Published in <u>Yearbook of South Asian Languages and Linguistics</u>, 2002, Sage, New Delhi/ London

Minimal Look-Ahead

Tanmoy Bhattacharya
Universität Leipzig
Zentrum für Kognitionswissenschaften
Brühl 34-50
D-04109 Leipzig
E-Mail: *bhattach@rz.uni-leipzig.de*

The following note points out a fallacy in one of the basic formulations of the Minimalist Program (MP) especially embodied in the later Chomsky manuscripts, Minimalist Inquiries or MI (1998) and Derivation by Phase (1999). As far as I am aware, this problem has not yet been widely noticed. However, a generous reading of this particular formulation (see (1)) that is attempted in this note in terms of the new concept of minimal look-ahead (ML-A) also opens up and addresses another deep-rooted problem of the MP: finding a trigger for the operation Merge. This latter problem is widely recognized as will be amply evident from the discussion of the relevant literature that follows and by Chomsky's own attempts (in MI) to provide a featural trigger for Merge (though inadequate) discussed in section 2.4.

The importance of this note lies in the simple fact that any sophisticated scientific field of enquiry (such as the MP) must resolve these basic fundamental theoretical issues (Merge being *the* basic building block of the MP) before making any empirical claims. The attempt at a "solution" in this note therefore is clearly devoid of any empirical implications. Consequently, the discussion by definition cannot be restricted to any particular language group or area.

1 A problem

Let us first introduce the problematic paradigm. Chomsky (1998: 14) identifies (1a,b) as steps a language L follows in order to specify the language and (1c,d) to derive a particular expression EXP:

(1)a. Select [F] from the universal set F

b. Select LEX, assembling features from [F]

c. Select LA from LEX

d. Map LA to EXP, with no recourse to [F] for narrow syntax

Minimalist Inquiries (MI) is a sophistication of the program designed to reduce architectural complexity of the overall model. Thus language L maps a subset of features [F] constructed out of the universal feature set F to a set of expression EXP by *one time selection*. Complexity is further reduced if L involves a one-time operation that "assembles" elements of [F] into a lexicon LEX. (This is revised as [F_L] in the 1999 (p7) manuscript but essentially the import remains identical.) A language L therefore maps ([F], LEX) to EXP. Access to this domain is further reduced by suggesting that [F] is not accessed at all in the rest of the steps which involve computation to LF, only LEX is accessed.

Reduction of complexity is thus achieved by making each step a one-time selection process. I see the following problem with this view. If steps a and b in (1) specify the language then it is not clear why a reduction in "operative complexity" is an issue. If it is an issue, then it must be assumed that both language change and language acquisition, both of which set as a goal of achieving a certain state of language (synchronic and acquisitional), proceed in such a fashion (i.e., by respecting steps a and b). It is far from certain that we are anywhere near such a unificatory treatment of the state of language reached through language change and achieved by acquisition. Step b therefore seems unnatural. Furthermore, without further modification (even when adequately footnoted) "assemble"[1] in step b must imply an umbrella operation for various smaller sub-assembles. Assemble in such a view is only *indirectly* a one-time operation.

[1] This idea of assemble incorporates the idea of assembling of lexical items from the numeration into different derivational space of Uriagareka (1999). His idea of "parallel" movement, based in turn on Nunes (1995), whereby prior to the merging of [*the man*] and [*saw* [*a woman*]] in (i), the subtrees must be assembled in separate derivational spaces.

(i) Numeration {*the*, *a*, *man*, *saw*, *woman*, ...}

{*the* {*the*, *man*}} and {*saw* {*saw*, {*a* {*a*, *woman*}}}}

However, my criticism of assemble in the MI framework extends equally to this model. In other words, assemble seems to need additional work of collecting the relevant LIs at one place. Instead, direct mapping from the Numeration of these subtrees will not take away a great deal from the theory proposed here, unless

Both of these objections about step b point towards a much more damaging question: why does the language system involve this elaborate array of steps instead of going directly from F to LA? That is, if one were playing god, why should one go through this complicated set of steps in order to endow the human species with Language?[2] There are two ways of responding to this question: the Evolutionary way and the Minimal Look Ahead way. I briefly discuss the former directly below but opt for the latter in section 2.

The evolutionary response has to do with the simple fact that going directly from F to LA is generative semantics. After such definitive parting of ways in the mid-60s, admitting to F to LA computation would imply that generative semantics was right all along. As we know, things went otherwise and (F,LEX) is really lexicalism in vaguely formal clothing. So, although the issue raised so far in this note with respect to the foundational steps in (1) may seem like making a mountain out of a molehill, deconstructing the way such formulations come about may in fact prepare us in recognizing ulterior shapes linguistic theories take.

The point to note about the lexical array LA in step (1d) is that it determines a PHASE as derivation of EXP proceeds by phases (or chunks of structures which are "propositional" and therefore "complete" in some sense). The same evolutionary (i.e. in terms of how the theory evolved) problem reappears here as it is not clear how if derivation proceeds by phases, computational complexity is reduced. Since phases seem to be required for a derivation to proceed optimally, they cannot be constructs of the lexicon but rather of the numeration.[3] If it were a question of reduction of complexity we should go the generative semantics way. That is, F to LA would be the simplest way to proceed.

---

of course it can be clearly shown that assembling is an essential prerequisite of building a Phase.

[2] The god metaphor came up while discussing with Juan Uriagereka the issued raised in this note, which has benefited much from the exchange we have had in sharpening the problem. However, he should not be held responsible for any mistakes in the "solution" in the form of Minimal Look-Ahead advanced here.

[3] Although there have been no visible efforts among the practitioners of the theory to address the issue (due perhaps to an overrated formal difficulty of admitting into the theory of certain pragmatic/ discourse terms), numeration may be roughly seen as speaker's intentions. This too then points towards an obvious globality within the various stages of a computation.

2 A solution?

Instead of the "evolutionary" reading of (1), it is possible to show (as I do in this section) that MP cannot avoid a certain notion of globality, which I identify as Minimal Look-Ahead (ML-A). In this connection, note that with regards to step d, a final reduction is suggested (Chomsky 1998:19-20) in terms of access to the LEX: at each stage of the derivation a subset LA_i is extracted out of LA and is placed in active memory (or the "work space" of the derivation). When LA_i is exhausted the computation may proceed if possible or it may return to LA and extract LA_j to continue. This, in principle, is different from the reduction observed in other steps as this alone involves multiple access to LA. Chomsky notes (p20) that "operative complexity *in some natural sense* is reduced" (emphasis mine).

Although it is not stated, this asymmetry, to my mind, implies that the language faculty must incorporate a version of look-ahead at some point to allow for the property of recursion in human language. Although this reduction in complexity is motivated by the concern to reduce the derivational space, it seems to also allow a weaker form of look-ahead by allowing multiple access to LA. In the rest of the paper, I show that ML-A is needed also to solve another deep-rooted globality problem in minimalism, related to the operation Merge. The "solution" offered exploits the inherent globality (ML-A) implied in Chomsky's system to provide a trigger for Merge.

2.1 Problems with Merge

Since Merge is a syntactic operation, it must be triggered. Merge is a basic operation in the MP whereby phrase structures are built up piece by piece as the computation proceeds.

(2) Merge $(\alpha,\beta) = \{\alpha,\beta\}$

The operation 'Merge' constructs a new syntactic object out of the pair (α, β). The operation is asymmetric, projecting either α or β. The element which projects becomes the label of the complex. In general, the new syntactic object must be of the form $\{\gamma, \{\alpha, \beta\}\}$, where γ identifies the type to which it belongs. Notice crucially that the order of the merged elements is irrelevant in this proposal. The notation {α, β} in (2) states precisely that. This may not be sufficient as the operation fails to identify the head of the derived constituent. Although finding the head is not an operation, it is assumed that it is calculated automatically at the time the constituent is formed by Merge – one simply finds the head from one of the daughters. Consider the following derivation from Collins (1997: 64) to see this more clearly:

(3)a. Select V

b. Select N

c. Merge (N,V) = {N,V}

Head ({N,V}) = V

d. Select Agr_o

e. Merge (Agr_o, {N,V}) = {Agr_o, {N,V}}

Head ({Agr_o, {N,V}}) = Agr_o

If instead, at (3c), N was chosen as the head, at LF we would have an NP with a V complement, which is assumed to be uninterpretable at this position. Such an assumption rests on a grammar model with a look-ahead facility which, according to Chomsky's manuscripts, is presumed to inflate the complexity of the computational component of the grammar. This too (i.e., apart from ML-A inherent in multiple access to LA), then points towards the direction of ML-A. But first, I will show how Collins' principle of INTEGRATION which is supposed to deal with the headedness problem falls short.

2.2 Integration

Consider the following partial derivation of *John left*:

(4)a. Select *John*

b. Select *left*

c. Merge (*John*, *left*) = {*John*, *left*}

The question that we have raised is what motivates the Merge in (4c). It is unlikely that a feature of either *John* or *left* is being checked through Merge. One possibility is that in selecting either of the two Lexical Items or LIs, a property of the LI concerned is being satisfied, namely, the property of being taken out of the Numeration (and consequently its associated integer reduced by one). This is rejected by Collins on the grounds that if two phrases (and not LIs) are merged, no appeal to the Numeration is made.

He assumes the alternative that Merge of α and β, whether lexical or not, is driven on the basis of the fact that both must be integrated into the clause. He calls this trigger for Merge, INTEGRATION, which involves no feature checking and defines it as follows:

(5) Every category (except the root) must be contained in another category.

(Collins,1997: 66)

The problem with the definition of root (a category not contained within any other category) is not addressed in Collins. Without such a definition, Integration as stated above is not meaningful. What I attempt in the rest of the paper is to provide a definition for ROOT (in terms of ML-A, the main construct of this paper) in order to make Integration a plausible trigger for Merge. However, Collins' non-featural view of Merge is an advancement which is ignored by Chomsky (1998). I offer a short critique of Chomsky's *new* formulations before returning to Integration in 2.4.

2.3 Merge in Chomsky (1998)

In this subsection I will briefly discuss the status of Merge (and its trigger) as presented in Chomsky (1998: 49-51). As will be clear from the discussion, the notion of Merge advanced in this refinement to MP is a feature-based, selectional view which has been independently shown by Collins (1997) to be inadequate.

As we have seen, one of the conditions for Merge is that it must perform the operation of constructing a new object. However, as in (2) above, there is no information available about the label of this new construct. Chomsky considers the possibility of predicting the label of a merger as follows.

First, he distinguishes between SET-MERGE for merger by substitution and PAIR-MERGE for merger by adjunction. Adjunction is inherently asymmetric (X is adjoined to Y) and leaves the category adjoined to unchanged. It is easy to see that pair-Merge of α to β will project the target β. Set-Merge as an operation is symmetric, so either label may project.

The result is either interpretable at LF or not. Notice, however, that such a formulation implies globality as part of the language design since Merge proceeds in the manner dictated by the success of the derivation at LF. In discussing the problems with Merge, we have rejected this approach as increasing the complexity of the computation. I think Chomsky sees this and attempts the following modification.

Set-Merge also has an inherent asymmetry since α, β merge in order to satisfy selectional requirements of one of them (the *selector*) but not both. Chomsky observes that the selector is *uniquely* determined (emphasis mine). In particular he opts for a featural account for Merge triggers. A feature F of one of the merged elements in $\{\alpha,\beta\}$ must be satisfied for the operation to take place. F is in the label of the selector and the label of the selector projects.

In conclusion, in the case of the asymmetric operation pair-Merge there is no selector whereas set-Merge has a unique and obligatory selector which determines the label of the construction. However, this selectional view is inadequate as shown below.

2.4 Lexical Integration

Note that the Integration approach presented earlier offers a non-featural trigger for Merge in contrast to the approach in Chomsky which crucially takes into account an F feature of the selector. Watanabe (1995) presented a version of Integration which pre-empts a selectional view as follows:

(6) Lexical Integration

Every constituent must either

(i) be dominated by another constituent, satisfying the lexical selectional property of the head of the dominating projection or

(ii) dominate every other constituent (the root)

In (6), selectional properties drive Merge. Collins (1997: 74) clearly demonstrates that a selectional view cannot explain what selectional properties, if any, are satisfied either by the DP or the V’ in the configuration (8) which is an intermediate representation of the VP in (7):

(7) *John threw the ball to Mary*

(8) [$_{VP}$ DP [$_{V'}$ V PP]]

If theta-role assignment is purely an interpretive process (i.e., operating at LF) we cannot accept the view that theta-roles drive the operation Merge. Similarly, if Case feature checking is also a property of the LF interface, then Case features of the DP cannot drive Merge either.

In particular, a selectional view of Merge appeals to an interpretable F feature, and thus cannot rid the system of the problem of look-ahead, since the theta-criterion, which appeals to lexically encoded semantic features of a head, is an LF interface principle. This being a deep rooted problem of MP, the best way to deal with which is perhaps a feature-based theta-theory as in Manzini and Roussou (1997) or Hornstein (1999).

Integration as defined in (5) above can take care of this. However, in order to adopt Integration, we have to first address the problem of defining a ROOT, topic of the next section.

3 Minimal look-ahead

This note suggests that one possible line of approach in defining the root could be in terms of look-ahead. Contrary to the attempts in Chomsky (1998, 1999) of reducing it, if look-ahead cannot be avoided then root is the point where there is no more look-ahead. The asymmetry pointed out earlier in connection with multiple access to LA, indicates the possibility of incorporating a certain amount of look-ahead in the grammar. It will be nice therefore to find a way to define ROOT in terms of this inherent globality.

However, fomalizing minimal look-ahead is no easy task. A possible first step in this direction is the following:

(9) A node R is a ROOT iff there is no other node N which has zero minimal look-ahead or, there is no other node N which is less embedded than R[4]

R is a ROOT $\leftrightarrow \neg\exists$ N | N$\neq$R $\wedge$ ML-A (N)=0[5]

[4] I adopt here the definition of Frank, Hagstrom and Vijay-Shanker (2000):
a node x is LESS EMBEDDED than a node y iff x does not c-command y, and every node which c-commands x also c-commands y.

However,

(9) does not distinguish between adjuncts and non-adjuncts.[6] That is, it cannot differentiate between the ROOT A and the adjunct B in $[_A B\ [_A C\ D]]$. So we need to cast (9) in terms of CATEGORIAL ROOT defined as in Frank, Hagstrom and Vijay-Shanker (FHV) (2000) as a node which does not c-command any other node. They identify (a single) adjoined structure as doubly rooted, one being a categorial root and the other the ATTACHMENT ROOT, which is not c-commanded by any other node and which identifies the site of cyclic attachment. The attachment root is therefore the one which does not project.

In terms of the notion of minimal look-ahead, the categorial root on the other hand would imply a position with zero ML-A.

(10) R is a categorial root iff minimal look-ahead at R is zero

R is a CATEGORIAL ROOT $\leftrightarrow$ML-A (R)=0

(10) has the following corollary:

(11) N is not a categorial root iff N has minimal look-ahead or N is an adjunct.

N is $\neg$ CATEGORIAL ROOT $\leftrightarrow$ ML-A (N)$\neq$0 $\wedge \neg \exists$ M $|$ M C N[7]

Finally, given the proposals put forward here, categorial root determines the unique root satisfying a well-formedness condition like (12) which is based on FHV's Category Identity Condition.

(12) Root Uniqueness Condition

A well-formed tree has a unique root with zero ML-A.

All other nodes must display ML-A in order to drive the computation.

---

[5] Indicating location in a graph where a particular property holds or does not hold seems to be quite problematic for set theoretic notations. Although Tree Modal Logic or Lattice Theory might be better equipped to deal with this, I have opted for treating ML-A as a function ranging over values from zero to n.

[6] For a theoretical discussion of non-adjuncts as the Main derivational Structure or MDS which are Incomplete Phases providing the "plugging address" for spelled-out phases to plug in during the course of the derivation, see Bhattacharya (2001).

[7] This is also the definition of an attachment root (i.e. there is no node M such that M c-commands N).

References

Bhattacharya, Tanmoy. 2001. Peripheral and clause-internal complementizers in Bangla: a case for remnant movement. *Western conference on linguistics* **12**, ed. by Vida Saamian. Fresno, CA.

Chomsky, Noam. 1998. Minimalist inquiries: the framework. *MIT occasional papers in linguistics*, Cambridge, MA: MIT.

Chomsky, Noam. 1999. Derivation by phase. *MIT working papers in linguistics* 18. Cambridge, MA: MIT.

Collins, Chris. 1997. *Local economy*. Cambridge, MA: MIT Press.

Frank, Robert, Paul Hagstrom and K. Vijay-Shanker (2000). Roots, constituents, and C-command. John Hopkins University and University of Delaware, ms.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30. 69-96.

Manzini, Rita and Anna Rouseau. 1997. A minimalist theory of A-movement and control. University College London and University of Wales, ms.

Nunes, Jairo. 1995. The copy theory of movement and linearization of chains in the Minimalist program. College Park, MD: University of Maryland dissertation.

Uriagereka, Juan. 1999. Multiple spell-out. *Working Minimalism*, ed. by Samuel David Epstein and Norbert Hornstein, 251-282. Cambridge, MA: MIT Press.

Watanabe, Akira. 1995. The conceptual basis of cyclicity. *Papers on Minimalist syntax*. *MIT working papers in Linguistics* 27, ed. by Robert Pensalfini and Hiroaki Ura.269-291. Cambridge, MA: MIT.