

Why MOVE ?

Adam Szczegielniak
Department of Linguistics
Harvard University
77 Dunster Street
Cambridge, MA, 02138
E-mail: SZCZEG@FAS.HARVARD.EDU
Version #7 (10/15/00)

1. The Operations¹

Chomsky (1995b,c) hypothesises that when elements are 'displaced' in a given language, it is a result of the application of certain transformational procedures, including²:

- (1)a. Move F, and
- b. Move α

(1b) is construed as consisting of an ordered set of operations. Hence, overt syntactic movement is performed by Move α which consists of the following ordered operations:

- (2) Move $\alpha \equiv$
 - a. Move F,
 - and b. Pied-piping of other features, (including P-features),
 - and c. Merge
 - and d. Repair operation.

Hence, Move α is not a single elementary operation, but is in fact an instance of Move F plus operations (2b,c) followed by (2d) necessary to satisfy the output conditions, imposed on PF representations. However, Move α does not only pied-pipe P-features, according to Chomsky(1995c), features which can be checked, in addition to features which must be checked, also move along³.

Chomsky (1995b) makes use of another operation called Select/Merge which is distinct from the operations mentioned in (2). Select selects a lexical item LI from a Numeration, understood as "a set of pairs (LI, *i*), where LI is an item of the lexicon and *i* is its index, understood to be the number of times that LI is selected" (Chomsky 1995b:225), and reduces its index by 1. Select introduces the item into the derivation, whereas Merge takes a pair of syntactic objects (SO_{*j*}, SO_{*i*}) provided to the computation by Select and replaces them with a new syntactic object SO_{*ji*}.

Move⁴ and Merge share many properties, each must satisfy the economy conditions⁵, both can 'deliver' uninterpretable features which must be checked and erased prior to spell out, at least in principle⁶. Additionally, Merge is cyclic and can operate only on the root, Move α , because it contains Merge and the pied-piping of phonetic features, is also cyclic.

2. The Numeration

I would like to propose the following central hypotheses:

- (3)a. A more optimal solution to the Bare Output Conditions is a C_{HL} which employs only Merge as a structure building operation.
- b. The Numeration (N_{HL}) is a 'buffer' between the Lexicon and the Computational System (C_{HL}).

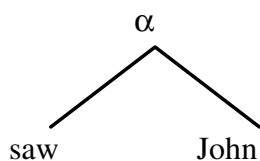
Chomsky (1995b:226-227) describes a Numeration N as the input to C_{HL} . Although it is not stated explicitly, it is also, the output of an indexing operation which determines how many times an identical Lexical Item (LI) will be accessed from the Numeration by C_{HL} . The mapping from N_{HL} to C_{HL} is straightforward, the latter accesses N_{HL} i times for every LI_i . However, this process cannot be taken literally where C_{HL} takes from the Numeration all the LI's until there is nothing there (all indices equal zero), since the Numeration, has another important role to play, namely, it is the reference set for comparing the economy of pending derivations, i.e. Global Economy (Chomsky 1995b:227). It seems then that there has to be a record of all the LI's and of their indices reflecting the initial state of N_{HL} before C_{HL} had any access to it and that record has to be accessible by C_{HL} until Spell-out.

I would like to propose that the Numeration N_{HL} has the following properties:

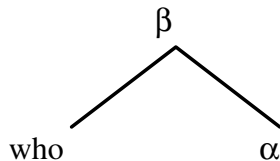
- (4) The Numeration N_{HL} is a set of pairs (LI {i,y}), where LI is a Lexical Item and {i,y} is its indices, where:
 - i=the number of times a given LI occurs in θ -role configuration,
 - y=the number of times a given LI occurs in a Checking configuration

It is important to note that C_{HL} is allowed to alter the indices in the order they are listed, or simultaneously⁷. The former situation when {i,y} cannot be changed by 1, simultaneously gives rise to the displacement of elements. It is worth noting that in the case of i the index operation is subtraction, every time an LI is chosen to satisfy Theta marking⁸.

(5) a.



b.



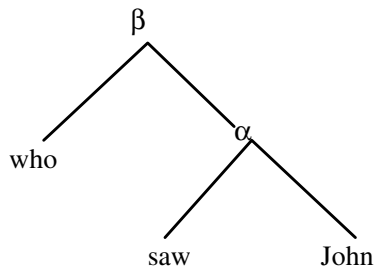
The relevant entries in the Numeration are:

	INITIAL	AFTER (5a)	AFTER (5b)
(6) a.	$LI_{\text{see}}\{2,y\}$	$LI_{\text{see}}\{1,0\}$	$LI_{\text{see}}\{0,0\}$
b.	$LI_{\text{who}}\{1,y\}$	-	$LI_{\text{who}}\{0,0\}$
c.	$LI_{\text{John}}\{1,y\}$	$LI_{\text{John}}\{0,0\}$	-

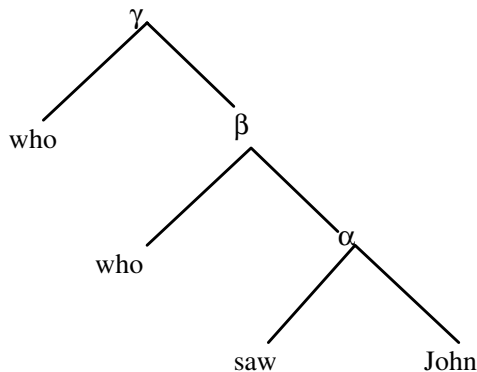
I will return to why the initial setting of y remains unspecified, presently we can assume that $y=0$. We can see that in (5b) all the selectional properties of the verb and both nouns have been satisfied (the i index is reduced to 0 in every case).

In the case of y , it is addition every time an LI is chosen to satisfy feature checking.⁹

(7) a.



b.



If this is the end of the derivation then the Numeration has been altered for the Lexical Item 'who'.

(8) After (7a)

$LI_{\text{who}}\{0,0\}$

After (7b)

$LI_{\text{who}}\{0,1\}$

There is no need for the kind of operations listed in (2), since there is a more optimal solution to the Bare Output Conditions, this is the independently motivated operation Select/Merge.

The difference between Merge and Move in Chomsky (1995b:316) is that the latter is responsible for the displacement property in language. As for the nature of the displacement mechanism, "on minimalist assumptions we want nothing more than an indication at LF that the position in which the displaced item is interpreted, that is chains are legitimate objects at LF."

Chomsky (1995b) argues that the operation Move is responsible for the introduction of chains. I propose a solution where the interaction of Numeration and Select/Merge is a sufficient and more optimal solution to the Bare Output Conditions.

Consider N_{HL} with its properties as in (4). First the operation Select/Merge satisfies the Selectional properties of LI's until $i=0$ for every $LI\{i,y\}$.¹⁰ Then the operation Select/Merge selects from N_{HL} Lexical Items, bearing an index $LI\{0,y\}$ in order to have the appropriate features of the attracting X^0 checked. It is important to stress that this is a derivational procedure and that all the properties of Move α as in Chomsky (1995b,c) can be formulated in terms of Merger of an $LI\{0,y\}$ from N_{HL} .

However, if the two analyses were completely inter-translatable (i.e. they were notational variants) there would be no reason to adopt one over the other.

In Chomsky (1995b) the existence of Merge, Select, Attract and of a Numeration is required for reasons independent of the ones which justify the existence of Move α . Hence, even if it turns out that the two approaches are just notational variants, then the one without Move α is on conceptual grounds a better solution to BOC's than an approach which has an additional entirely superfluous operation.

The question is, however, are the two approaches empirically equivalent?

3. The properties of Merge

Two properties of the approach adopted here seem to fall out automatically:

- (9)a. Merge is always cyclic.
- b. Merge can apply only to elements which are X^{0max} , or higher.

Chomsky (1995b) must postulate that (9a) and (9b) are a result of the fact that Move α is a set of sub-operations as stated in (2) above.¹¹

However, the model proposed here does raise some questions. For example:

- (10) a. How are chains formed?
- b. Move α can take as its input LI's and products of Merge, how does Merge 'know' when to chose what?
- c. In what configurations is feature checking possible.

Chains are a property of N_{HL} . Hence, a valid chain at LF is:

- (11) CH = {A, B},
 A = $LI_i\{i=0, y=\max\}$,
 B = $LI_j\{i=\max, y=0\}$
 where $LI_i = LI_j$

The chain in (11) is formed in a trivial fashion: the second member of the chain $LI\{i=\max, y=0\}$ is the position where LI has satisfied all its selectional properties (where it either receives or assigns θ -roles). This corresponds to the application of Merge as in Chomsky (1995b), and is a position prior to any movement. This member of the chain is the Copy of the LI. The first member of the chain $LI\{i=0, y=\max\}$ indicates the position of LI after it has entered into all the necessary checking configurations. This corresponds to the displaced 'original' and is the head of the chain CH. Hence, the chain CH resulting from Wh-movement in (7) is $CH = (LI_{who}\{0,1\}, LI_{who}\{0,0\})$.

As far as (10c) is concerned, the notion of Checking Domain is reformulated in terms of 'closeness'. A LI can check its features with a head H if its is close to it, close = SPEC of H, or in SPEC of its projection (with the exception of adjunction).

However, question (10b) resulting from the property of Merge in (9b) can only be answered by stating that Merge primarily has to satisfy the selectional properties of elements, and only later can it satisfy the feature checking requirement of LI's. Notice that the same question does arise in Chomsky (1995b), mainly why Merge applies before Move α .

More problematic is the instance when Merge has to apply to its own products. For example let us take a sentence 'Which man killed John' where the subject is a DP = [DPwhich man]. Once the DP is formed by merging the LI's 'which' and 'man' it functions as a syntactic

object, hence, a proper English sentence would involve the wh-movement of the whole DP giving:

(12) Which man ___ killed John

Before analysing whether (12) can be generated in the framework presented above, Let us take a closer look at the derivation of (12) along the lines of Chomsky (1995a,b,c).

4. Merge and Move

Let us derive (12a) along the lines proposed by Chomsky (1995a,b,c).¹²

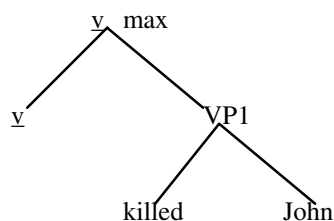
The Numeration N_{HL} consists of the following LI's:

(13)

LI ₁ = John	LI ₃ = kill
LI ₂ = <u>v</u>	LI ₄ = which
LI ₅ = man	LI ₆ = T
LI ₇ = C	

The Operation Select selects 'John' from N_{H_L} and Merge merges it with 'kill' (also selected from N_{H_L}) creating a set A = {<kill>, {kill, John}} denoted in (14) as VP1.¹³ The operation Select then selects v from N_{H_L} and merges it with A = {<kill> {John, kill}} yielding B = {<v>, {v, A}} = {<v>, {v, {<kill>, {kill, John}}}}, denoted in (14) as v^{max}. Both instances of Merge are motivated by Selectional requirements.

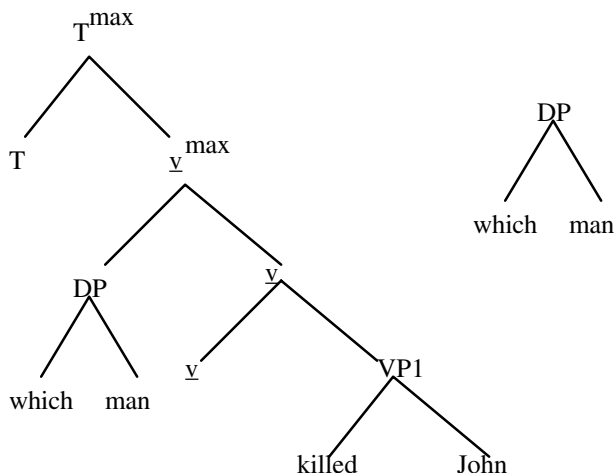
(14)



The Selectional properties of B = {<v>, {v, A}} require that it be merged with an element to which it can assign a θ role (that of an AGENT). There are two possibilities, either 'John' is selected, or 'which man'.¹⁴

In the model adopted by Chomsky (1995b) Merge is preferred to Move, and hence 'which man', which has undergone merger forming K = {<which> {which, man}}, is selected creating D = {<v>, {K, B}} = {<v>, {{<which>, {which, man}}, {<v>, {v, {<kill>, {kill, John}}}}}} shown in (15) as v^{max}¹⁵. Merge then merges T with v^{max} to form E = {<T>, {T, D}}, indicated as T^{max} in (15).

(15)

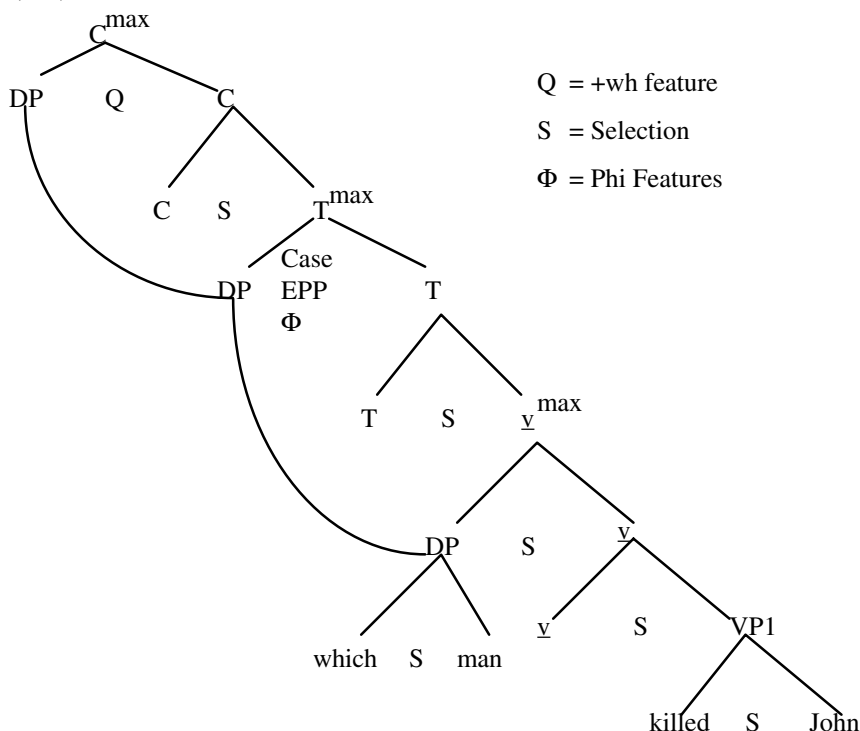


In English, T has at least one SPEC.¹⁶ Following Chomsky's recent proposals let us assume that T in English cannot 'see' the embedded 'John' and hence can only Attract the DP [which man].¹⁷

Thus the Case, Φ , and D features Move to T, expanding the head T and creating T^0_{max} , shown in (16) as movement 1¹⁸. Notice that the expansion of the head does not alter the structure which is the same as in (15), since it is only the head which has been expanded through an operation whose effects are not visible to C_{HL} .

However, since we do not find natural languages which have single features displaced, but we do find whole words and phrases which are displaced, Chomsky assumes that Feature Movement is followed by pied-piping of the remaining features to the SPEC of T, indicated as 2 in (16), in order for a repair operation to take place which will restore the moved features in their original place.¹⁹ Hence K merges with E to form $F = \{ \langle T \rangle, \{ K, E \} \}$.²⁰ This combination of operations Chomsky (1995b) calls Move α as defined in (2). The application of Move α leaves a copy of the moved element in its original position and forms a chain $CH = \{ T, \underline{v} \}$, defined in terms of sisters of the moved element.

(17)



The model proposed in sections 2 and 3 is incapable of generating (12), or to be more precise it is not capable of operating on its own products and consequently it would predict that the word order in (12) is obtained by Merging 'man' to 'Which man killed John', and then Merging 'which' to 'man which man killed John'.

Obviously we are missing a fundamental generalisation that 'which man' functions as a constituent. Also the the merging of 'man' to 'which man killed' is not motivated by any instance of feature checking, since no wh-features are being checked²². Additionally, there would be two chains involved in the movement. One headed by the LI 'man' and the other by the LI 'which'.

Consequently, either the model presented above is incompatible with the data, or it is incomplete. At present, let us assume the latter, and see if it can be refined.

5. Two kinds of Merge

It is obvious that Merge must be allowed not only to operate on Lexical Items, but also on its own products. However, this requires that we postulate *a different* kind of Merge, since it would no longer interact with N_{HL} and would operate on its own products.

Let us call the two instance of Merge M_1 and M_2 .

(18)

Operation/Properties	Merge1(M1)	Merge2 (M2)
a. Operate's on LI's	+	+
b. Motivated by properties of the target	+	-
c. Motivated by properties of the head of the target	-	+
d. Forms a Chain	-	+
e. Cyclic ²³	+	+

M_1 merges elements in order to satisfy their selectional properties to be exact the selectional properties of the target. Hence, considering the example in (17), let us consider the stage of CH_L when the subject 'which man' is merged with \underline{v}^{max} . The operation Select selects $K = \{<which>, \{which, man\}\}$ and M_1 targets $B = \{<\underline{v}>, \{\underline{v}, \{<kill>, \{kill, John\}\}\}\}$ and merges K with B forming $D = \{<\underline{v}>, \{K, B\}\}$. Note that K is merged with B because of the Selectional properties of B and not of its label.²⁴ M_1 always applies at the root and is always cyclic.

M_2 is responsible for satisfying the feature checking requirements of elements, or to be more precise of their heads.

Consider the example in (17), the last stage of the derivation involves movement of the DP [which man] from $SPEC-T^{max}$ to $SPEC-C^{max}$.

M_2 targets $K = \{<which>, \{which, man\}\}$ and merges it with $G = \{<C>, \{C, F\}\}$ forming $H = \{<C>, \{K, G\}\}$.²⁵ Note that it is the [+wh] feature of C , that is the label of G which triggers M_2 . A copy of K is also left in its original position and a Chain $CH = \{C, \underline{v}\}$, which is defined in terms of *sisters* of the Merged element is formed.²⁶

Note that the application of M_1 will not always reduce the first index of the Numeration N_{HL} . the merger of a DP 'which man' with \underline{v} is an instance of M_1 . However, the Numeration as defined in (4) is not capable of capturing the generalisation that the q role is assigned by \underline{v} to the whole DP and not to *which* and *man* separately.

This introduces an unwanted redundancy in the system and we could return to the notion of a Numeration as defined in Chomsky (1995b). However, we shall see that the Numeration as stated in (4) will be useful in defining *head movement* as opposed to XP movement.²⁷ Hence, for the time being let us assume that the Numeration has the properties and structure as stated in Chomsky (1995b).

Note that we do not have to specify which instance of Merge operates on LI's only. If Merge selects an item from C_{HL} which has already received a θ role and merges it again for selectional reasons then the derivation will crash at LF.²⁸

We do, however, need a preference of one instance of Merge over the other. Returning to the derivation in (17) the DP [which man] is merged with \underline{v} . However, another possible analysis would involve the merger of the object 'John'. We are faced with a similar problem in constructions involving expletives which are discussed in Chomsky (1995b), where we have an option of either merging the expletive or moving (or performing M_2) the subject. Consider the following example:

- (19) a. There seems ___ to be someone in the room
 X_____X
 b. Someone seems __to be __in the room

There is a stage in the derivation in (19) where C_{HL} is faced with a choice of either deriving (19a) or (19b). However, when 'there' is in N_{HL} then (19a) is the only correct derivation. The question is what prevents the movement of 'someone' as in (19b). Chomsky (1995b,c) assumes that this is because Merge is preferred to Move, since the former is less costly. In our model, possible solution is to assume that C_{HL} is primarily concerned with the reduction of all indices in N_{HL} (as defined in Chomsky 1995b) to zero. To the best of my knowledge expletives are cross-linguistically Lexical elements, hence the merger of an expletive will always be preferred from any other operation.

The choice of subject merger over movement of the object to SPEC- \underline{v} is more problematic. Chomsky (1995b,c) assumes that Merge is less costly than Move α , which would account for the expletive construction and for the preference of subject merger over object movement.

We can stipulate something similar, and state that the application of M_1 is cheaper than the application of M_2 , since the latter entails chain formation.

Notice, however, that such a model does not differ conceptually from the proposals of Chomsky (1993). The difference is notational. Chomsky (1993) calls M_2 Move α . It seems that the difference between assuming a model proposed here where we have two different kinds of Merge and a model proposed by Chomsky (1995b) is the same as that of the approach in Chomsky (1993) and Chomsky (1995a, b,c). The approach here and the one adopted in Chomsky (1993) essentially constitute notational variants.

However, if there is feature movement in LF, then Move α is a reflex of a more general operation MoveF. The complexity of the rule then is not inherent, but a compromise resulting from the requirements of the sensory-motor system.

5. Terms

Bobaljik (1995) develops a model in which there is also only one operation Merge. He proposes that the operation Merge decreases the number of root nodes and increases the number of terms by one.²⁹

Consider the derivation of (12) repeated here as (20)

(20) Which man killed John.

The numeration is the same as in (13). The terms are:

(21) {which}
 {man}
 {killed}
 {John}
 {v}
 {T}
 {C}

Merge operates on the above terms creating new terms out of them. Hence the relevant steps of the derivation are (in order to save space I have called the initial set of terms in 21 α):

- (22) a. Merger of {which} and {man}=
 α
 {{which}, {man}}
- b. Merger of {killed} and {John}=
 α
 {{which}, {man}}
 {{killed}, {John}}
- c. Merger of {v} with {{killed}, {John}} =
 α
 {{which}, {man}}
 {{killed}, {John}}
 {{v}, {killed}, {John}}
- d. Merger of {{which}, {man}} with {{v}, {killed}, {John}} =
 α
 {{which}, {man}}
 {{killed}, {John}}
 {{v}, {killed}, {John}}
 {{which}, {man}, {v}, {killed}, {John}}
- e. Merger of {T} with {{which}, {man}, {v}, {killed}, {John}} =
 α
 {{which}, {man}}
 {{killed}, {John}}
 {{v}, {killed}, {John}}
 {{which}, {man}, {v}, {killed}, {John}}
 {{T}, {which}, {man}, {v}, {killed}, {John}}
- f. Merger of {{which}, {man}} with {{T}, {which}, {man}, {v}, {killed}, {John}} =
 α
 {{which}, {man}}
 {{killed}, {John}}
 {{v}, {killed}, {John}}

$\{\{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$

g. Merger of $\{\text{C}\}$ with $\{\{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\} =$

α

$\{\{\text{which}\}, \{\text{man}\}\}$
 $\{\{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{C}\}, \{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$

h. Merger of $\{\{\text{which}\}, \{\text{man}\}\}$ with $\{\{\text{C}\}, \{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\} =$

α

$\{\{\text{which}\}, \{\text{man}\}\}$
 $\{\{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{C}\}, \{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{which}\}, \{\text{man}\}, \{\text{C}\}, \{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$

There are two conditions which an output of the derivation must meet:

- (23) At any point of the derivation, select any term and feed it to the interface with the other computational components (LF, PF). This term must:
- a. include all the terms in the list,
 where a term τ includes a term ρ if ρ is a term or sub-term of τ ,
 and

- b. contain no unchecked features.

Chomsky's Greed/Last Resort is reformulated to apply solely to Merge:

(24) Last Resort:

Merge (α , β) is a legitimate operation iff Formal properties of either α or β are thereby satisfied.

As Bobaljik (1995:60) puts it, "the significant difference between this version and the version advocated by Chomsky is that the operation Merge for me is subject to this requirement as much as the operation Move (i.e. since they are not distinct).

Let us see how the above proposals cope with the preference of subject merger over object raising. The relevant point of the derivation is (22f) repeated here as (25):

(25) Merger of $\{\{\text{which}\}, \{\text{man}\}\}$ with $\{\{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\} =$
 $\{\text{which}\}$
 $\{\text{man}\}$
 $\{\text{killed}\}$
 $\{\text{John}\}$
 $\{\underline{\text{v}}\}$
 $\{\text{T}\}$
 $\{\text{C}\}$
 $\{\{\text{which}\}, \{\text{man}\}\}$
 $\{\{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$
 $\{\{\text{which}\}, \{\text{man}\}, \{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}, \{\text{John}\}\}$

The principles in (23) and (24) are also satisfied if we instead of merging $\{\{\text{which}\}, \{\text{man}\}\}$, we merge $\{\text{John}\}$ with $\{\{\text{T}\}, \{\text{which}\}, \{\text{man}\}, \{\underline{\text{v}}\}, \{\text{killed}\}\}$.

In this model we have no straightforward way of indicating the fact that elements which are embedded more deeply in the structure are less accessible to the operation merge. Using the

terminology of Chomsky (1995a,b,c), we have to ask why does T attract the subject DP in SPEC- \bar{y} and not the object DP. (23b) will not rule out such a derivation since T requires overtly any DP in its SPEC. Case and other features can be checked by the subject covertly. The solution adopted in Chomsky (1995b,c) is that T is *not strong* enough in certain languages like English to 'look down further' that the SPEC- \bar{y} position.³⁰

The condition (23a) also imposes on the derivation that it should be capable of seeing all terms of the at a given stage of the derivation, this, however, implies that at the end of the derivation just before Spell-Out each step of the derivation must be visible.

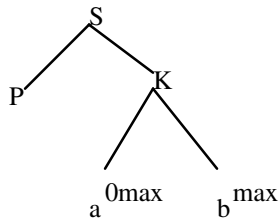
Note that the model proposed by Bobaljik (1995) there "is no need to appeal to a distinct operation Copy to relate one stage of the derivation to the next, or even to define complex terms. Even though the output term (or any complex term) may contain multiple copies of other terms, the effect of copying falls out from the simple mechanical application of (Merge)".

6. Sets

I would like to propose a similar approach to that of Bobaljik (1995), where the displacement property of language and the fact that there are dominance/precedence relations visible in linguistic structures can be accounted for with the help of one simple procedure.

Chomsky (1995a) assumes that the operation Select/Merge creates out of two terms a, b selected from the Numeration a set $K = \{ \langle L \rangle, \{a, b\} \}$. Let us take this definition literally, hence Merge creates a set out of two elements. This set has two members L, {a,b}. The second member is a two element set. L being the label of the set determines its properties (whether it is nominal, verbal, etc.) and either equals a or b. When you merge two sets the procedure is similar. You select a set K and a set P and you merge them creating $S = \{ \langle M \rangle, \{K, P\} \}$. M is again either K, or P which translates to being equal to either the label of K, or the label of P. Let us assume the former, hence $S = \{ \langle a \rangle, \{K, P\} \}$. This corresponds to the following tree:

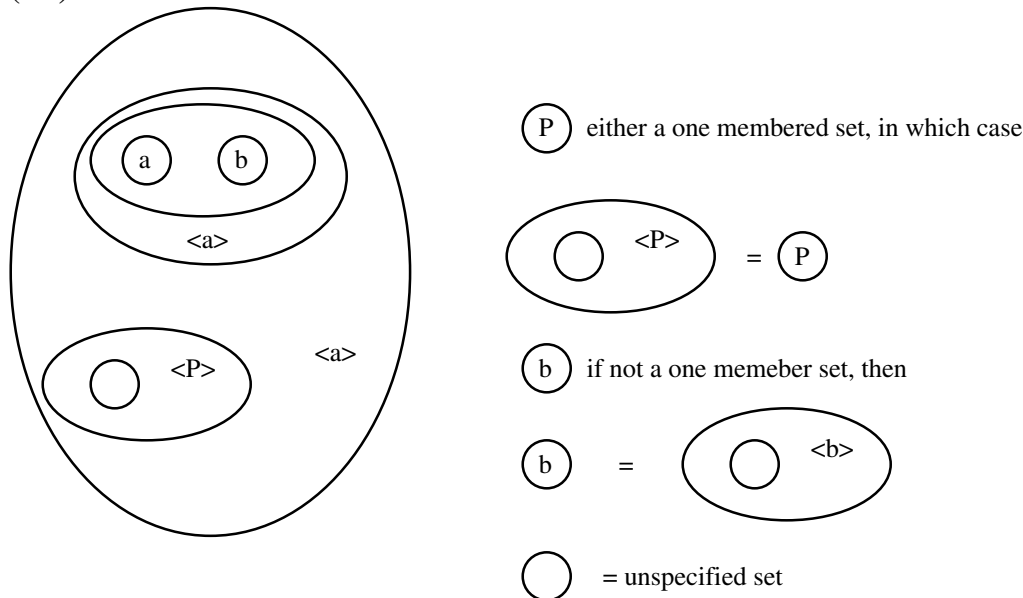
(26)



Note that instead of labels we have heads which project. The term b has to be a maximal projection as well as a minimal one. For example 'John' in (12). P can be a maximal projection, but it could also be a head.

If we translate (26) into sets, we obtain:

(27)



A tentative proposal which I would like to put forward is a compromise between the models adopted in Bobaljik (1995) and Chomsky (1995a,b,c).

Note that CHL is primarily concerned with building a structure which would meet the conditions imposed by the Bare Output Conditions.

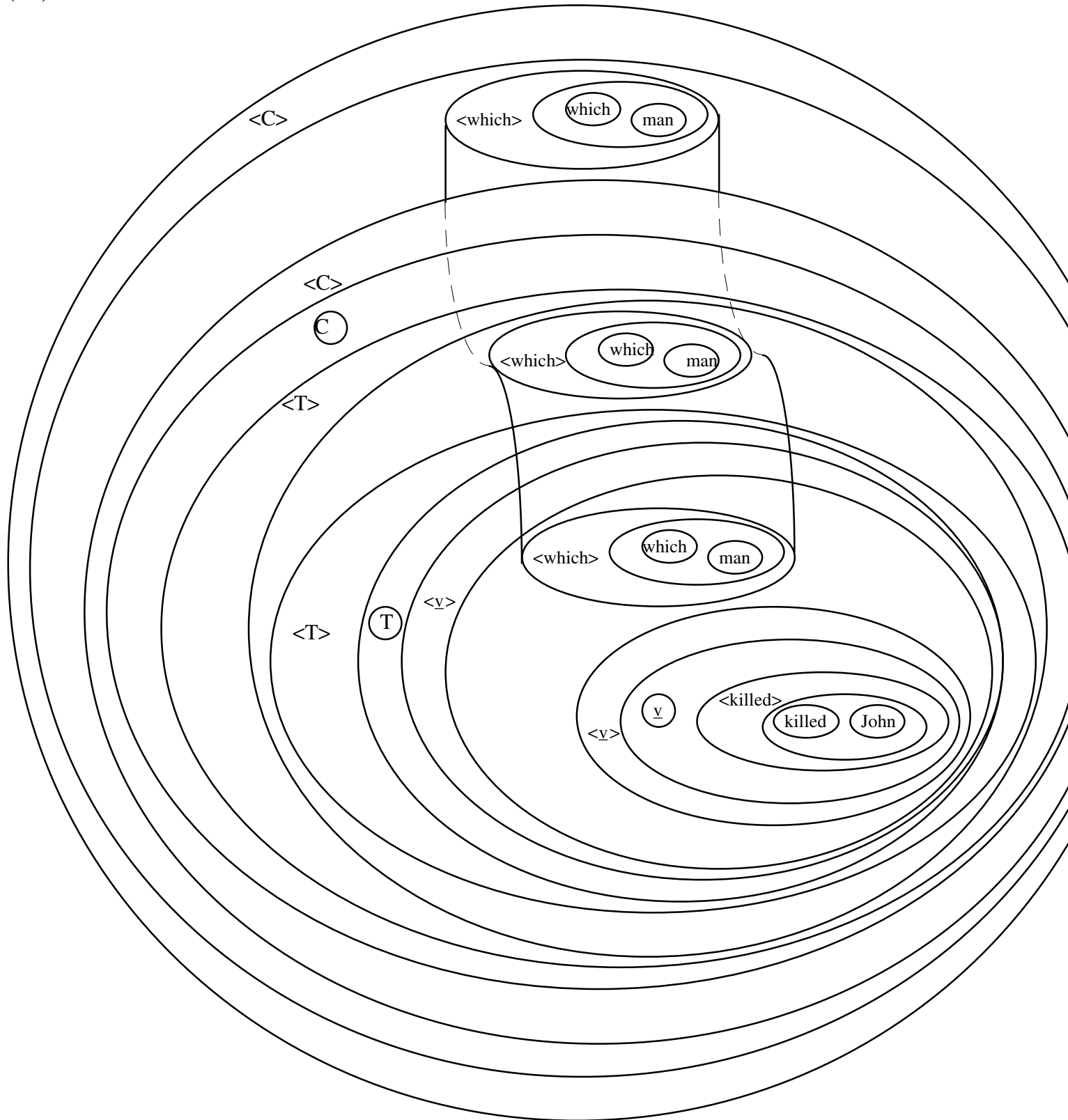
It has to build a structure which could after Spell-Out be read by the PF interface, such a structure primarily cannot have features of one lexical head (X^{0max}) dispersed, hence the pied-piping of features in the model adopted by Chomsky (1995a,b,c). Notably PF cannot read certain kinds of features which are considered - interpretable. These features have to be checked in order to be made invisible.³¹ The case features of T hence can be checked by the case features of a DP moved to it from SPEC- \bar{v} (see the analysis of "which man killed John" above). This gives rise to what Chomsky calls movement. There is not enough elements in

the workspace to satisfy the above condition, and hence things have to move, leaving a copy behind. Bobaljik (1995) proposes a model which would eliminate the need for an operation Copy, and consequently eliminate Move.

Looking at the model proposed by Chomsky (1995a,b,c) in terms of not trees, but of sets eliminates the need to have copies and automatically creates a link between displaced elements.

Consider a representation of the derivation of (12) in terms of sets:

(28)



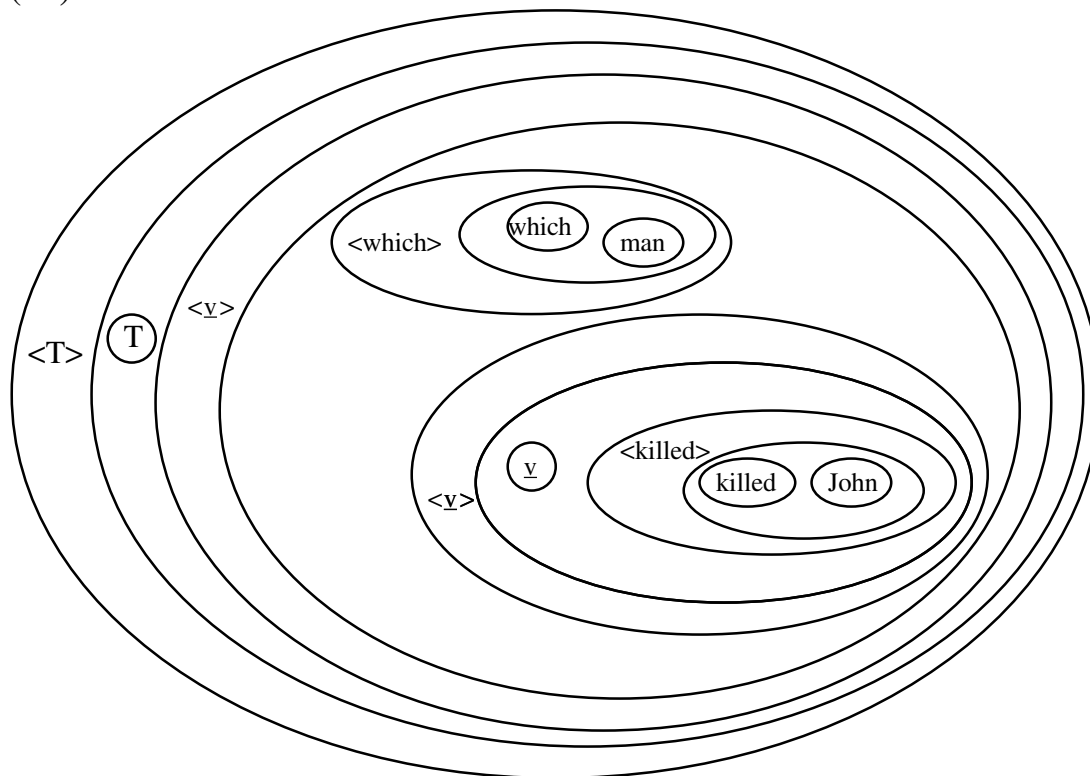
Note that (28) is equivalent to (17). However, I would like to propose that in fact the displacement property exhibited by all natural languages is caused by the fact that the set $K = \{ \langle \text{which} \rangle \{ \text{which, man} \} \}$ is present in three positions. However, there are no copies, or chains. The set is in those three places simultaneously. Each time in a given configuration the existing set has to be merged with K it does so. This can be motivated by the selectional properties of items, or the need to assign θ - roles, or it can be motivated by feature checking requirements of heads.

The first merger of K with the set $\{ \langle \underline{v} \rangle, \{ \underline{v}, \{ \langle \text{kill} \rangle, \{ \text{kill, John} \} \} \}$ is in order to satisfy selectional and θ - role requirements. The merger of K with $\{ \langle T \rangle, \{ T, \{ \langle \underline{v} \rangle, \{ \underline{v}, \{ \langle \text{kill} \rangle, \{ \text{kill, John} \} \} \} \}$ satisfies EPP and checks Case and Φ features of the head (which is the label) of the targeted set.

Note that as in previous analyses we run into the problem why did C_{HL} merge K and not the object John. The same problem arises when the computation is faced with a choice of either merging the expletive, or moving the subject as in (19).

If we adopt that the operation which selects a set which is already embedded in a larger set is more costly than an operation which merges separate sets. T for example attracts a DP, because of EPP.³² In English T is strong enough to attract the D features of K , but not strong enough to attract 'John'. Consider:

(29)

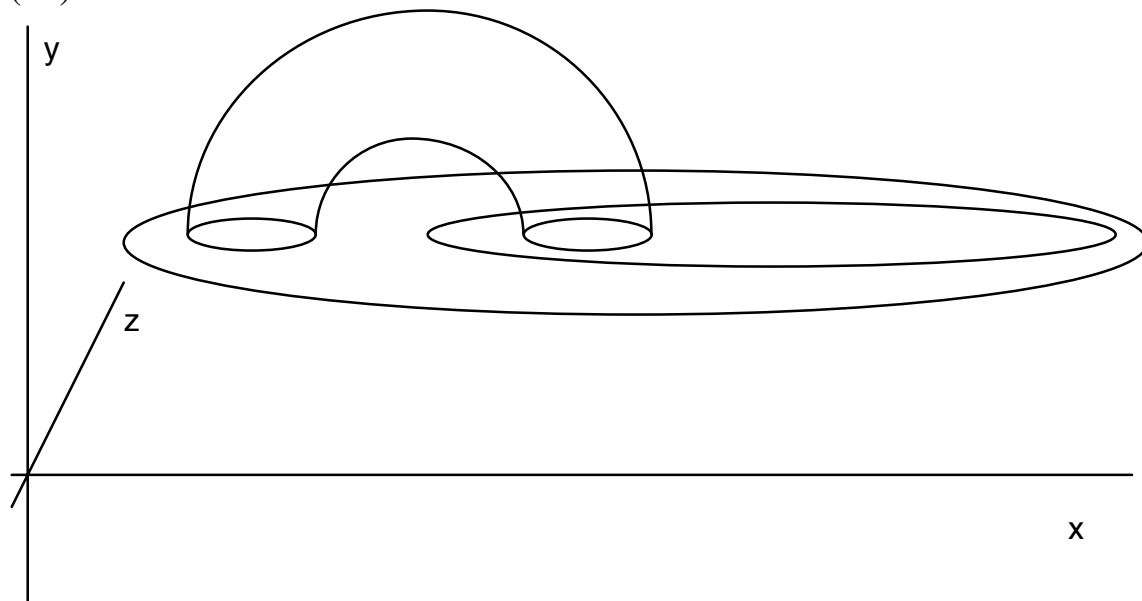


The shaded area indicates a set is not available for attraction

The analysis presented above crucially assumes that merger resulting from feature checking is triggered by the a head attracting a set containing the needed features.

Note that graphically the merger of subset with its superset can be denoted as extending the subset in a dimension $n+1$, where n is the dimensions occupied by the superset.

(30)



If we assume that merger of the type M_1 or what Chomsky simply calls Merge is an operation operating in 2 dimensions, let us say the plane X-Z, then M_2 or Move extends the set X into the dimension $n+1$, that is in our case 3-dimension depicted as X-Y-Z. Both operations are structure building and cyclic since both are in fact instances of Merge as defined by Chomsky (1995a,b).

7. Head Movement

Chomsky (1995c) suggests that head movement might be outside the Computational System. Head movement, for example is not cyclic (does not meet the extension requirement), it is semantically vacuous.

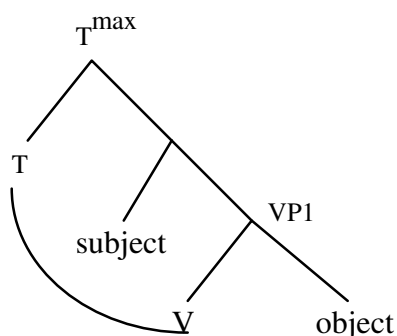
I would like to propose that Head movement is outside the model proposed in 6 since heads are not sets as defined in 6, and consequently cannot be subject to Merge.

I will assume that head movement operates post syntactically, but before PF and in essence is the kind of movement described in sections 2 and 3.

To be more precise there is an operation, like Merge, but which can access from the Numeration LI1 $\{0,y\}$ and merges it with LI2 $\{0,y\}$ forming not a set but a single object which has the properties of one of either LI1, or LI2.

Consider the case of French raising of V to T.

(30)



The V in the Numeration is listed as $LI_1 = V\{0,0\}$, it then merged with T, forming a single category $[V-T]$. In French this is done before PF, since the T has uninterpretable features which have to be checked before PF. Note that movement of V to T is in essence morphological, it is the affixation of T to V. What is different in V raising to T form affixation in the Lexicon is that this process leaves a trace and copy in the way it was proposed in sections 2 and 3. Hence, the relevant chain $CH_{LI} = V\{0,1\}; V\{0,0\}$.

What about LI's which are X^{0max} and X^0 at the same time, for example 'John' in (12). Note that if the object were to merge with C for example the result would be a new set $\{<C>, \{C, John\}\}$. The output of the derivation would suggest that this is not head movement, however, John is also an LI, hence could be an input to head movement.

However, if we assume that Lexical Items can function as sets, then we can eliminate this ambiguity. Let us adopt the following condition:

(31) $LI\{i,y\}$, $i \neq 0$ is understood to be a set $L = \{, \{LI, \emptyset\}\}$

Interestingly enough (31) allows us to eliminate the difference between M_1 and M_2 in one respect, namely both operations take as their input sets and create sets which are not subsets of any other set.

This is a considerable step towards eliminating the need to have two different structure building operations as stated in (18). Notably if we assume that a single operation Merge can be triggered of by two kinds of attractors: heads, and sets, then we are able to stipulate the existence of one operation Merge. Its properties are:

(32) The operation Merge is a procedure which:
a. takes as its input any pair sets: A and B, where

- either the head (identified by the label) of A has a feature to be checked by the head of B, or
 - the set A has to Merge with the a set of the type B (type = label) in order to satisfy selectional requirement or to enter into a θ marking configuration.
- b. The output of the merger of the sets A and B is a set $S = \{ \langle L \rangle, \{A, B\} \}$. L = label of S, and L= label of A or B.
 - c. S cannot be a subset of any other set.
 - d. Merge can merge A with $S = \{ \langle L \rangle, \{A, B\} \}$ forming $H = \{ \langle L \rangle, \{A, S\} \}$. Such merger entails that the a subset of S can merge with S to form a new set. The position of the set A is described in terms of sisterhood. A is a sister of B when A and B are merged to form S. A is also a sister of S, since A and S are merged to form H.
 - e. The merger of A with B where A is a subset of some set S such that S is not a superset of B, involves the creation of a chain CH defined in terms of sisters of A. The notion Chain is understood here as the extension of the set A into the dimension $n+1$, where n is the dimension of sets B and S.
 - f. The operation in e also requires that the label of the set S is strong enough to differentiate its own internal structure.³³
 - g. From e and f we can conclude that The merger of a set A with a set B, where neither A or B is a subset of any other set is less costly than the operation in e.

8. Conslusions

It is far from obvious whether the proposals put forward here will turn out to be a better and more minimal solution to the conditions imposed by the Bare Output Conditions.

I have not touched many problems posed by my proposals in sections 7 and 8. For example, the problem of Relativised Minimality, or the nature of covert movement, nor have I pursued the notion of c-command, so prevalent in works which have a similar goal of minimising the amount of operations in CHL (see Epstein 1995, Groat 1995, O'Neil 1995).

I have not also touched upon the difference between approaches which assume that the displacement property of languages is a result of an operation Attract, or Move. The model I propose seems to fit in more with an approach where the target attracts the displaced element.

¹ I would like to thank Noam Chomsky and Sam Epstein for their invaluable comments.

²In referring to Chomsky (1995b) I primarily refer to Chapter of the book.

³Hence, for example in English when the D features move to T (Move F), the Case and Φ features also move.

⁴From hereon when referring to Move, I will mean overt movement.

⁵In principle Merge is 'costless', however, the application of Merge is not (Chomsky 1995c).

⁶Following Chomsky (1995b:235): "If α has a strong feature F, it triggers an operation OP that checks F before the formation of a larger category that is not a projection of α . The operation OP may be Merge or Move."

⁷It is not clear, if there is any overlap between Checking and θ -role configurations between LI's. However, this seems to be the case when an expletive is merged (as opposed to being moved) in order to satisfy EPP

⁸I assume following Chomsky (1995b,c) that it is the target that always projects, hence for the time being I will assume that the 'saw' is the target in (5a) and α in (5b). I also put aside the problems connected with Adjunction.

⁹I assume below that the wh-phrase checks its features off with β . This is, of course a oversimplification, but one which has no bearing on the argument. The need for something in between β and γ is motivated by, among other things, the standard assumption that V is not the head of γ .

¹⁰I am not specifying if it is C-selection, S-selection, or some other mechanism.

¹¹Where pied-piping and the repair operation are motivated by BOC's. The cyclicity of Move α in Chomsky (1995b) is a result of the pied-piping of Phonetic features.

¹²Where relevant I assume a most recently adopted framework, mainly (Chomsky 1995b, (Chapter 4) and Chomsky 1995c).

¹³Chomsky (1995b,c,d) adopts the framework where a can merge with b to form $K = \{ \langle c \rangle \{ a, b \} \}$, where c is the *Label* of K indicating its properties. $c=a$ if a projects, c is the Head of K. Note that this is irrelevant of whether Merge is triggered by Move/Attract, or by Merge/Select.

¹⁴Notice that the selection of 'man' is also theoretically possible. However, such a derivation would crash, since the selectional properties requiring that 'man' be a DP would not be satisfied in the derivation.

¹⁵The preference of Merge over Move is obvious if we assume that Move consists of Merge and other operations.

¹⁶In order to satisfy EPP, check case and ϕ features (see Chomsky 1995b).

¹⁷An interesting question is why does not T select only 'man'. Unfortunately, I do not have a straightforward answer to this, recent research seems to indicate that this is connected with the definiteness of the DP as opposed 'man' itself.

¹⁸As stated previously, T requires a D feature (EPP), however, Chomsky (1995b,c) assumes that the Case and Φ -features move along for free, because they can be checked.

¹⁹This is an enormous oversimplification of the procedure and the problems associated with it. The nature of the repair operation is not syntactic. Note that the whole phrase 'which man' has to be pied piped to SPEC-T and not just 'man', since the D feature of T (EPP) was checked by a feature from 'which'.

²⁰To be more precise, $K = \{ \langle \text{which} \rangle, \{ \text{which}, \text{man} \} \}$ minus the moved to T^0 features.

²¹The format is simple a merges with b to create $\{ \langle c \rangle \{ a, b \} \}$ where c is the label of the set (understood as either a, or b depending on which projects).

²²This is more obvious in cases where we move more complex DP's like: 'Which man who you saw yesterday with your brother in Madrid', where obviously it is highly unlikely that each LI in the above phrase moves separately.

²³I assume that by being cyclic an operation meets the Extension Condition, as stated in Chomsky (1995b).

²⁴I will use the term 'label' meaning 'head'. The difference is only significant in cases of adjunction, which I do not discuss here.

²⁵A same procedure is adopted for the merger of K with T in order to satisfy EPP.

²⁶This is following recent proposals in Chomsky (1995c). The definition of Chains for M_2 and for Move α is identical

²⁷In recent works it will be X^{0max} movement and X^{max} movement respectively.

²⁸It is also possible to postulate that all the LI in N_{HL} have to be selected by C_{HL} by the end of the derivation, or it crashes.

²⁹Bobaljik defines terms in a similar fashion to Chomsky (1995a,b):

(i) For any structure K:

K is a term of K,

if L is a term of K, then the members of L are terms of K (Bobaljik 1995:fn. 6)

³⁰This a very crude oversimplification. The crucial assumption is that T universally (EPP) licences one SPEC position which has to be filled by a DP.

³¹There is a difference between feature erasure and feature deletion. See Chomsky (1995b, Chapter 4)

³²For the sake of clarity, I will use the tree notation whenever possible.

³³The notion strength is very vague, however, intuitively it refers to the morphological strength of a given set, or to be precise its label, to re analyse its own structure, hence attract its own subset with the relevant features. The set in English $J = \{ \langle T \rangle, \{ T, \underline{v} \} \}$ is able to only analyse the structure of the subset \underline{v} , and hence T can only merge with the subject = Su, since $\underline{v} = \{ \langle \underline{v} \rangle, \{ Su, \underline{v} \} \}$. This is identical to the proposals in Chomsky (1995b,c) where T can only attract the nearest DP.

References:

- Bobaljik, J. D. 1995. In terms of Merge. In Pensalfini R. and H. Ura (eds) *The MIT working Papers in Linguistics*. Volume 27. Cambridge, MIT. 41-64.
- Chomsky, N. 1993. A minimalist program for linguistic theory. In Hale, K and S.J. Keyser (eds) *The view from building 20*. Cambridge, Mass.: MIT Press.
- 1995a. Bare phrase structure. In Webelhuth, G. (ed) *Government and Binding in the Minimalist Program*. Oxford: Blackwell.
- 1995b. *The minimalist program*. Cambridge, Mass.: MIT Press.
- 1995c. Class lectures. Cambridge, Mass.: MIT, Autumn 1995.
- Epstein, S. 1995. The derivation of syntactic relations. Ms Harvard University
- Groat, E. 1995. On the redundancy of syntactic representations. Talk at GLOW, Tromsø.
- O'Neil, J. 1995. The nature of syntactic structure. In Samuel d. Epstein, Hoskuldur Thrainsson and Steve Peter (eds) *Harvard Working Papers in Linguistics*, Vol.5, 93-111.

Why Move ? - main ideas:

1. Chomsky (1995) proposes that:
 - a. Syntactic structures are formed by the operation Merge which creates a Syntactic Object SO_{ij} out of two Syntactic Objects: SO_j and SO_i .
 - b. the displacement property of language is triggered by the fact that certain features of heads (H1) are -interpretable at PF, and hence have to be checked off by compatible features of a different head (H2) by moving to H1. This carried out by Move F.
 - c. PF cannot read constructions with displaced features as in b, hence the remaining sets of features pied pipe targeting the maximal projection of H1 and merge with it forming a Spec-H. The moved to H1 feature reconstruct back to H2, and the structure is readable at PF.
 - d. The movement of H2 leaves a copy of H2 in its original position.
 - e. steps b-d constitute the operation Move α
 - f. Move α forms a chain CH connecting the displaced by Move α element with its copy. CH is read at LF, and is necessary to prevent the original and copy receiving different q roles.
 - f. Merge and Move are two different operations. The latter is sensitive to the feature configurations of heads. Merge is not, with the exception of expletive merger to satisfy EPP.
 - f. Merge is cyclic, Move α is also cyclic by virtue that it has in it as a sub-operation Merge and pied-piping.
 - g. Merge is preferred to Move α , since it is cheaper. Hence constructions where we have a choice of either merging a subject (for example), or moving an object, the former always block the application of the latter.

I. Bobaljik (1995)

- a. Only one operation is responsible for creating syntactic structures and for the displacement property of language.
- b. The operation Merge applies to terms creating new terms out of them.
- c. The output of this operation has to satisfy the condition that no features are left unchecked.
- d. The output of this operation has to also satisfy the condition that the spelled out term contains all the previous terms in it.
- e. There is no operation creating Copies, since the set of initial terms (called a Numeration in Chomsky (1995)) can be accessed by Merge recursively. Merger can also re-access terms created by it.
- f. Chains are defined internally to each term, hence each step of the derivation defines a new term.
- g. Head movement?

3. The proposal contained herein.

- a. Syntactic operations apply to sets.
- b. There is one operation Merge responsible for creating syntactic structures. It takes two sets, or LI's and creates a new set.
- c. Merge creates sets out of LI's provided by the Numeration.
- e. The Numeration is a set NHL consisting of sets $\{LI, \{y, i\}\}$. LI= Lexical Entry,
 - i=the number of times a given LI occurs in θ -role, or selectional configuration,
 - y=the number of times a given LI occurs in a Checking configuration
- f. Merge operates on two kinds of elements, Lexical Entries, or sets.
- g. Merge which operates on sets and LI's = M2.

- h. Merge which operates on LI's = M1.
- i. Both M1 and M2 have the same output: sets which cannot be subsets of other sets. When merging x and y a set S is created, where $S = \langle H \rangle, \{x, y\}$. H is the label of the set, and determines its properties. H = either x, or y if x and y are not sets themselves. If they are, then H = the label of x, or y.
- j. M1 reduces the index i in NHL by 1, but it cannot access LI once $i=0$.
- k. M2 can access the same set K, n-times, where n is the number of times K is needed in order to check - interpretable features of the head H1 = label of the set T which is targeted.
- l. H2 attracts K which has a head H2 which contains features which have to check off with the features in H1.
- m. When K is a proper subset of T then M2 can access K which is a subset of some set R, which is a subset of T, and make K a subset of a set U which was created by merger of K and T.
- .
- n. There is no operation Form Chain
- o. There is no operation creating copies
- p. The position of a given set is established on the basis of sisterhood. Two sets are sisters if both sets A and D are subsets of the set Q created by the merger of A and D. A can also be a subset of D.
- q. The difference between M1 and M2 can be eliminated if we assume that Merge operates only on sets and that LI $\{i, y\}$, $i \neq 0$ is understood to be a set $L = \langle LI \rangle, \{LI, \emptyset\}$
- r. There is no movement of LI $\{0, y\}$ in overt syntax. It is Carried out post-syntactically either at PF or during Spell-out.
- s. The movement of LI $\{0, y\}$ (let us refer to it as head movement) alters the y index in the Numeration by 1, and thus creates a chain: CH1 = LI $\{0, \max\}$; LI $\{\max, 0\}$

