's system. Where substitution permits structure to be added only along the frontier of an elementary tree, i.e. along the edges, adjoining allows the insertion of a piece of phrase structure inside the body of another elementary tree. To do this, adjoining requires the use of an elementary object having a special form: one which is recursive in that its root is categorially identical to one of the nodes along its frontier. Elementary trees having this property we will call *auxiliary trees*, and the recursive node on the frontier we will call the *foot node* of the auxiliary tree. Elementary trees which are not auxiliary trees are sometimes called *initial trees*. An example of an auxiliary tree, recursive on YP, is given in (13).

(13)

```
            YP
          /    \
        Y       ZP
              /    \
            Z       YP
```
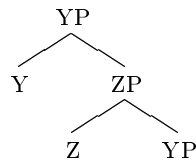
Now, given an auxiliary tree A, recursive on a node YP, the adjoining operation can apply to some other elementary tree T containing a node labeled YP by removing the subtree of T dominated by YP leaving behind just a copy of the YP node at the root of this subtree, attaching the root A to this YP node, and reattaching the subtree of T to the foot node of A. The result of adjoining (13) into the elementary tree in (14)a at the YP node, then, is given in (14)b.

(14) a.

```
            XP
          /    \
        X       YP
              /    \
            Y       ZP
```

b.

```
              XP
            /    \
          X       YP
                /    \
              Y       ZP
                    /    \
                  Z       YP
                        /    \
                      Y       ZP
```
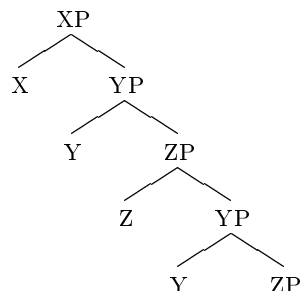
The recursive character of auxiliary trees provides us with another view of the adjoining operation: it is a domination-preserving expansion of a single node in a piece of phrase structure into a larger structure. If the inserted structure did not have identically labeled root and foot nodes as in an auxiliary tree, then either the original parent of the adjoining site would come to dominate a node of a different label, or the label of the root of the subtree excised during the adjoining operation would not be preserved.
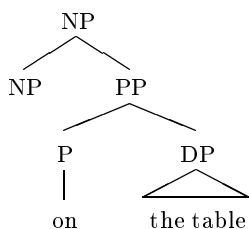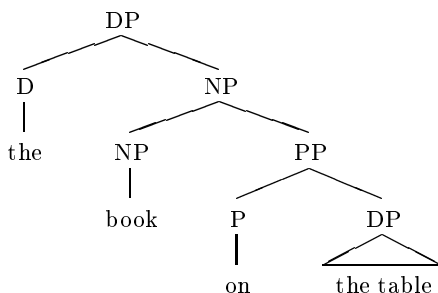
Note that in case the root node of an auxiliary tree directly dominates the foot node, the adjoining operation functions in a fashion similar to Chomsky's adjunction operation. The recursion in the auxiliary tree serves only to introduce an additional segment of a given structural node. Thus, to attach a prepositional phrase to an NP it modifies, we might make use of the following auxiliary tree:

(15)

```
              NP
            /    \
          NP      PP
                /    \
              P       DP
              |      /  \
             on   the table
```

Application of the adjoining operation using this auxiliary to the NP node in (12)b yields the following derived structure:

(16)

```
              DP
            /    \
          D       NP
          |      /    \
         the   NP      PP
               |     /    \
             book   P      DP
                    |     /  \
                   on  the table
```

This derived structure might then be substituted into a DP argument slot, as before.

Of course, auxiliary trees need not be limited to such restricted forms. Thus, by taking advantage of the fact that the node label of clausal complements is identical with the root of an embedding clause, i.e., CP, we can utilize the adjoining operation to introduce instances of clausal complementation. An auxiliary tree representing a clause with the predicate *think* might look as follows:

(17)

```
              CP
            /    \
           C     IP
                /   \
              DP     I'
                    /  \
                   I    VP
                       /  \
                      V    CP
                      |
                    think
```
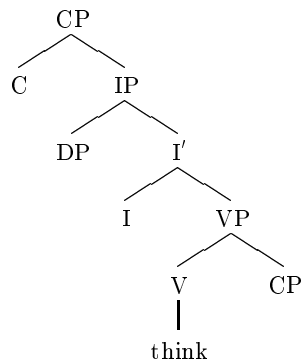
During the derivation of a sentence like (18), this CP auxiliary might adjoin to the root CP node of the elementary tree in (12)a to produce the structure in (19).

(18)   Lisa thinks that I read the book.

(19)

```
              CP
            /    \
           C     IP
                /   \
              DP     I'
                    /  \
                   I    VP
                       /   \
                      V     CP
                      |    /  \
                   think  C    IP
                             /   \
                           DP     I'
                                 /  \
                                I    VP
                                    /  \
                                   V    DP
                                   |
                                 read
```
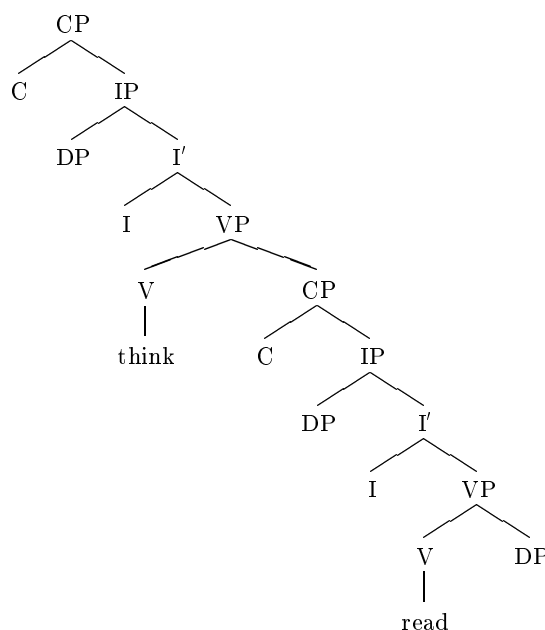
When adjoining takes place at the root of an elementary tree, the resultant structure can also be produced using substitution. In this case, the subordinate clause CP elementary tree could be substituted into the matrix elementary tree at the CP frontier node. A derived structure, therefore, does not unambiguously determine a sequence of derivational steps.[5] It is important for us, therefore, to keep separate the concept of derived structure from that of derivation structure, a representation of the history of a derivation. The notion of derivation structure has a long history within generative grammar going back to the T-marker in the theory of Chomsky (1955, 1957). However, such structures have largely been ignored in recent work, perhaps because in a system which builds structure one level of projection at a time, such as Chomsky's, derived and derivation structure are indistinguishable. Consequently, it is simply unclear whether it is the derived structure

---

[5]Note that under the system of feature percolation proposed for the TAG framework by Vijay-Shanker (1987), the manner in which features are percolated in the two derivations of (19) will be distinct, and the resultant structures will be distinct. For empirical evidence supporting distinctions in these two modes of feature percolation, see Frank and Kroch (1994).

or the derivation structure that is the object of grammatical interest.[6]

As the reader may have noted, we use the term *adjoining* uniquely to denote the TAG operation, while *adjunction* refers to Chomsky's generalized transformation.[7] In Chomsky's system, adjunction structures, i.e., those in which there are two segments of a single category, require the use of the adjunction operation. Indeed, all applications of the adjunction operation will result in adjunction structures. In a TAG, however, applications of adjoining need not give rise to adjunction structures. This is clear in derivations using auxiliary trees like (17) where the foot node is not the daughter of the root node. The converse question of whether all instances of adjunction structures in derived representations necessarily involve the adjoining operation, we leave open, though this identification seems attractive.

Even in the restricted cases where the foot node is the daughter of the root, so that adjoining and adjunction produce the same class of structures, adjoining nonetheless differs from Chomsky's adjunction operation and in much the same way that our substitution operation differs from GT. In the case of substitution, we saw that when a DP structure was to be inserted as complement to a verb, the DP node was already present in the elementary tree containing the verb. On the other hand, prior to the application of GT there is simply no DP node in the verb's syntactic projection. Comparing adjoining and adjunction, we see that the operation of adjunction adds a segment to the node at which the adjunction takes place. The recursive structure of an auxiliary tree, though, guarantees that no nodes are added in adjoining. The foot node of a "modification" auxiliary of the type in (15) will be identified with the node to which the adjoining takes place and the root node will become the additional segment of the projection to which an adjunction structure is being added. In both of these cases, then, we see that the TAG operations are "structure preserving" in a strong sense: local configurational relations, for example the relationship of sisterhood, which are determined at the level of the elementary tree never change during the course of a derivation.

With the formal machinery of TAG now laid out, the intuition behind a derivation in the system should be clear: a derivation is simply a sequence of combinations of elementary trees using the operations of adjoining and substitution. To make this intuition precise, we will make use of the notion of derivation structure mentioned above, the TAG analog of the Chomsky's (1955, 1957) T-marker. A TAG derivation structure is a tree in which each node n corresponds to some elementary tree τ and each daughter of n, call it m, represents another elementary tree τ' which is either adjoined or substituted into τ. In addition, the branch connecting a parent and daughter in such a derivation structure is annotated with the node in the parent elementary tree that is the locus of adjoining or substitution. Since the derivation structure is a tree, the daughters of n in the derivation structure, i.e., the m's above, may have daughters of their own, i.e., τ' may be the locus for the adjoining or substitution of other elementary trees. However, to guarantee that the derivation structure is coherently definable, it must be the case that the elementary tree τ' could independently have substituted or adjoined into the elementary tree τ. This restriction on possible derivations imposes a context-free character on these derivation structures, and indeed it is the case that TAG derivation structures, though not the derived structures, are strongly context-free.[8] Beyond this formal characterization of a derivation, we can impose other restrictions on derivations, for example that they must produce a structure rooted in CP, so as to guarantee the generation of sentences. Basically, however, the system operates freely once it is given a set of elementary trees

---

[6]In fact, we believe that it is the derivation structure which is paramount and derived structure plays no rule other than serving as the input to the phonological component. Pursuing this matter here would take us far afield.

[7]To complicate this terminological mare's nest, we should point out that in other work within the TAG framework, the term *adjunction* is sometimes used to refer to the TAG operation. We will, however, maintain the strict distinction in nomenclature between adjunction and adjoining, since it is crucial for the current discussion.

[8]For formal discussion of the definition of the derivation tree, see Vijay-Shanker (1987).

to operate on.

Thus far, we have said nothing about the character of the atomic objects, the elementary trees, which the TAG operations manipulate during the course of a derivation. Indeed, the TAG formalism says nothing whatsoever about this topic. TAG provides only the machinery to combine elementary trees once they are specified. It is thus the responsibility of a theory of grammar which exploits these formally defined operations to provide substantive conditions governing the well-formedness of elementary trees. Consequently, the adoption of the TAG operations does not implicate any particular theory of grammar any more than the adoption of the operation of GT does. Indeed there has been work in the TAG formalism from the differing perspectives of Lexicon Grammar (Abeillé 1988, Abeillé and Schabes 1989), Head-driven Phrase Structure Grammar (Kasper 1992), and Government-Binding Theory (Kroch 1989b, Kroch and Santorini 1991, Frank 1992, Hegarty 1993).

In our own work, we adopt a theory of grammar falling under the general rubric of "principles and parameters theory". Language universal principles, as instantiated by the values of parameters set for a given language, determine which elementary trees are licit in a derivation: all and only those trees which satisfy these well-formedness conditions. From this perspective, the set of elementary trees that can be used in the derivations for the grammatical sentences in a given language has no particular status. Such a set is an entirely derivative object and focus on it obscures the primacy of the underlying principles of grammar.

What then are the principles which govern the structure of elementary trees? Many answers to this question are possible, and it is too early in the development of our research program for us to offer a definitive answer here. Most of our research, however, has treated the elementary tree as a relatively large structure, corresponding in the canonical case to a simplex clause. For purposes of this paper, we will adopt the Condition on Elementary Tree Minimality (CETM) proposed in Frank (1992), which requires that each elementary tree correspond to the extended projection of a single lexical head, in the sense of Grimshaw (1991). The notion of extended projection is intended to provide an enlarged domain over which a lexical head imposes its syntactic requirements by including the syntactic projections of the functional elements which are associated with it. Thus, the extended projection of a verbal head will (optionally) include projections of I (or its constituent projections T and Agr, depending on one's view of clausal structure) and C, but may not extend either further up into a superordinate clausal domain or down below into the projection of a phrase appearing in a complement or specifier position within the extended projection. Likewise, an extended projection of N may include projections of D and perhaps P, but may not include structure which embeds this nominal extended projection, such as a VP.

The use of extended projections as the objects which are combined by generalized transformations during the course of the derivation extends Chomsky's proposal that the fundamental units of the derivation should be X-bar projections. See Frank (1992) for arguments that the enlargement of this minimal domain to extended projections is appropriate. There it is argued that this enlargement enables us to capture various dependencies as local constraints within the confines of the elementary tree.

In addition to the CETM, we will also require that a modified version of the projection principle hold for each elementary tree:[9]

(20)   Projection Principle (TAG version): If α is a maximal projection which appears along the frontier of an elementary tree τ, then α is a member of a chain which has a member that is selected in τ, through either theta role assignment or predication.

---

[9]The licensing of nodes via predication is intended to cover the licensing of foot nodes in "modification auxiliary trees" such as (15), and also the subject position of predicates such as *tough* and *easy*.

This version of the projection principle guarantees that all non-terminals which are present in an elementary tree, as a site of substitution or as a foot node, must be independently licensed. We also impose the theta criterion as a well-formedness condition on elementary trees.

(21)   Theta Criterion (TAG version): Given an elementary tree τ, for every theta role R assigned by a predicate in τ, there must be a unique chain in τ having a unique member to which R is assigned.

Note that the theta criterion and projection principles are statable as constraints on elementary trees exactly as a result of the property of local structure preservation that TAG derivations exhibit. If non-terminal nodes could not appear on the frontier of the atomic objects that are composed during the derivation, then we would clearly be unable to require that thematic structure be defined on such representations. Rather, we would need to delay until the point at which the phrase marker had been constructed to check, for example, whether an element had been assigned a theta role. This is exactly the state of affairs in system of Chomsky (1993). Instead, by including these nodes in the atomic objects, we can understand the inherently local nature of theta role assignment, and node licensing generally: such processes occur internal to the elementary trees. Indeed, one of the goals of our enterprise is to understand what the character of locality conditions is. One attractive possibility that the model of grammar we are exploring suggests is that all grammatical principles must be stated as well-formedness conditions on the atomic objects of the derivational system. In this view, once well-formed elementary trees are constructed, the compositional operations apply essentially blindly without further constraints on the derived representation.

Note that in both the theta criterion and projection principle we have referred to chains within an elementary tree τ. As we will discuss in section 5, we assume that a process of chain formation does take place in grammatical derivations, but is entirely separable from the process of phrase structure composition. Among other things, this guarantees that only so-called economical derivations will be licit without resorting to additional requirements such as shortest move.

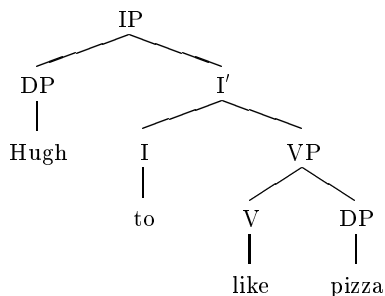## 4   Cyclicity without the Extension Requirement

With the basics of the TAG framework laid out, let us examine the fate within this framework of the problems for Chomsky's GT and adjunction posed by the super raising examples discussed in section 2. In the analyses given by Kroch and Joshi (1985) and Frank (1992), subject to subject "movement" is not accomplished through a movement operation at all, but rather is the result of the adjoining operation. Consider the derivation of the example in (22).

(22)   Hugh seemed to like pizza.

We represent the subordinate infinitival clause by the elementary tree in (23). As a result of the projection principle and theta criterion, all of the arguments to the predicate *like* are represented inside of this structure.[10]

---

[10]For the sake of readability, DP arguments are filled in these elementary trees. However, according to the CETM they could not be present in the same elementary tree as a verbal head. Thus, we will assume that they are inserted using the substitution operation during the course of the derivation, prior to the application of the adjoining of interest in this case.

(23)

```
              IP
          ___/  \___
        DP          I′
        |        __/  \__
       Hugh     I         VP
                |       __/  \__
                to     V         DP
                       |         |
                      like     pizza
```

The elementary tree representing the matrix raising predicate is given in (24).

(24)

```
           I′
         _/  \_
        I      VP
             _/  \_
            V      I′
            |
         seemed
```
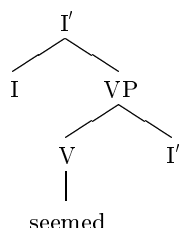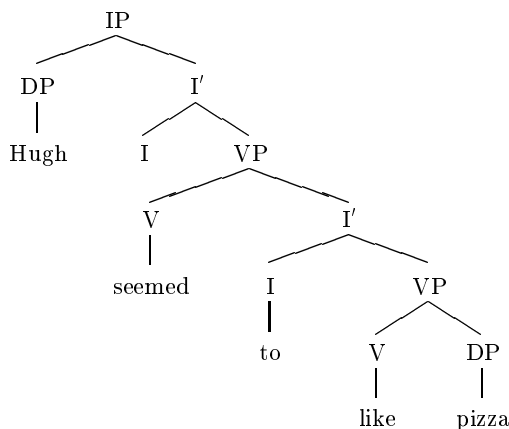
Note that this structure lacks a position for the matrix subject. This follows from the fact that the projection principle as stated in (20) applies to elementary trees. Such a subject position would be unlicensed in the matrix tree, and therefore may not appear. This lack of a [SPEC,IP] position coupled with the recursive structure of auxiliary trees forces us to allow predicates to take X′ level projections as complements. In particular, a head like *seem*, which does not license a subject position and therefore only tolerates an I′ extended projection as its elementary tree, must be allowed to take an I′ complement.

We can now apply the adjoining operation, inserting this auxiliary to the I′ node in the elementary tree in (23). The resultant structure is as in (25):

(25)

```
              IP
          ___/  \___
        DP          I′
        |        __/  \__
       Hugh     I         VP
                       __/  \__
                      V         I′
                      |       __/  \__
                   seemed    I         VP
                             |       __/  \__
                             to     V         DP
                                    |         |
                                   like     pizza
```

In this derivation, there is no application of a movement transformation to take the subject from one clause to the next, nor is there a trace left behind by the subject in its original position.[11]

---

[11] Of course, if we adopt the VP-internal subject hypothesis, then there remains the possibility of real transformational movement internal to the elementary tree from the VP-internal position to the [SPEC,IP] position. For an alternative TAG-based proposal in which movement does not obtain even in this case, see Hegarty (1993).

Rather, what has happened is that the originally local structural relation between the subject and its associated clause has been stretched during the adjoining operation. Note, though, that since the assignment of thematic roles is done at the level of the elementary trees, the absence of a trace has no detrimental effect on our ability to interpret the derived structure.[12],[13]

Now consider how we might try to derive the ungrammatical example in (7) in the TAG-based theory we are exploring. The representation of the subordinate infinitival clause must remain as in (23) since the structure was dictated by the combination of the CETM, projection principle and theta criterion. In order to "move" the subject *Hugh* to the front of this example, we will evidently need to adjoin in structure which expresses the lexical material *seemed it is certain* between *Hugh* and *to*, something like the structure in (26).

(26)

```
                    I'
                  /    \
                 I      VP
                      /    \
                     V      IP
                     |    /    \
                 seemed  DP     I'
                         |    /    \
                         it  I     AdjP
                             |    /    \
                            is  Adj    I'
                                 |
                              certain
```
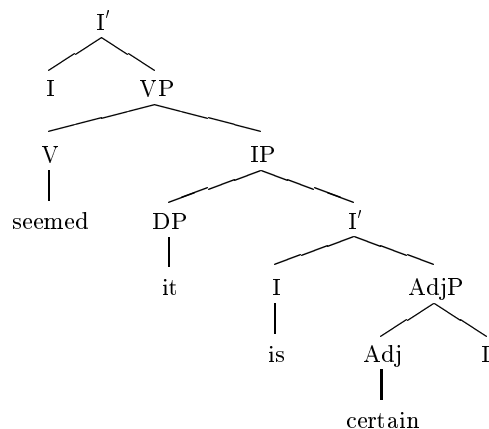
Under the conception of elementary trees as individual extended projections that the CETM provides, this structure cannot be a single elementary tree. The lexical predicates *seem* and *certain*

---

[12]Nor does the absence of transformational raising block the analysis of familiar cases of scope ambiguity such as (i):

(i) A unicorn seems to be in the garden.

In the elementary tree of the IP complement in (i) (that of *A unicorn to be in the garden*), the subject of the IP has scope over the I' at which adjoining takes place. This relationship can be taken to capture the wide scope interpretation of *a unicorn* within this elementary tree. Under the assumption that the subject starts out in a VP-internal subject position and then moves to the derived subject position within its own elementary tree, the VP-internal trace would capture the narrow scope reading. To implement this solution, we assume the augmentation of TAG with feature structure descriptions proposed by Vijay-Shanker (1987). Pursuing this analysis would require a rather extensive tangent to the main discussion, but no issues of principle arise.

[13]The analysis as given in the text does not block examples such as (i):

(i) * Hugh seems likes pizza.

There are a number of possible explanations for the ungrammaticality of this case. Perhaps the most straightforward would invoke case in some fashion: either *likes* (or more accurately the tensed Infl) is unable to assign its nominative case or *Hugh* is improperly assigned case twice. Alternatively, we might understand the problem with (i) as a problem related to Chomsky's (1993) notion of GREED: no adjoining is needed, so none may be performed. In order to express either of these solutions, we will need some way of imposing constraints on the applications of adjoining. A range of systems for integrating constraints into the TAG formalism have been proposed (Joshi 1985, Vijay-Shanker and Joshi 1985, Vijay-Shanker 1987). The GREED solution can be expressed simply by imposing what Vijay-Shanker and Joshi (1985) called a *null adjoining constraint* (which blocks the possibility of adjoining) at the I' node of the *Hugh likes pizza* elementary tree. Either of the case solutions can be expressed in terms of the proposal of Vijay-Shanker (1987) which augments the nodes of an elementary tree with feature bundles that must be compatible with those of other nodes which become local during adjoining. (See Frank (1992) for further discussion.) We will not pursue the question of which analysis is superior as it is largely orthogonal to our current concerns.

must head distinct extended projections, and so the structure which is to be adjoined in must be composed from distinct elementary trees. Suppose that the auxiliary tree representing the clause headed by *seemed* is as in (24). We must combine this structure with the elementary tree for the *certain* clause and then adjoin the result to the I′ node of the elementary tree in (23). However, the representation of the *certain* clause must be rooted in IP: this clause licenses the expletive subject *it*, presumably by some association with the clausal complement of *certain*. Thus, we cannot simply adjoin the *seem* I′ auxiliary tree to the root of a *certain* IP auxiliary, as shown in (27), even if we allow *certain* to take IP complements, since the node labels do not match.

(27)
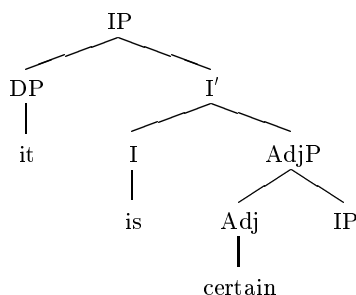
Also, there is no way to insert the expletive *it* into an I′ structure as [SPEC,IP] during the course of the derivation, since the adjoining operation can only introduce recursive structures.[14]

The only other way to derive the super-raising structure would be to combine the *seem* and *certain* clauses using substitution and then adjoin the result. That is, we would use elementary trees such as the following:

(28) a.

---
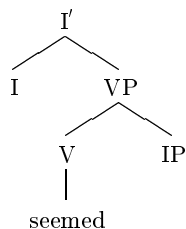
[14]As an anonymous reviewer points out, both our analysis and Chomsky's ignore the possibility suggested by Bennis (1986) that the expletive *it* in "non-raised" versions of raising sentences is generated within the complement domain of the raising predicate and is moved to the higher subject position. This idea could be adapted into our TAG framework by assuming following structure for the complement clause:

    (i) [$_{IP}$ it [$_{I'}$ I [$_{CP}$ …]]]

A derivation of a an example like (ii) would then be identical to that for simple raising cases like (22).

    (ii) It seems that Hugh likes pizza.

The reader may observe that this analysis will still block the generation of the superraising cases. Unfortunately, the structure in (i) violates the CETM and would represent a substantial departure from our current structural assumptions. Thus, we put aside this possibility for the present.

b.

```
                        IP
                   ____/  \____
                 DP            I′
                 |          __/  \__
                 it        I       AdjP
                           |      __/  \__
                           is   Adj       I′
                                 |
                              certain
```
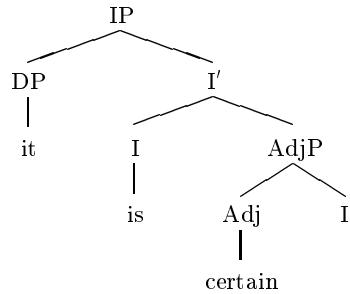
We could substitute the elementary tree in (28)b into the IP node on the frontier of (28)a yielding a derived structure like that in (26). The structure would have an I′ root node and an I′ node on the frontier, and we might try to adjoin this structure to the I′ node in (23) to derive the example (7). However, this derivation is impossible. TAG derivations require that the recursive pair of root and foot nodes of auxiliary trees must be present in the auxiliary tree prior to its entry into the derivation. To see why this must be so, recall that a TAG derivation structure is a tree each of whose nodes η represents an elementary tree τ, and where the children of η correspond to other elementary trees which have been adjoined or substituted into τ. In the derivation we are currently considering, the structure which is adjoined into the elementary tree in (23) is not an elementary tree, but is instead a derived structure. The question, then, is how the derivation structure looks in this case. Do we attach the node corresponding to the elementary tree in (28)a as daughter of the node corresponding for (23) or attach the node for the elementary tree in (28)b? Neither of these choices is adequate since neither of these elementary trees could by itself adjoin into (23). However, the derivation structure must be a tree, and hence have a unique root. Therefore, no well-formed derivation structure can be constructed in this case, and the derivation is blocked.

We are not blocked from performing adjoinings into an auxiliary tree α prior to the adjoining of α into some other elementary tree, so long as the root and foot of the auxiliary α are preserved. Indeed, such a sequence of operations would be needed in the derivation of examples involving iterated subject to subject raising, like (29).

(29)   Hugh seemed [$_{I′}$ to be certain [$_{I′}$ to like pizza]]

Here we use two I′ auxiliary trees for each of the raising predicates, adjoining the *seem* tree first to the root of the *certain* tree, and then adjoining the result into the *like* tree. Further, we can construct a well-formed derivation structure of this sequence: the node for the *like* tree dominates the node for the *certain* tree which itself dominates the node for the *seemed* tree.

We see, then, that the principles of elementary tree well-formedness prevent us from forming a derived auxiliary tree of the appropriate size to derive the ungrammatical (7). The auxiliary tree must be recursive on I′, but the embedded clause must be a full IP. This, coupled with the restrictive TAG machinery prevents us from deriving instances of super raising. Thus, the effects of Chomsky's extension constraint are captured without explicit stipulation.

Raising is but one example of a construction in which an element is moved successively from one clause to the next, in each clause occupying an identical structural position. Using the adjoining operation in a similar manner, we can account for all instances of putative successive cyclic movement without transformational movement, but rather through the adjoining in of intervening material. Thus, to handle cases of successive cyclic wh-movement, we will, following Kroch (1987, 1989b) and Frank (1992), use a derivation similar to the one given above for raising, but where the adjoining stretches the relation between an element in [SPEC,CP] position and its base generated

clause.[15] So long as elementary trees are constrained to have at most one landing site of a given type, this type of derivation of successive cyclic constructions will have the same empirical coverage as one which uses substitution with the extension constraint. Since use of adjoining allows us to eliminate the extension constraint, we are free to eliminate the distinction between substitution and adjunction that Chomsky posited. Thus, neither adjunct structures nor substitution structures, recast into their TAG guises, will be required to extend their targets.

5   The Local Nature of Form-Chain

In addition to the extension requirement, we have now eliminated the necessity for the derivational operation of move-α in cases of unbounded dependencies usually accomplished through iterated movement through identical positions in successively higher clauses. Continuing in this spirit, let us investigate the hypothesis that move-α is completely absent as a derivational operation. In order to pursue this we must consider the residue of movement not thus far replaced by applications of the adjoining operation: movement which gives rise to "hybrid" dependencies, instances of movement from an argument position to a non-argument position.

In the case of wh-movement, such hybrid dependencies abound. The fronted element must somehow get from its base position to the [SPEC,CP] position in which it takes scope. As is especially clear in cases of object extraction, there is no way that an instance of adjoining could accomplish this dislocation. Interestingly, under the current conception these hybrid dependencies can be localized entirely within an elementary tree. In the case of wh-movement, the hybrid dependency entered into by the wh-element is between its base position and the [SPEC,CP] position of its own clause. The remaining dependencies are all iterated movements through positions of the same type, i.e., [SPEC,CP] positions of higher clauses. These can all be handled using applications of adjoining as discussed in the previous section.

Now, to account for these local hybrid dependencies, we can adopt Chomsky's idea that there is a fundamental "Form-Chain" operation, and for the purposes of this paper we will do so.[16] However, we will differ from him in limiting the operation to applying within a single elementary object, i.e., an elementary tree, mapping that tree to another elementary tree. For present purposes, the following suffices:

(30)   Form-Chain: Within an elementary tree, coindex an element in SPEC with a c-commanded θ-position.[17]

To derive an instance of successive cyclic wh-movement, we can apply the Form-Chain operation to the elementary tree in (31)a to produce the elementary tree in (31)b.[18]
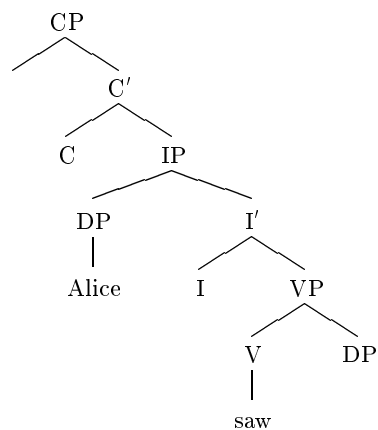
---

[15]This case will be discussed in more detail in the next section.

[16]Other possibilities exist for creating filler–gap dependencies and have been discussed in other TAG-based linguistic work (cf. Kroch 1987, 1989b, Frank 1992, Hegarty 1993).
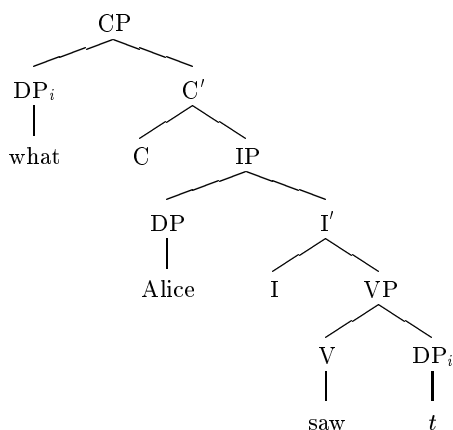
[17]Note that if we were to alter our assumptions regarding phrase structure so that an A-chain containing a subject could include traces or copies in [SPEC,VP], [SPEC,TP] and [SPEC,AGRP$_S$] positions, we might want to extend this formulation of Form-Chain to allow more than one chain link to be generated in a single application (assuming that economy considerations dictate that the formation of such a chain is comparable in "cost" to A-chains having only a single link). Nonetheless, the number of chain links which Form-Chain could create in such an extension would remain bounded by some small number, as opposed to the potentially unbounded number of links created by the Form-Chain operation as defined by Chomsky.

[18]Note again that the DP nodes have been filled in in these elementary trees for readability. The CETM will, however, require that such DPs be separate elementary trees since they constitute independent elementary trees. The coindexation that results from the application of Form-Chain, though, will be represented entirely via the non-terminals of the elementary tree, present as a result of the theta criterion and projection principle.
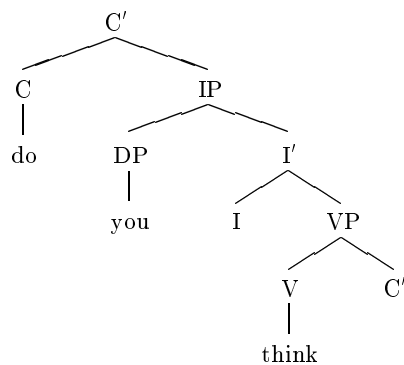
(31) a.

```
              CP
               |
               C'
              /  \
             C    IP
                 /  \
               DP    I'
                |   /  \
              Alice I   VP
                       /  \
                      V    DP
                      |
                     saw
```

b.

```
              CP
             /    \
          DP_i     C'
           |      /  \
         what    C    IP
                     /  \
                   DP    I'
                    |   /  \
                  Alice I   VP
                           /  \
                          V    DP_i
                          |     |
                         saw    t
```

We can then adjoin an auxiliary tree like the one in (32)a, to the C$'$ node in (31)b, yielding the structure in (32)b.[19]

(32) a.

```
              C'
             /  \
            C    IP
            |   /  \
           do  DP    I'
                |   /  \
              you  I   VP
                       /  \
                      V    C'
                      |
                    think
```

---

[19]For the structure in (32)a, we assume that *do* does not head its own VP projection but rather that it is generated within the projection of some functional head. Thus, a single elementary tree may host *do* as well as another lexical verb.

b.

```
                        CP
                   ┌────┴────┐
                  DP_i       C'
                   │      ┌───┴───┐
                 what     C       IP
                          │    ┌──┴──┐
                          do  DP     I'
                               │  ┌──┴──┐
                             you  I     VP
                                     ┌──┴──┐
                                     V     C'
                                     │  ┌──┴──┐
                                   think C    IP
                                         │ ┌──┴──┐
                                        DP      I'
                                         │   ┌──┴──┐
                                       Alice  I    VP
                                                ┌──┴──┐
                                                V    DP_i
                                                │     │
                                               saw    t
```
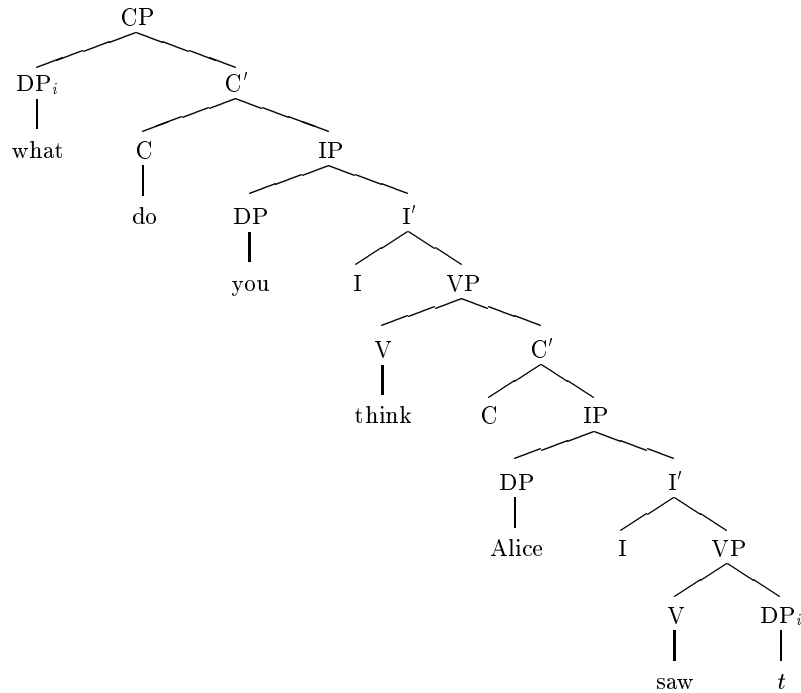
As in the case of raising, we have here a predicate that takes a non-maximal projection as its complement. In the raising case, the complement is an I′ (cf. the auxiliary tree in (24)) while here we see that the verb *think* can select for a C′. Again, the fact that this auxiliary tree root node must be C′ follows from the requirement that auxiliary trees be recursive structures. Frank (1992) suggests that the possibility of selecting a non-maximally projected clause, a C′, is exactly what distinguishes the class of bridge verbs from the class of non-bridge verbs. Bridge verbs may select for a clausal complement with a defective complementizer system, i.e., a C′, and are therefore able to head an auxiliary tree like the one in (32)a, and thereby enter into a derivation like the one here. Non-bridge verbs and verbs which take indirect questions, on the other hand, select a full CP, and for reasons essentially analogous to those considered above for the super raising case cannot take part in such a derivation.[20] Acceptable instances of extraction from non-bridge verbs and wh-islands employ a different sort of derivation, one involving multi-component auxiliary tree sets. See section 7 for some discussion, and also Kroch (1989b) and Frank (1992) for further details.

The Form-Chain operation given in (30) differs from Chomsky's operation in that only one chain link, and not an unbounded number of them, may be created in a single application. We believe, however, that this operation is nonetheless equivalent in its role in our theory to Form-Chain (as opposed to move-α) in Chomsky's theory since it is sufficient to generate in one application all chains

---

[20]Under Chomsky's (1994) system in which the notion of maximal projection is defined relationally, the distinction we exploit between maximally projected CPs and non-maximally projected C′s is not directly recoverable: either CP or C′ may be maximal, so long as it is not further projected. (Note that this is not true under Muysken's (1982) proposal where a C′ projection could be characterized as [+projected,-maximal]). Further, if we require for a well-formed auxiliary tree only that both root and foot nodes be maximal projections in the relational sense, our analysis of locality effects collapses. Observe, however, that even under Chomsky's proposal, what we call CP and C′ projections must nonetheless remain distinct, at least for languages like English. Assuming that projections of C^0 may contain only a single specifier, some property, presumably relating to feature content, must distinguish C′ from CP so that only the former tolerates the attachment of an element as specifier. By requiring that root and foot nodes be identical in this property, whatever it turns out to be, we can reconstruct our accounts of locality within Chomsky's bare phrase structure view.

that occur. This follows from the fact that, under our assumptions, any single clausal domain will contain only a single link of any non-trivial chain, assuming instances of head movement involve adjunction of successively larger complexes rather than substitution (cf. Chomsky 1992, 1994) and that Wh-movement proceeds directly to [SPEC,CP] without an intermediate A′ landing sites. Since we generate unbounded dependencies by stretching such local dependencies through adjoining, a chain with more than one link will simply never arise.[21]

This property of our version of Form-Chain yields the benefit of resolving the tension in Chomsky's system between two distinct notions of economy of derivation. Chomsky points out the problem posed by successive cyclic movement: if we are trying to minimize the number of operations in a derivation, we will have longer movements; if we try to minimize the length of chain links, more operations will result. His resolution of this conflict is to replace the transformational operation move-α by Form-Chain. However, as we noted above, Form-Chain must be defined so as to generate arbitrarily many instances of iterated successive cyclic movement, all in one step with unit cost. That is, given the representation in (33)a, a single application of Form-Chain suffices to produce the representation in (33)b.

(33) a. [$_{CP}$ e do you think [$_{CP}$ e that Bill said [$_{CP}$ e that Alice saw what]]]
      b. [$_{CP}$ what$_i$ do you think [$_{CP}$ t_i'' that Bill said [$_{CP}$ t_i' that Alice saw t_i]]]

In derivations using the TAG adjoining operation, this bit of creative accounting is no longer necessary. Form-Chain is constrained to apply entirely locally, and creates only a single operator–variable dependency in its application. Thus, Form-Chain applies in cases of successive cyclic wh-movement just as it does in cases of local, i.e., intra-clausal, wh-movement: it moves the wh-element to the local [SPEC,CP] position. Consequently, it introduces the same degree of cost in each case. In the framework we are adopting, the remaining cost associated with the derivation of an example like (33) is associated with the application of the adjoining operation to produce the layers of clausal embedding. Thus, we will need first to adjoin the auxiliary tree in (32)a to the root node of the auxiliary tree in (34)a producing the derived structure in (34)b, and then adjoin this to the C′ node of the elementary tree in (31)b as before.

(34) a.

---

[21]See note 17.

b.
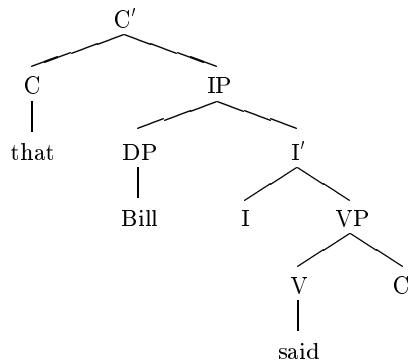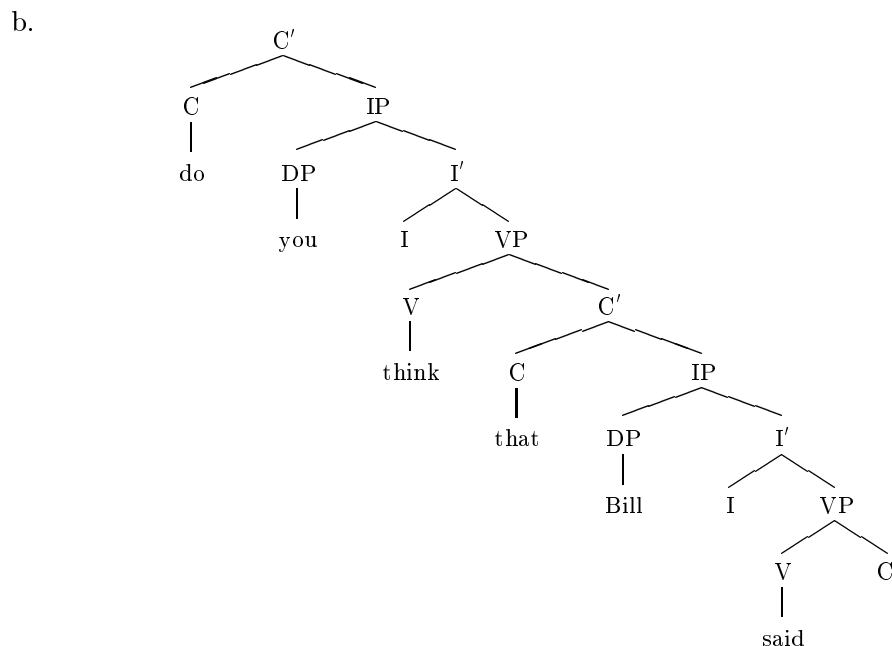
It is clear that any sentences with the same amount of clausal embedding will require the same degree of derivational complexity, whether there is a fronted wh-element or not. The embedded clauses must simply be integrated into the output structure using the system of generalized transformations that the grammar allows. This complexity is unaffected by the fact that a local dependency is being stretched during the course of an adjoining.[22]

6 Eliminating Shortest Move

Note that we have not imposed restrictions on movement or chain formation of the sort suggested by Chomsky, e.g., the shortest move requirement, the minimal link condition, or the like. Since our version of Form-Chain operates only over individual elementary tree domains, this is perhaps unsurprising: the examples which motivated Chomsky (1993) to invoke such conditions were instances of inter-clausal dependencies exhibiting relativized minimality type effects (e.g., super raising, wh-island violations, etc.).[23] Under our analysis, these effects are derived from the character of the adjoining operation coupled with our restrictive notion of well-formed auxiliary trees. We take this to be a positive step. By conflating the locality of long distance (or successive cyclic) dependencies with that of clause-bounded phenomena, such as movement of the subject and object to case checking positions, Chomsky (1993) is forced to complicate considerably the idea of shortest move, a notion whose elegant conceptual basis would have led us to expect a simpler realization. In particular, Chomsky defines a notion of minimal domain such that if two elements fall within the same minimal domain, they count as equidistant for the purposes of shortest move. In the absence of head movement, a minimal domain includes precisely the inhabitants of a single XP projection. Thus, movement which is shortest corresponds directly to our intuitive notion: each specifier is in

---

[22]However, see Frank (1992) where it is claimed that differences in the *processing* complexity of combinations of elementary trees using substitution and using adjoining leads to differences in the complexity of acquisition of a range of constructions.

[23]An exception to this statement is the locality effects inherent in head movement. However, see below for further discussion of the ways in which inter-clausal head movement must be distinguished from these cases of inter-clausal dependencies.

a separate minimal domain, so no two are equidistant, and consequently movement to a specifier position must truly be to the closest one. When a head X adjoins to another one Y, however, the minimal domain is extended to include the occupants of both XP and YP.

Chomsky motivates this complication of shortest move through an intra-clausal dependency, that involved in the raising of an object DP for case purposes. Under Chomsky's assumptions, the object raises past the closest specifier, i.e., [SPEC,VP], to the position in which its accusative case is checked, i.e., [SPEC,AGRP$_O$], in apparent violation of the shortest move requirement. If, however, the verb adjoins to Agr$_O$, the minimal domain of the verb is extended to include both [SPEC,VP] and [SPEC,AGRP$_O$], rendering them equidistant. Thus, the object may move freely to the higher position without incurring any economy violation. Jonas and Bobaljik (1993) present what looks like strong empirical support for this domain extension machinery in their analysis of Icelandic object shift. They observe that the restriction of object shift (which they take to be overt movement to [SPEC,AGRP$_O$]) to clauses in which the verb has raised, follows directly from the need to extend the appropriate minimal domain.

Note that even if Chomsky is correct that the additional complication in the notion shortest move is necessary, it is interesting to observe that it has effects only in intra-clausal and not in inter-clausal dependencies. To our knowledge, there have been no cases of inter-clausal movement cited anywhere in the literature where the notion of shortest move must be extended beyond the the intuitive one with the mechanisms of domain extension. This is not to say that such a case could not exist. In particular, there is a way in which we could use this machinery to generate instances of extractions from wh-islands like (35).

(35)   * Who$_i$ do you wonder what$_j$ t_i ate t_j?

To do this, suppose that the derivation has thus far produced the following structure:

(36)   [$_{CP}$ what$_j$ [$_{C'}$ C [$_{IP}$ who ate t_j]]]

Now, we select another C^0 head from the lexicon and apply GT targeting this head with the previously derived CP.

(37)   [$_{C'}$ C [$_{CP}$ what$_j$ [$_{C'}$ C [$_{IP}$ who ate t_j]]]]

If, at this point, we try to raise the subject wh-element *who* the specifier of this higher CP we would violate shortest move. However, if we adjoin the lower C^0 to the higher C^0, this will extend the minimal domain for the subject in [SPEC,IP] to include both [SPEC,CP] positions. Thus, we can produce the following structure without incurring any violations:

(38)   [$_{CP}$ who$_i$ [$_{C'}$ [$_C$ C$_k$ C] [$_{CP}$ what$_j$ [$_{C'}$ t_k [$_{IP}$ t_i ate t_j]]]]]

All that remains to produce (35) is to extend this structure with the matrix clause and perform the uncontroversial movement of *who* into the matrix [SPEC,CP]. The structure in (39) results:

(39)   [$_{CP}$ who$_i$ do you wonder [$_{CP}$ t_i [$_{C'}$ [$_C$ C$_k$ C] [$_{CP}$ what$_j$ [$_{C'}$ t_k [$_{IP}$ t_i ate t_j]]]]]]

There are a number of ways in which this derivation might be blocked setting aside the shortest move requirement. We might, for instance, appeal to the fact that indirect question complements do not tolerate CP recursion (Iatridou 1991, Iatridou and Kroch 1992). Nonetheless, given the existence of the mechanism for redefining what counts as the closest position, we should expect that it is at least sometimes exploited in inter-clausal dependencies. We take the fact that it is not to indicate that Chomsky's conflation of inter-clausal and intra-clausal locality to be incorrect. Rather, we suggest that what Chomsky described as the shortest move requirement is correct in its simple and conceptually elegant guise, but holds only for cases of inter-clausal movement and is not relevant for intra-clausal dependencies.

With Chomsky's system, this division is surprising and perhaps even unexpressible. However, given that division within our TAG framework between dependencies derived via Form-Chain and those derived with adjoining, it is entirely comprehensible: the effects of the shortest move requirement are entirely derivative from the properties of adjoining. The shortest move requirement itself has no theoretical status and is simply a descriptive statement. Consequently, we predict that Form-Chain, the operation which creates intra-clausal dependencies, will not show signs of obeying a shortest move requirement in any form. This prediction is confirmed from cases of object shift, i.e., movement to [SPEC,AGRP$_O$], which, in contrast to the situation in Icelandic, are not tied to verb raising. Two such cases are specific DP scrambling in Dutch (Koopman 1995) and object preposing in Mandarin Chinese (Ernst and Wang 1995).[24] Further support is provided by the locality of head movement. Under Chomsky's system, this intra-clausal dependency should obey the strictest version of the shortest move requirement, i.e., one without domain extension, since head movement cannot affect the relative closeness of head positions (since they do not belong to a minimal domain, cf. Chomsky 1993). The phenomenon of long head movement (Rivero 1991) seen in the Balkan and Slavic languages, where participles and infinitives may precede auxiliaries in main clauses, provides evidence that this is not true. Rivero analyzes this as verb movement across an intervening head, occupied by the auxiliary, to C^0. As just noted, Chomsky's proposal of minimal domains does not extend to this case (but see Rivero 1993 for a non-shortest-move based alternative within the Chomsky's framework). Strikingly, however, long head movement appears to be uniformly clause bounded, and therefore may be properly analyzed as an instance of Form-Chain applying within an elementary tree domain. Once again, this points to the fact that intra-clausal locality must be distinguished from inter-clausal locality in precisely the manner we have suggested.

7   Long Movement

Thus far, the long distance dependencies we have considered are those ordinarily generated using successive cyclic movement. As is well known, there is another set of cases, commonly called "long movement", in which licit extraction takes place out of an island or skips some intermediate landing site. Extraction of complements from wh-islands, as in (40), is one example.

(40)   Which car did Sally wonder how to fix?

Chomsky's extension constraint on substitutions – coupled with the requirement that movement always be to the closest landing site – uniformly blocks the derivation of such examples. If the DP *which car* moves first to the matrix [SPEC,CP] skipping past the intermediate C projection, subsequent movement of the wh-element *how* to the intermediate [SPEC,CP] will violate the extension requirement. If instead the element *how* is inserted first, then fronting of *which car* will be impossible without violating the shortest move requirement. Indeed, Chomsky takes this as a virtue since this conspiracy of constraints blocks the derivation of illicit examples like (41):

(41)   * How did Sally wonder which car to fix?

However, if we are to make progress in understanding the difference between the cases in (40) and (41), it seems unlikely that the ungrammaticality of (41) should be derived from such "deep" and invariant principles of the grammar.

Within the TAG conception of long-distance dependencies, this particular problem for generating examples like (40) simply does not arise since it does not include the extension requirement. One might wonder, however, whether the analysis of wh-movement in TAG might not suffer from its

---

[24]Of course, this means that we must reanalyze the linkage between verb raising and object shift in Icelandic.

own problems. Let us consider how a TAG-based analysis of long movement cases might look. Following the assumptions we have made, deriving example (40) requires applying Form-Chain to some elementary structure such that we can insert the element *which car* into the [SPEC,CP] position and also posit a c-commanded coindexed trace. This elementary structure to which Form-Chain applies cannot be the embedded clause, however: its presumably unique [SPEC,CP] position must be filled by the wh-element *how*. The only other possibility is that Form-Chain applies to the elementary tree representing the matrix clause. However, a problem arises in this application of Form-Chain as well. According to the formulation given above in (30), we must not only place an element in the specifier position, but must also posit an empty category in some c-commanded position. While there is an available [SPEC,CP] position this time, there is no position in which we can place a DP empty category without violating requirements of the projection principle and theta theory: there will be an argument chain which lacks a position that is assigned a theta role. Thus, we are stuck.
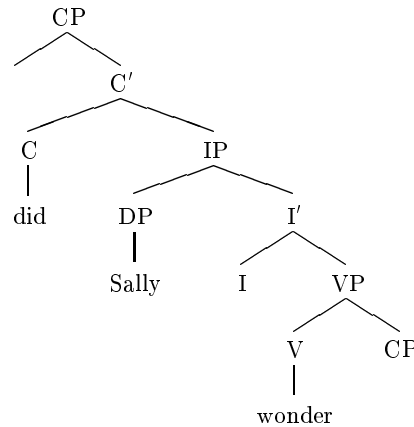
It is curious that Chomsky's system, with its shortest move and extension requirements, and ours, with its seemingly quite different definition of adjoining and CETM, run into such similar problems in the analysis of long movement. We think that this is suggestive of a deep similarity between the two, though we leave a precise explication of this convergence for future work. What we will instead do in the remainder of this section, is consider possible extensions or relaxations of our current assumptions which will allow us to account for the contrast between (40) and (41). We leave open the degree to which our suggestions can be transported into Chomsky's framework.

As just mentioned, instances of long movement pose the following dilemma for the application of Form-Chain: either we entirely avoid positing an empty category coindexed with the element placed in [SPEC,CP], which might result in a failure of interpretation of the wh-operator, or else we continue to posit an empty category but somehow relax the assumption that it be within the narrow domain of the elementary tree in which the operator is placed. The first of these options might be applicable in the case of elements which are inherently operators, and are therefore directly interpretable in the specifier position in which they are placed during Form-Chain. Indeed, this might be exactly what takes place in cases of adjunct extraction (cf. Williams 1993). If this view is correct, it provides a hint of an understanding for the strict locality of adjunct extraction. We will not pursue this further here.

For wh-elements which are not inherently semantic operators, matters of semantic interpretation force us to continue to posit the existence of a trace so as to form an operator–variable chain. Thus, we need to pursue the second of the options suggested above, that is allow for a relaxation of the domain in which the coindexed empty category is posited. The analyses of Kroch (1986,1989b) and Frank (1992) do just this. These analyses suggest that in cases like that in (40), Form-Chain applies to the auxiliary tree representing the matrix clause, such as (42).[25]
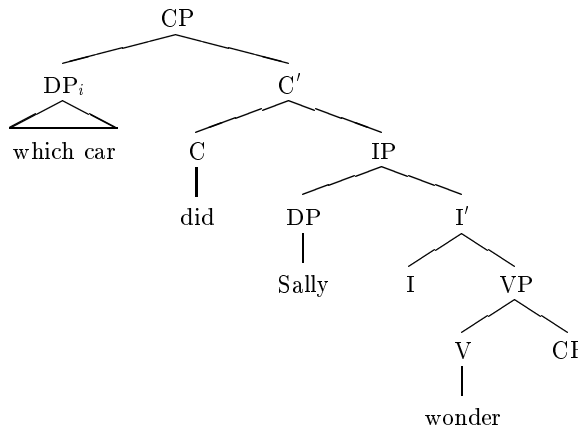
---

[25]Note that Frank and Kroch in the papers just cited do not use the Form-Chain operation. However, we can interpret their analyses in the terms of this paper as utilizing Form-Chain in the fashion suggested in the text.

(42)

```
                    CP
                    |
                    C′
                 ┌──┴──┐
                 C     IP
                 |   ┌──┴──┐
                did  DP    I′
                 |  ┌──┴──┐
               Sally I     VP
                        ┌──┴──┐
                        V     CP
                        |
                      wonder
```

.

The result of this operation is a pair of auxiliary trees rather than a single elementary tree. In particular, an element is inserted in the [SPEC,CP] position in (42) to form the auxiliary in (43)a, and the coindexed empty category is realized through the degenerate single node auxiliary tree in (43)b which bears the index of the moved element. We call this a degenerate auxiliary tree since its foot and root nodes are not only identical in label, but are the same node.

(43) a.

```
                    CP
               ┌─────┴─────┐
             DP_i          C′
             / \        ┌──┴──┐
         which car      C     IP
                        |   ┌──┴──┐
                       did  DP    I′
                        |  ┌──┴──┐
                      Sally I     VP
                               ┌──┴──┐
                               V     CP
                               |
                             wonder
```

b. DP$_i$

These two trees form a new kind of elementary object in the formal system, what we will call a *multi-component tree set*. The adjoining operation can now be generalized to operate over such tree sets so that a multi-component auxiliary tree set, i.e., a tree set consisting of auxiliary trees, may adjoin into a single elementary tree when all the elements of the tree set are simultaneously adjoined into the same underived elementary tree. Note that all pieces of the auxiliary tree set must be adjoined into a single elementary tree, and may not be dispersed throughout another auxiliary tree set.[26] To encode the requirement that the empty category be c-commanded by the wh-moved element, we will posit a link which connects the degenerate auxiliary in (43)b to the foot node of the auxiliary in (43)a which guarantees that when the two halves of the tree set are adjoined, this domination relation is preserved.

---

[26]So long as this definition of adjoining over multi-component tree sets is obeyed, it can be shown that certain basic formal properties of the TAG system, such as weak generative capacity, are unaffected by this generalization. A number of other possible ways of generalizing the adjoining operation to the multi-component case, some more powerful than the "benign" extension we adopt here, are considered in Weir (1988).

Observe that the use of multi-component auxiliary trees in this fashion requires that we modify the statement of the theta criterion and projection principle give above. In particular, the licensing requirements on the chain (in the projection principle) and the assignment of a theta role (in the theta criterion) will obtain only over the domain of the tree set and not over its constituent elementary trees. We will, however, continue to impose the CETM as a requirement which all component elementary trees in a multi-component tree set must satisfy.[27]

The multi-component version of the adjoining operation can now be used to derive example (40) by adjoining the two halves of the multi-component auxiliary in (43) into the elementary tree in (44)a at the CP root and at the DP complement position, yielding the structure in (44)b. To avoid irrelevant complications (see Frank (1992) for details), we omit traces of the wh-adverbial element *how* in the following representations.
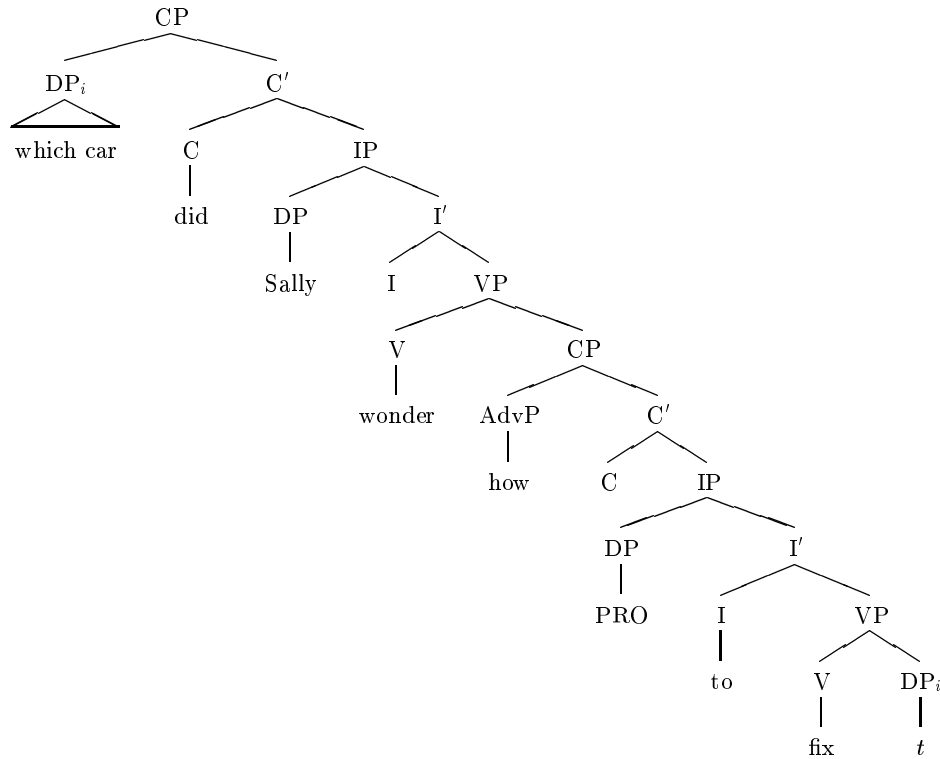
(44) a.

---
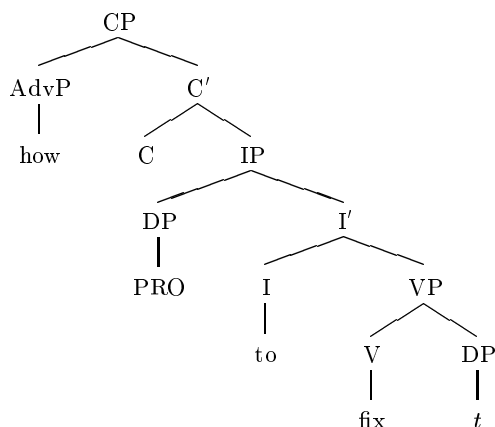
[27] In cases like (i), we must allow a single extended projection to be spread out among the trees of a multi-component auxiliary:

(i) What did Hugh seem to like?

Here, a multi-component set consisting of *did* and *seem* will adjoin into an initial tree for the lexical material of *What Hugh to like*. See Frank (1992) for further details of this analysis. We leave open the question of whether multi-component auxiliaries must in general consist collectively of only a single extended projection. See Bleam (1994) for a proposal to this effect.

b.

```
                              CP
                    ┌─────────┴─────────┐
                  DP_i                  C'
                  ╱────╲          ┌──────┴──────┐
                which car         C             IP
                                  │      ┌───────┴───────┐
                                 did    DP              I'
                                        │          ┌─────┴─────┐
                                      Sally        I          VP
                                                          ┌────┴─────┐
                                                          V          CP
                                                          │     ┌─────┴─────┐
                                                        wonder  AdvP        C'
                                                                 │     ┌─────┴─────┐
                                                                how    C          IP
                                                                           ┌───────┴───────┐
                                                                          DP              I'
                                                                           │         ┌─────┴─────┐
                                                                          PRO        I          VP
                                                                                     │      ┌────┴────┐
                                                                                     to     V       DP_i
                                                                                            │        │
                                                                                           fix       t
```

A point of interest in this derivation is that the empty category is present in the elementary tree in (44)a in spite of the fact that Form-Chain has not applied to this tree. This need not be cause for concern because, as we have commented in a number of places above, we have inserted the material below the DP argument nodes in our examples only for the sake of readability. In actuality, the material dominated by DP in a verbally headed elementary tree must be inserted during the derivation by substitution. Thus, even in cases of insertion of an empty category in a simple elementary tree, we must assume the existence of a DP elementary tree headed by an empty category, as in (45):

(45)

```
        DP
        │
        D'
        │
        D
        │
        t
```
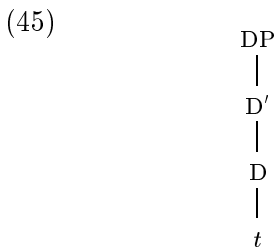
Thus, in the cases discussed in the last section, e.g., the derivation of example (32)b, the trace, like all other DP elements, will be inserted into the elementary tree in (31)b by substitution. We must therefore understand our statement of Form-Chain in (30) to govern directly the presence of non-terminal maximal projections having appropriate features rather than the terminal strings. The result of Form-Chain, then, is the presence of two non-terminals bearing identical indices in an elementary object. To guarantee that a wh-element and an empty category are inserted into the appropriate sites, we will assume that substitution and adjoining are further constrained by matching of sets of features at the site of the adjoining or substitution, essentially an extension of

the requirement that the substituted or adjoined structure must bear matching node labels.

The function of the degenerate auxiliary in the multi-component auxiliary tree set in (43) is now clear. It serves to coindex the empty category already present in the elementary tree in (44)a with the wh-operator in (43)a. A question that now arises is what licenses the presence of the empty category in object position in (44)a if not an application of Form-Chain. Let us suppose that we allow such empty elements to be freely generated, but impose certain well-formedness conditions on their presence. Kroch (1989b) suggests that such elements must satisfy a version of the ECP. In particular, he gives a formulation of the ECP which governs the appearance of empty elements in general, i.e., both traces produces by Form-Chain and the foot nodes of auxiliary trees. His formulation is as follows:

(46)   Empty Category Principle (TAG version): For any node X in an elementary tree α, if X is empty, then it must either be properly governed or be the head of an athematic auxiliary tree [i.e., an adjunction structure like that in (15)].

Kroch defines proper government as follows:

(47)   For any node X in an elementary tree α, X is *properly governed* if and only if one of the following conditions is satisfied:

1. the maximal government domain of X is the root node of α;

2. X is coindexed with a "local" c-commanding antecedent in α.

This statement of the ECP is satisfied in one of two ways, then. Either antecedent government obtains, where antecedent government is trivially defined as binding within a single elementary tree. Alternatively, an empty element can satisfy a head government requirement by having its maximal government domain be identical with that of the root of the elementary tree. The notion of *maximal government domain* of a node mentioned here corresponds to the highest projection p such that the path from p down to the projection of the governor of the node in question passes only through maximal projections which are themselves governed. The maximal government domain of an object DP trace in a simple sentence, then, will be the CP, since the path from this CP down to the trace proceeds through the IP, which is governed by C^0, and the VP which is governed by I^0, and it is the head of this VP which itself governs the trace. In essential respects, this corresponds to the notion of g-projection of Kayne (1984).

Returning to our discussion of long movement, the empty category in object position in (44)a satisfies the ECP via head government. Its maximal government domain is identical to the root CP, since the head governor of this trace is the verb whose projection is (canonically) governed by I, whose projection is in turn (canonically) governed by C. Antecedent government is impossible since there is no antecedent within the same elementary tree.

This version of the ECP predicts then that this type of derivation for long movement will be blocked whenever the unindexed trace in the subordinate clause is generated in a position whose maximal government domain is not the root node. Such a case arises in extraction from subject position. Indeed, such sentences are ungrammatical as the following example shows:

(48)   * Which car$_i$ did Sally wonder how t_i was fixed?

Similar considerations apply in the case of long-movement of an adjunct, as in (41). If we assume with Kroch that the adjunct trace is part of the subordinate clause elementary tree, then the fact that it is not head governed results in an ECP violation and hence ungrammaticality.[28] Finally,

---

[28]Frank (1992) proposes an alternative formulation of the ECP, which preserves the basic structure of Kroch's analysis, but avoids generating the traces of adjuncts in verbally headed elementary trees, a move which seems at odds with the view of elementary trees embodied in the CETM. Under Frank's conjunctive formulation of the ECP,

observe that the ECP also blocks the possibility of using multi-component adjoining to derive instances of super-raising, since the trace in the lower subject position would not be properly governed, just as in (48).

8   An Aside on Intermediate Traces

A salient property of the derivations and representations we have been discussing is the absence of intermediate traces. Since dependencies beyond the domain of a single clause are neither the result of applying move-α nor Form-Chain, there can be no intermediate traces left by the movement from [SPEC,CP] to [SPEC,CP]. Intermediate traces have carried a fairly substantial explanatory burden in grammatical theory. If we can succeed in showing that our TAG analyses mimic the beneficial effects of intermediate traces without positing their existence, we will be rid of the troublesome problems they raise, such as the question of their relevance to the ECP.

The most prominent motivation for intermediate traces has been their role in the characterization of locality. In the past few sections, we have seen briefly how our TAG-based analyses maintain locality of wh-dependencies in spite of the absence of intermediate traces. See Kroch (1987, 1989b) and Frank (1992) for further discussion. Another argument for the existence of intermediate traces has stemmed from the phenomenon of complementizer agreement, as seen in Irish for example. Frank (1992) shows, however, that this phenomenon can in fact be handled within a TAG-based derivation of wh-movement along the lines sketched here. Further, the phenomenon of French stylistic inversion, which was argued by Kayne and Pollock (1978) to be sensitive to the presence of an intermediate trace, has similarly been reanalyzed in a TAG-based analysis by Kroch and Joshi (1985). In each of these cases, the "derivational effects" of intermediate traces can be replicated within a TAG-based theory. What our TAG-based theory prevents us from stating is fortunately never required: a truly global condition which is sensitive to the simultaneous presence of all the traces of successive cyclic movement.

In the remainder of this section, we will investigate two more phenomena whose analyses have relied crucially the existence of intermediate traces. The first of these involves so-called connectivity or reconstruction effects, as discussed by Barss (1986) among others. The second concerns the relationship between the locality of syntactic movement and scope interpretation. We will argue that the absence of intermediate traces has no ill consequences in either of these empirical domains, and that a TAG-based treatment may even be empirically superior.

8.1   Connectivity Effects and Intermediate Traces

It has been observed that the set of possible binders of an element can be altered by the movement of a phrase containing it. In example (49)a, for instance, the reflexive *herself* cannot be bound by the matrix subject. However, if the DP containing it is fronted, this binding becomes possible, as seen in (49)b.[29]

---

it is a failure of antecedent government which is the relevant factor in the case of long-movement of adjuncts. See also Hegarty (1993) for another alternative to locality theory in a TAG-based model.

[29]An anonymous reviewer suggests that the contrast in (49) derives from the fact that *herself* is construed as an anaphor in (49)a and therefore must be bound to the closest subject, while in (49)b it is construed as a logophor and is bound to the higher subject. The reviewer argues that these two possibilities are distinguished lexically in Dutch, where the anaphor is *zichzelf*, but the logophor is *d'rzelf/'m zelf*. Thus, the Dutch translation of example (49)b is good only with *d'rzelf* (cf. Koster 1985 for examples and related discussion). While the semantico-pragmatic notion of logophoricity no doubt plays a significant role in the phenomenon of reflexivity, we do not see how it would explain the contrast induced by the wh-movement of the picture NP. In particular, it is unclear what blocks the logophoric interpretation of *herself* in (49)a, but permits it in (49)b. Note that the putative Dutch logophors are also impossible

(49) a. * Marsha$_i$ thought that I painted a picture of herself$_i$.
     b. Which picture of herself$_i$ did Marsha$_i$ think that I painted?

It has been suggested that data like this can be accommodated by allowing the transformational movement to be partially reversed or *reconstructed* at LF (cf. Langendoen and Battistella 1982, van Riemsdijk and Williams 1986, Clark 1992, *inter alia*). Applying reconstruction in (49)b, moving the wh-phrase back to the intermediate [SPEC,CP], yields a representation in which the reflexive may be locally bound by the matrix subject. Another account of these facts uses traces left by movement as potential loci into which binding may occur (cf. Guéron 1984, Barss 1986, Hornstein 1984, *inter alia*). On such an analysis, the trace in the intermediate [SPEC,CP] provides a position from which the matrix subject is an accessible binder for an anaphor, as shown in the following example of overt wh-movement to this intermediate position:

(50)   Marsha$_i$ wonders which picture of herself$_i$ I painted.

Of course, in a TAG-based analysis of wh-movement, which neither exploits intermediate traces nor allows the interaction of transformational movement with the derivation, neither of these solutions is possible. Therefore, if either of these accounts for such reconstruction effects were correct, we would be at a loss to explain the binding possibilities in (49)b. Consider, however, the example in (51).

(51)   Which picture of herself$_i$ did Marsha$_i$ wonder how/whether I had painted?

In this example, wh-movement has taken place out of a wh-island, and therefore has not left an intermediate trace in [SPEC,CP]. However, the binding possibilities are unaffected: the anaphor *herself* may be bound by the matrix subject, just as in (49)b, suggesting that both successive cyclic movement through an intermediate CP and any resulting intermediate trace are irrelevant for the purposes of anaphor binding in both (49)b and (51).[30] We conclude that what is going on in these cases seems not to depend on the existence of intermediate traces, or on any process of syntactic reconstruction.

    Further evidence against syntactic reconstruction comes from changes in binding preferences. Consider a reflexive within a "picture NP" in the direct object position of a ditransitive verb:

(52)   Alice$_i$ gave Helen$_j$ a picture of herself$_{j=i}$?

Both bindings for the anaphor seem equally possible here (which we indicate by the notation $j = i$). However, if this "picture NP" is fronted, we find that preferences shift.

(53)   Which picture of herself$_{i>j}$ did Alice$_i$ give Helen$_j$?

Here, by far the more prominent reading of the anaphor is the one in which it is coreferent with the subject (as shown by the notation $i > j$). If these cases involved reconstruction, then at the level at which binding applies (52) and (53) would be indistinguishable. Hence we would not expect there to be a difference in binding preferences.

---

in cases like (49)a (Koster 1985:157). Interestingly, however, clearly logophoric elements, such as the Ewe logophor *ye* (Clements 1975), are indeed possible in such contexts, but are impossible in cases like (49)b (C. Collins p.c.). It seems to us, therefore, that the distinction between the Dutch forms, and thus between the English cases in (49), is not tied to logophoricity, but rather to some structural notion, as Koster's (1985) structural analysis of the *zichzelf/d'rzelf* contrast suggests.

[30]A similar point has been made for Spanish in Campos (1993). Uriagereka (p.c.) points out to us that Barss notes such data and takes them to argue for the proposal of Chomsky (1986) where wh-movement proceeds by adjunction to VP. On such a view, it is via the VP-adjoined position, that chain binding takes place. However, with Cinque (1990) and Chomsky and Lasnik (1993) we reject the possibility that wh-movement may adjoin to VP on both empirical and theoretical grounds; hence we leave this option unexplored.

Instead of a process of syntactic reconstruction or a device like chain binding, we will account for these cases by allowing a subject to bind into an element in [SPEC,CP] of its own clause, following Reinhart (1981).[31] Reinhart argues that the ungrammaticality of examples like those in (54) derives from the fact that the subject is able to directly bind into the preposed PP, yielding a principle C violation.

(54)   a. * Near Dan$_i$ he$_i$ saw a snake.
       b. * For Ben's$_i$ car, he's$_i$ asking three grand.
       c. * Ben's$_i$ problems, he$_i$ won't talk about.

To show that the ill-formedness of the examples in (54) is not due to the configuration prior to movement or some process of reconstruction, or equivalently to the relationship between the subject and the trace, Reinhart cites examples similar to those discussed by Lakoff (1968) which demonstrate that an object which c-commands an element in its base position and thereby induces a principle C effect if it remains in situ, does not do so if the element is fronted.

(55)   a. * Rosa showed him$_i$ her new tricks in Dan$_i$'s apartment.
       b. * I'm willing to give him$_i$ two grand for Ben$_i$'s car.
       c. * You can't talk to him$_i$ about Ben$_i$'s problems.
(56)   a. In Dan$_i$'s apartment, Rosa showed him$_i$ her new tricks.
       b. For Ben$_i$'s car, I'm willing to give him$_i$ two grand.
       c. Ben$_i$'s problems, you can't talk to him$_i$ about.

This subject/object asymmetry has been overlooked in much recent work. It is inconsistent with standard accounts of reconstruction, but is captured directly under Reinhart's suggestion that the subject but not the object c-commands the fronted element. Reinhart's proposal allows us to account for the possibility of binding of the anaphor by the subject in (49)b and (51) since the matrix subject is now able to bind into the fronted wh-phrase. Similarly, the configuration in which the subject binds the anaphor in (53) is distinct from that in which it binds it in (52) and we can link the change in prominence of the readings to this difference.

Nevertheless, the grammaticality of examples like the following seems to suggest that we must still sometimes allow binding to obtain via a trace in the base position:

(57)   Which picture of himself$_i$ do you think that Herb$_i$ likes best?

However, within our TAG-based analysis of successive cyclic dependencies, this fact follows without use of chain-binding, reconstruction or the like. Recall that configurations relevant for anaphor binding cannot be determined on the basis of the derived structure as this structure plays no role in our model of grammar. We must instead suppose that the relevant configurations are determined by the structure of the elementary trees. The translation of Reinhart's proposal into this framework, then, will allow a subject in [SPEC,IP] to command a fronted element, in [SPEC,CP] say, when the two are both generated within a single elementary tree. The straightforward derivation of (57) involves the use of an elementary tree containing both the lower subject and the matrix [SPEC,CP] positions (cf. the derivation using (31)b above). The matrix clause is simply inserted between these two elements during the course of the derivation using the adjoining operation. Thus, the desired binding obtains. Accounting for cases like (51) is similarly straightforward. Recall from section 7 that cases of extraction from wh-islands make use of multi-component tree sets in which the auxiliary tree corresponding to the matrix clause already contains the position for the fronted

---

[31]Of course, this description of Reinhart's proposal is a bit anachronistic as she made rather different theoretical assumptions and therefore did not discuss principle C or [SPEC,CP]. Nonetheless, the translation of her analysis into our current framework is clear. We leave open, however, the precise characterization of the command relation which allows elements in [SPEC,CP] to be commanded by subjects.

wh-element (cf. (43)). Consequently, the matrix subject is able to command and hence bind the fronted anaphor.

Consider next the example in (49)b, which now appears problematic. Here we seem to have a case of a simple adjoining derivation in which the matrix subject is nonetheless able to bind an anaphor fronted from the embedded clause. However, nothing in our discussion thus far prevents cases derivable using simple adjoining from having alternative derivations which involve the multi-component derivations that are necessarily used for wh-island cases. Under such a derivation for (49)b, the binding possibilities are predicted to be exactly like those in (51), thereby allowing the matrix subject binding.

With these problems behind us, there remains one case we have not covered.

(58)   Which picture of herself$_i$ do you wonder whether Alice$_i$ likes?

Here, the extraction from the wh-island forces a multi-component adjoining derivation. Consequently, the embedded subject cannot command the wh-phrase in its fronted position since they do not reside in the same elementary tree. In order to resolve this conflict, we will adopt a weakened version of the binding via trace of movement account, following Barss *et al.*, only for the case of anaphor binding.[32] In particular, we will allow an element α to command an element β if α commands the tail of a chain whose head γ dominates β. This definition is needed in any case to generate the second reading for an example like (53). Locality conditions on binding will be characterized in the usual fashion. Using this extended definition of command allows the embedded subject *Alice* to command the reflexive *herself* since the subject commands the tail of the chain of the entire wh-phrase. Note that this command condition remains a condition determined by the structure of elementary trees since the set of elements dominated by the head of the chain whose tail is the object trace is determined entirely by the choice of which multi-component elementary tree is adjoined to the embedded clause.

This extension to our account of anaphor binding makes a number of predictions. First is the apparently correct prediction that wh-movement never removes any possibilities for anaphor binding that exist when the moved element is in its base position. The second prediction is that certain reflexives should allow split antecedents. This prediction derives from the extra possibility of local command between an NP and an anaphor resulting from a subject's ability to bind into a fronted wh-phrase. Thus, if we front a wh-phrase containing a reflexive but no local binder, the reflexive can be locally bound either by the elementary-tree-local subject or by the subject of the clause in which the trace resides. In derivations using simple adjoining, these two will not be distinguishable. However, in multi-component derivations we should witness the simultaneous binding by an embedded subject and a higher subject. Surprisingly, this seems true, as shown in the following case:

(59)   Which picture of themselves$_{i+j}$ did John$_i$ think that Mary$_j$ had bought?

Here, we assume a multi-component derivation, so that the matrix subject *John* binds the reflexive via command into its local [SPEC,CP], while the lower subject binds the reflexive via the trace in the embedded clause. This case contrasts in acceptability with one in which the two binders are both subjects of higher clauses.

(60)   * Which picture of themselves$_{i+j}$ did John$_i$ think that Mary$_j$ said that I had bought?

---

[32]Examples like those in (56) show that such chain binding or reconstruction cannot be allowed to obtain for the purposes of principle C. This dichotomy is perhaps indicative of a divergence in the nature of principles A and C. We might take principle A to be a purely syntactic condition and therefore potentially sensitive to details of syntactic representation such as the presence of traces, while principle C is non-syntactic, but instead a constraint on possible interpretations (cf. Chomsky 1992).

Here, the dual binding is not possible since the trace is not elementary-tree-local to either of the potential binders, and the wh-phrase cannot have simultaneously been generated in the same elementary tree as both of the subjects. If we alter this example slightly by placing the binders in the lowest and highest subject positions, it improves markedly.

(61)   Which picture of themselves$_{i+j}$ did John$_i$ think that I said that Mary$_j$ had bought?

Here, the lowest subject can bind via the trace, while the matrix subject can bind directly into the [SPEC,CP] position. Also grammatical, as predicted, is the following example where the local command relation between the middle subject *Mary* and the wh-element is present in the elementary tree, but is interrupted during the course of the derivation by the adjoining of the matrix clause.

(62)   Which picture of themselves$_{i+j}$ did you think that Mary$_i$ said that John$_j$ had bought?

8.2   Scope Asymmetries in Long and Short Movement

Let us now turn to a second case in which the existence of intermediate traces has been exploited, that of scope dependencies. Consider the derivations of filler–gap dependencies induced by our version of the Form-Chain operation. Those which require only the use of the simple adjoining operation correspond exactly to the cases which are standardly assumed to involve successive cyclic movement. Derivations requiring the use of multi-component adjoining are exactly those cases which involve long movement. There is an important difference between the structures which are induced by Form-Chain in the two cases. In "local" applications of Form-Chain, the "moved" element is placed in the specifier of its own clause, while in non-local applications the moved element is placed in an operator position of a clause higher than its own.

Let us suppose that we make the following minimal assumption concerning scope of quantifica-tional expressions: they receive scope in the position in which the application of Form-Chain places them. We now predict that there should be asymmetries in scope interpretation between instances of long and of successive cyclic movement of quantifiers. In particular, we predict that all cases of long-moved quantifiers must take scope outside of the clause in which they are generated.

Data concerning the extraction of amount quantifiers (AQs) suggests that this is indeed true. As discussed by Cinque (1990) and Longobardi (1990), the possible interpretations of an extracted AQ such as *how many books* depend upon the nature of the extraction. In (63), where the extraction proceeds successive cyclically, or via simple adjoining, the sentence may be understood with either of the interpretations in (64).

(63)   How many books did the editors decide to publish?
(64)   a. For what number x, for some set of books b of cardinality x, the editors decided to publish b.
        b. For what number x, the editors decided that for some set of books b of cardinality x they will publish b.

In the first of these, there is a particular set of books that is presupposed by the questioner – this is evidenced by the wide scope existential quantifier – while in the latter reading only a number is at issue and no set is presupposed by the speaker. Cinque labels these the referential and quantificational readings, respectively.

Turning to cases of long movement of an AQ, we see that only the first of these readings is possible.

(65)   How many books did the editors wonder whether to publish?

(66)   a. For what number x, for some set of books b of cardinality x, the editors wondered whether to publish b.
       b. * For what number x, the editors wondered whether for some set of books b of cardinality x, they should publish b.

Cinque argues that the quantificational, but not the referential, interpretation of an AQ requires antecedent government to obtain, and hence this reading is impossible in cases of long movement. We can provide an alternative account of this asymmetry, observing that the difference between the two readings lies in differences in the scope of the existential quantifier over the set of books provided by the AQ *how many books*. Following our assumption concerning scope assignment, the facts now reduce to the statement that an existential quantifier receives scope in the clause in which it is generated under Form-Chain. In (63), the operator may be generated in either the matrix or the embedded [SPEC,CP] positions, leading to either simple or multi-component derivations. On the other hand, in (65) the only position for the generation of the operator is in the matrix [SPEC,CP]; hence it must be interpreted as having wide scope.

   It is interesting to observe that the recursive character of the TAG adjoining operation forces us to analyze certain cases of extraction as involving multi-component adjoining which we might otherwise analyze as successive cyclic movement. In particular, extraction from nominals will necessarily involve an application of Form-chain which yields a multi-component tree set, since the target [SPEC,CP] position cannot be present in the same elementary tree as the base position. A [SPEC,CP] position would necessitate the presence of a CP projection, and hence a verbal extended projection, while the extraction site is, by hypothesis, from the argument of a noun, and hence presupposes a nominal extended projection, a violation of the CETM if both are present in a single elementary tree.[33] This leads us to the prediction that AQ interpretation should be unambiguously wide scope, referential in Cinque's terms. This is borne out, as shown by the following data from Italian (modified slightly from Longobardi 1990) and English.

(67)   a. Di quanti   autori   è   facile recensire i libri?
          Of how many authors is it easy   to review the books?
       b. How many books did she decide on the publication of?

In both English and Italian, only the interpretation which presupposes the existence of some set of authors or books is available.

   If this analysis is on the right track, it suggests that it may not be the fronted wh-element which provides the interrogative force of a wh-question since matrix question force is retained regardless of whether the derivation involves the generation of the wh-element in the subordinate clause, resulting in simple adjoining, or in the matrix clause, resulting in multi-component adjoining. The semantic force of the AQs we have been considering is rather the existential quantifier whose scope we have seen to vary. Thus, we might suppose that the interrogative force of a wh-question resides in a head bearing appropriate features, perhaps the Q-morpheme of Baker (1970). This view is consonant with proposals such as that of Rizzi (1991), where the [+WH] features on the C^0 head attract the movement of the wh-operator. Such a proposal might also allow us to understand the curious phenomenon of wh-imperatives discussed by Reis and Rosengren (1992) where a fronted wh-element in German takes subordinate clause scope, despite its sentence initial position.

---

[33]In fact, Stowell (1989) does analyze extraction from nominals as movement through a [SPEC,DP] escape hatch in exactly parallel fashion to movement through the clausal [spec,cp] escape hatch. The facts concerning AQ interpretation remain unexplained on his account, needless to say. We should point out also that extraction of AQs from gerunds behaves like that from clauses and not nominals. In other work (Frank and Kroch 1994), we have used this and other data to suggest an IP analysis of gerund structures.

(68)   Welches Buch sag mir daß du gelesen hast
        which book   tell me that you read   have
        Tell me which book you have read!

9   Conclusions and Directions

To sum up, this paper has investigated some of the consequences of an alternative system of generalized transformations to that proposed in Chomsky (1993). Our TAG-based proposal, based on the operations of adjoining and substitution, shares with Chomsky's system the elimination of a D-structure level of representation. However, it solves certain conceptual inelegancies of Chomsky's extension requirement by eliminating entirely the grammatical relevance of such a requirement. Further, we have explored the analysis of so-called successive cyclic and long wh-dependencies within our framework and found that the operation which creates antecedent–trace dependencies, what we call Form-Chain following Chomsky, may be uncoupled from the process of phrase structure composition. This separation allows us to understand the locality of such dependencies without out appeal to economy conditions: locality follows from the very architecture of the grammar.

One question which we've left open here concerns the interface between syntax and the conceptual-intentional systems, what in Chomsky's system appears as the level of Logical Form. This is no accident: our derivations do not explicitly include a syntactic level of LF.[34] It is worth noting that, thus far, the only role we have assigned to the phrase marker that is constructed during a TAG derivation is that of determining the order of the words in the sentence. We have appealed to no grammatical well-formedness conditions which constrain the character of the derived phrase marker. Indeed, we believe that feeding the phonology is the only role of this derived structure. As for the interface to conceptual-intentional systems, our proposal would be to follow the line taken in the earliest work in generative grammar (cf. Chomsky 1955, 1957): the object relevant to interpretation is the structure which records the history of derivational steps, the T-marker or derivation structure.[35] Naturally, this assumption has a great many consequences that go beyond the scope of the present article and we are pursuing them in on-going work.

---

[34]We should point out that it is possible to enrich the system of derivations we have presented thus far so as to incorporate a level of LF. This is done by utilizing the mechanism of synchronous TAG proposed by Schabes and Shieber (1990). In this way, it is straightforward to recast virtually all of the standard LF analyses.

[35]A historical aside is perhaps appropriate at this point. Along with the demise of generalized transformations in Chomsky (1965) came the removal of T-markers. As we discussed in the introduction to this paper, the change resulted to a large degree from the excessive expressive power that T-markers allow. In particular, one would expect there to exist within the grammar constraints on the T-markers themselves, perhaps ordering constraints among the generalized and singular transformations, which exhibit cross-linguistic variation, as had been observed with singular transformations alone. Yet, such conditions were found to be unnecessary, and it was concluded that T-markers were a representational excess. However, an alternative line might have been followed, as suggested by Fillmore (1963). He proposed a model of grammar which utilized generalized transformations in a somewhat restricted fashion. One feature of his model was that the application of generalized and singular transformations could not intermingle in arbitrary ways, and in fact intermingled only in the ways actually observed. Bach (1977) presented further empirical support for Fillmore's proposal.

Our TAG-based model differs from the pre-*Aspects* proposals in blocking interaction between generalized and singular transformations (in a way rather different from Fillmore) since the latter do not exist as derivational operations in the same way as the former. From this, we can see that the analog of the T-marker in our theory the derivation structure, is an extremely restricted formal object. (It can, in fact, be shown to be a context-free generable tree structure). Given the return of generalized transformations in the minimalist framework, one wonders what the status of that framework's correlate of the T-marker might be.

References

Abeillé, Anne. 1988. Parsing French with tree adjoining grammar: Some linguistic accounts. In *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest.

Abeillé, Anne and Yves Schabes. 1989. Parsing idioms with a lexicalized tree adjoining grammar. In *Proceedings of the European Conference of the Association for Computational Linguistics*, Manchester.

Bach, Emmon. 1977. "The position of embedding transformations in a grammar" revisited. In A. Zampolli (Ed.), *Linguistic Structures Processing*. North Holland.

Baker, C. Lee. 1970. Notes on the description of English questions: the role of an abstract question morpheme. *Foundations of Language* 6.

Baltin, Mark and Anthony Kroch (Eds.). 1989. *Alternative Conceptions of Phrase Structure*. Chicago, IL: University of Chicago Press.

Barss, Andrew. 1986. *Chains and Anaphoric Dependence*. PhD thesis, MIT, Cambridge, MA.

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

Bleam, Tonia. 1994. Clitic climbing in spanish. Manuscript, University of Delaware.

Campos, Hector. 1993. Reconstruction and picture nouns in Spanish. Manuscript, Georgetown University.

Chomsky, Noam. 1955. *The Logical Structure of Linguistic Theory*. Distributed by Indiana University Linguistics Club. Published in part by Plenum, 1975.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

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

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Kenneth Hale and Samuel Jay Keyser (Eds.), *The View from Building 20*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 1994. Bare phrase structure. MIT Occasional Papers in Linguistics 5.

Chomsky, Noam and Howard Lasnik. 1993. Principles and parameters theory. In J. Jacobs, A. von Stechow, W. Sternefeld, and T. Vennemann (Eds.), *Syntax: an International Handbook of Contemporary Research*, 506–569. Berlin: de Gruyter.

Cinque, Guglielmo. 1990. *Types of $\overline{A}$-Dependencies*. Cambridge, MA: MIT Press.

Clark, Robin. 1992. Scope assignment and modification. *Linguistic Inquiry* 23:1–28.

Clements, George N. 1975. The logophoric pronoun in Ewe: Its role in discourse. *Journal of West African Languages* 10:141–177.

Ernst, Thomas and Chengchi Wang. 1995. Object preposing in Mandarin Chinese. *Journal of East Asian Linguistics* 4(4).

Fillmore, Charles. 1963. The position of embedding transformations in a grammar. *Word* 19.

Frank, Robert. 1992. *Syntactic Locality and Tree Adjoining Grammar: Grammatical, Acquisition and Processing Perspectives*. PhD thesis, University of Pennsylvania, Philadelphia, PA.

Frank, Robert and Anthony Kroch. 1994. Nominal structures and structural recursion. *Computational Intelligence* 10(4):453–470.

Grimshaw, Jane. 1991. Extended projection. Manuscript, Brandeis University.

Guéron, Jacqueline. 1984. Topicalisation structures and constraints on coreference. *Lingua* 63.

Hegarty, Michael. 1993. Wh fronting and the composition of phrase structure in Tree Adjoining Grammar. Manuscript, University of Pennsylvania.

Hornstein, Norbert. 1984. *Logic as Grammar*. Cambridge, MA: MIT Press.

Iatridou, Sabine. 1991. *Topics in Conditionals*. PhD thesis, MIT, Cambridge, MA.

Iatridou, Sabine and Anthony Kroch. 1992. Cp-recursion and its relevance to the Germanic verb-

second phenomenon. *Working Papers in Scandinavian Syntax* 50:1–24.

Jonas, Diane and Jonathan Bobaljik. 1993. Specs for subjects: the role of TP in Icelandic. In J. Bobaljik and C. Phillips (Eds.), *Papers on Case and Agreement I, MIT Working Papers in Linguistics 19*, 59–98. MIT Department of Linguistics and Philosophy.

Joshi, Aravind K. 1985. How much context-sensitivity is required to provide reasonable structural descriptions: tree adjoining grammars. In David Dowty, Lauri Kartunnen, and Arnold Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*. Cambridge: Cambridge University Press.

Joshi, Aravind K., Leon Levy, and Masako Takahashi. 1975. Tree adjunct grammars. *Journal of the Computer and System Sciences* 10(1):136–163.

Kasper, Robert. 1992. Compiling Head-driven Phrase Structure Grammar into Lexicalized Tree Adjoining Grammar. Presented at the TAG+ Workshop, University of Pennsylvania.

Kayne, Richard. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris.

Koopman, Hilda. 1995. On verbs that fail to undergo V-second. *Linguistic Inquiry* 26(1):137–163.

Koster, Jan. 1985. Reflexives in Dutch. In J. Guéron, H.-G. Obenauer, and J.-Y. Pollock (Eds.), *Grammatical Representation*, 141–167. Dordrecht: Foris.

Kroch, Anthony. 1987. Unbounded dependencies and subjacency in a tree adjoining grammar. In Alexis Manaster-Ramer (Ed.), *The Mathematics of Language*, 143–172. Amsterdam: John Benjamins.

Kroch, Anthony. 1989. Asymmetries in long distance extraction in a tree adjoining grammar. In (Baltin and Kroch 1989).

Kroch, Anthony and Aravind K. Joshi. 1985. The linguistic relevance of tree adjoining grammar. Technical Report MS-CS-85-16, Department of Computer and Information Sciences, University of Pennsylvania.

Kroch, Anthony and Aravind K. Joshi. 1987. Analyzing extraposition in a tree adjoining grammar. In Geoffrey Huck and Almerindo Ojeda (Eds.), *Discontinuous Constituents*, Vol. 20 of *Syntax and Semantics*. New York: Academic Press.

Kroch, Anthony and Beatrice Santorini. 1991. The derived constituent structure of the West Germanic verb raising construction. In Robert Freidin (Ed.), *Principles and parameters in comparative grammar*, 269–338. Cambridge, MA: MIT Press.

Lakoff, George. 1969. Pronouns and reference. Distributed by Indiana University Linguistics Club.

Langendoen, Terrence and Edward Battistella. 1982. The interpretation of predicate reflexive and reciprocal expressions in english. In *Proceedings of the 12th Annual Meeting of the North Eastern Linguistic Society*. Graduate Linguistics Students Association, University of Massachusetts, Amherst.

Longobardi, Giuseppe. 1990. Extraction from NP and the proper notion of head government. In A. Giorgi and G. Longobardi (Eds.), *The Syntax of Noun Phrases*. Cambridge: Cambridge University Press.

Muysken, Pieter. 1982. Parametrizing the notion 'head'. *Journal of Linguistic Research* 2:57–75.

Reinhart, Tanya. 1981. Definite NP anaphora and c-command domains. *Linguistic Inquiry* 12(4):605–635.

Reis, Marga and Inger Rosengren. 1992. What do WH-imperatives tell us about WH-movement. *Natural Language and Linguistic Theory* 10:79–118.

Rivero, María-Luisa. 1991. Long head movement and negation: Serbo-Croatian vs. Slovak vs. Czech. *The Linguistic Review* 8:319–351.

Rivero, María-Luisa. 1993. Finiteness and second position in long head movement languages: Breton and Slavic. Manuscript, University of Ottawa.

Rizzi, Luigi. 1991. Residual verb second and the WH criterion. Manuscript, Université de Genève.

Schabes, Yves and Stuart Shieber. 1990. Synchronous tree adjoining grammars. In *Proceedings of the 13th International Conference on Computational Linguistics*, Helsinki.

Stowell, Tim. 1981. *Origins of Phrase Structure*. PhD thesis, MIT, Cambridge, MA.

Stowell, Tim. 1989. Subjects, specifiers and X-bar theory. In (Baltin and Kroch 1989).

van Riemsdijk, Henk and Edwin Williams. 1986. *Introduction to the Theory of Grammar*. Cambridge, MA: MIT Press.

Vijay-Shanker, K. 1987. *A Study of Tree Adjoining Grammars*. PhD thesis, University of Pennsylvania, Philadelphia, PA.

Vijay-Shanker, K. and Aravind K. Joshi. 1985. Some computational properties of tree adjoining grammar. Technical report, Department of Computer and Information Science, University of Pennsylvania.

Weir, David. 1988. *Characterizing Mildly Context-Sensitive Grammar Formalisms*. PhD thesis, University of Pennsylvania, Philadelphia, PA.

Williams, Edwin. 1993. The scopal ECP. *GLOW Newsletter* 30. Paper presented at the 16th GLOW Colloquium.