Implications of a Restrictive Theory of Transformations  
H. Lasnik  
April 1981

I. The problem: What is the nature of L.A.D. such that the following procedure can take place?

1) **DATA (D) → L.A.D. → GRAMMAR (G)**

2) We would expect that L.A.D. provides only a limited class of possible grammars.

If any logically possible grammar were available to L.A.D., language acquisition could not take place in the way that it does. The data would drastically underdetermine the grammar, and a variety of children in the same linguistic environment could arrive at mutually inconsistent G's which all agree on the presented data. There is no reason to believe that the situation in 3 occurs in any substantial way.

3) $D \rightarrow L.A.D. \rightarrow G_1$  $D \rightarrow L.A.D. \rightarrow G_2$  $D \rightarrow L.A.D. \rightarrow G_3$ 

Where $G_1 \rightarrow L_1$  $G_2 \rightarrow L_2$  $G_3 \rightarrow L_3$ 

And $L_1 = D + E_1$  $L_2 = D + E_2$  $L_3 = D + E_3$ 

And $E_1, E_2, E_3$ ... are all different.

4) We would expect that the 'decisions' made necessary by L.A.D. require only limited sorts of data.

We should be suspicious of a theory that for example makes 2 G's available such that the choice between them requires sentences with 5 layers of embedding, since such sentences presumably do not occur in the data base. Similarly, we would reject a theory that requires extensive use of 'negative evidence' in learning.

5) Finally, we would expect that L.A.D. contains an evaluation measure (EM) and that the class of possible G's is well-suited to evaluation. Unless the class of G's is so heavily restricted that L.A.D. provides only one G compatible with any D, an EM will be needed.

II. Some restrictions on the class of grammars

1) Transformations are structure-dependent

Thus, (2) is not available as a hypothesis for 'Question Formation'.
2) Move the first auxiliary verb to the front of the sentence. (4) rather than
(5) is the question corresponding to (3). Further, this is an overwhelmingly
strong intuition even though data leading to the correct result is surely
not uniformly available. In addition, (5) is not the kind of error children make.

3) The man who is here is tall.

4) Is the man who is here tall?

5) *Is the man who here is tall.

Further, in accord with (1), no language seems to have rules like (6) or (7).

6) Move the last word to the beginning of the sentence.

7) Interchange the 4th and 5th words of the sentence.

#1 has always been assumed. Below are several recent proposals for even heavier
restrictions. Lasnik and Kupin (1977) formalizes a theory incorporating these
restrictions. Note that such a theory provides a finite (and rather small) number
of tranformational components.

8) Every term a constituent. 9) 'Adjacency' cannot be freely stipulated. 10) Limited
number of terms (3). 11) Only one elementary operation per T. 12) Only 3 possible
elementaries = substitution, (Chomsky) adjunction, permutaion [The 3rd I assume to
be heavily constrained.]. 13) No quantificational or Boolean conditions. 14) No
'traffic rules' = ordering statements, optional v. obligatory rules, relative obligatori-
ness.

III. Restrictions limiting the types of data required.

Consider the following 'particle movement' data.
1) John called up Harry.
2) John called Harry up.
3) *John called up him.
4) John called him up.

One analysis (see Chomsky 1957) would have 2 rules of particle movement, an optional one
giving (2) when it applies; and an obligatory one giving (4) but never (3). What would
motivate this analysis for the child? If the child makes the wrong guess that both
rules are optional (and in fact identical), to discover that he is wrong, he would need
to know that (3) is ungrammatical. In fact, all of the properties ruled out by (14)
would require such negative evidence.

5) I gave my assignment to the teacher.

6) I gave the teacher my assignment.

7) I gave it to the teacher.

8) *I gave the teacher it.
(8) and (3) seem rather similar in their properties, and seem to tell us something
general about pronouns. If this can be factored out, a complicated set of
transformational statements will be greatly simplified. Similarly, virtually all
of the 'traffic rules' in Chomsky (1957) can be factored out and replaced with
something like (9).

9) Affixes cannot occur 'free' in surface structure.

Stated this way, the traffic rules don't even seem to have a syntactic role. Rather, 
they all conspire to guarantee a result that the morphological component will plausibly 
require independently.

IV. Two analyses of the English auxiliary

A. Chomsky (1957).

1. \[ \begin{align*}
\text{S.A.} & \left\{ \begin{array}{l}
\text{NP - C - VX} \\
\text{NP - CM - X} \\
\text{NP - C have - X} \\
\text{NP - C be - X}
\end{array} \right\} \quad \text{Precedes} \\
\text{S.C.} & X_1 - X_2 - X_3 \\
\text{Optional}
\end{align*} \]

2. \[ \begin{align*}
\text{S.A.} & \text{same as 1.} \\
\text{S.C.} & X_1 - X_2 - X_3 \quad \text{Precedes} \quad X_2 - X_1 - X_3 \\
& \text{Optional}
\end{align*} \]

3. \[ \begin{align*}
\text{S.A.} & \left\{ \begin{array}{l}
\text{S} \quad \emptyset \\
\text{Past} - \text{have} \\
\text{en} - \text{ing}
\end{array} \right\} \quad \text{Precedes} \quad Y \\
\text{S.C.} & X_1 - X_2 - X_3 - X_4 \\
X_1 - X_3 - X_2 \# - X_4 \\
& \text{Obligatory}
\end{align*} \]

4. \[ \begin{align*}
\text{S.A.} & \text{X-Y (where } X \neq v \text{ or } Y \neq Af) \quad \text{Precedes} \quad X_1 - \# X_2 \\
\text{S.C.} & X_1 - X_2 \\
& \text{Obligatory}
\end{align*} \]

5. \[ \begin{align*}
\text{S.A.} & \# - Af \\
\text{S.C.} & X_1 - X_2 \\
X_1 - \text{do} + X_2 \\
& \text{Obligatory} \\
& \text{[I assume that end variables should be added in 4 and 5. ]}
\end{align*} \]


All of the following are optional and unordered.

6. contraction \[ \begin{align*}
\left[ \begin{array}{l}
\text{+v} \\
\text{-n}
\end{array} \right] \\
\text{not}
\end{align*} \]

7. subject auxiliary inversion \[ \begin{align*}
\text{NP} \quad \text{Aux} \\
\text{Permute}
\end{align*} \]

2+1 (i.e. right adjoin 2 to 1)
8. affix hopping
   [+ Af] [+ n]
   2

9. v raising
   [+ tense] [+n] [+n]
   [-n] [-n]
   [+aux] [+aux]

(i.e.: replace 2 by 3)

10. do support
    [+v]
    [-n]
    [+aux]

Major P.S. rules:

11. Aux → [+Af] ( [+n] [+aux] )

12. V → Aspect V

13. V → [+v] [...]

14. Aspect → (not) (perfect) (progressive) (?passive)

15.

B. assumes Kiparsky's 'Elsewhere Condition' as a principle of rule application.
This replaces a complex set of traffic rules, namely, that v raising is obligatory,
and is ordered before do support.

16. If T₁ and T₂ are both applicable to P, and the set of potential inputs to T₁
    is properly included in the set of potential inputs to T₂, then T₂ (the
    more general rule) may not apply to P.

17. John (has not) left

18. (Hasn't) John left
19. Harry likes beans  
20. Does Harry like beans  
21. Harry does not like beans
22. Harry does like beans
23. * are careful [Imperative sentences are not finite]
24. * not be foolish [Imperative sentences have an affix that falls under the stranding prohibition]

Some relevant data
25. John is cautious  
26. John is not cautious  
27. *John does not be cautious
28. Be cautious  
29. *Be not cautious  
30. Do not be cautious
31. Assuming principle (15), a simplified v raising rule with no context term would both overgenerate (producing *29) and undergenerate (failing to allow for 30). Presumably, undergeneration is easily corrected.

V. We assume that U.G. is coherent, hence that the restrictions it embodies make sense.
1) Locality principles make sense: very long S's are not needed in D.
2) Heavy restrictions on traffic rules make sense: The need for negative evidence is greatly reduced.
3) E 8-13 above make sense: They inter-relate in a natural way, and result in a small and nicely scattered class of G's.

Some restrictions that don't make sense:
4) 'No T may have precisely eleven terms'. This wouldn't make a real dent in the class, and doesn't seem to follow from anything. (Unlike II 10 above, which follows rather naturally from II 11.).
5) 'T's that move material to the left in English cannot apply in both NP's and S's'. To the extent that this is true, we would surely want it to follow from something.
6) Restriction to some inner circle in the Chomsky hierarchy. Note that the right theory couldn't be a circle in this picture at all. For example, no human language could be finite, but every circle allows G's for such L's.

Note also that the goal of reducing the number of G's is completely orthogonal to this notion of 'power'. A theory like 'Z' might have a very small number of G's (and they might be nicely scattered). Lasnik and Kupin (1977) is such a theory.

Note also that a theory allowing, say, the entire class of C.F.G's would be virtually useless as a model of U.G. It would allow many G's with the wrong properties; would allow too many G's; would not obviously provide the needed scattering.
To the extent that restrictions are based on learnability considerations, we might expect them to be possibly overridden under special circumstances in which there is no learnability problem. The context term in IV 9 above was one example. The theory of markedness provides other cases. Thus, there can be both optional and obligatory rules as long as obligatory is the unmarked case, since any error will be in the direction of undergeneration.

References

