Learnability in Linguistic Theory: The Role of Restrictiveness and the Evaluation Metric

H. Lasnik
March 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.

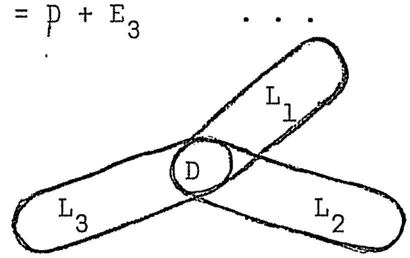
If any logically possible grammar were available to L.A.D., language acquisition could not take place in the way that it does. The data would drastically underdetermine the grammar, and a variety of children in the same linguistic environment could arrive at mutually inconsistent G's which all agree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

3) $D \longrightarrow L.A.D._1 \longrightarrow G_1$ $D \longrightarrow L.A.D._2 \longrightarrow G_2$ $D \longrightarrow L.A.D._3 \longrightarrow G_3$ ...

Where $G_1 \overset{\text{generates}}{\Longrightarrow} L_1$ $G_2 \Longrightarrow L_2$ $G_3 \Longrightarrow L_3$ . . .

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, permutation [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 obligatoriness.

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 <u>for the child</u>? 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 <u>ungrammatical</u>. 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 im 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. Proposals for the evaluation metric and their implications

1) S versus $\bar{S}$ for subjacency has been argued by Rizzi to be a parameter of core grammar.

The S (English) option rules out (2), (3).

2) *What [Do you know [who[saw]]]

3) *Who do you know what saw

Sentences analogous to these are apparently grammatical in Italian, but crossing 2 $\bar{S}$'s is impossible even in that language. Mona Anderson points out that if this analysis is correct, (4) must be a principle of U.G.

4) S is the unmarked case for subjacency

Similarly, the theory could countenance both optional and obligatory T's, but only if (5) is a principle of U.G..

5) Obligatory is the unmarked case for a T.

If non-universal filters exist, an impossible learnability problem would seem to arise. A rather stupid solution would be (6).

6) The unmarked case is for a G to have every filter.

Ken Wexler has a much more plausible approach:

7) The unmarked case is for derivations to observe 'uniqueness' = each D.S. gives rise to only one S.S..

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) II 8-13 above make sense: They inter-relate in a natural way, and result in a small and nicely scattered class of G's.

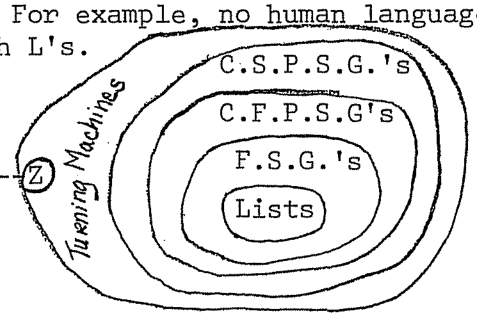
Some restrictions that don't make sense:

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 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 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 V 9 below turns out to be 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.

V. Two analyses of the English auxiliary

A. Chomsky (1957).

1. S.A. $\left\{\begin{array}{l} NP - C - VX \\ NP - CM - X \\ NP - C\ have - X \\ NP - C\ be - X \end{array}\right\}$ Precedes

$X_1 - X_2 - X_3 \longrightarrow$

S.C. $X_1 - X_2 + 'nt - X_3$

Optional

2. S.A. same as 1.
S.C. $X_1 - X_2 - X_3 \longrightarrow$ Precedes

$X_2 - X_1 - X_3$

Optional

3. S.A. $X - \left\{\begin{array}{l} S \\ \emptyset \\ Past \\ en \\ ing \end{array}\right\} - \left\{\begin{array}{l} M \\ V \\ have \\ be \end{array}\right\} - Y$ Precedes

S.C. $X_1 - X_2 - X_3 - X_4 \longrightarrow$

$X_1 - X_3 - X_2\# - X_4$

Obligatory

4. S.A. X-Y (where $X \neq v$ or $Y \neq Af$) Preced

S.C. $X_1 - X_2 \longrightarrow X_1 - \#X_2$

Obligatory

5. S.A. # - Af
   S.C. $X_1 - X_2 \rightarrow X_1$ - do + X_2

   Obligatory

[I assume that end variables should be added in 4 and 5.]

B. Lasnik (1981), in the framework of Lasnik and Kupin (1977).

All of the following are optional and unordered.

6. $\begin{bmatrix} +v \\ -n \end{bmatrix}$ neg

   $2 +_r 1$ (i.e. right adjoin 2 to 1)

7. NP Aux

   Permute

8. [+ Af] $\begin{bmatrix} + v \\ - n \end{bmatrix}$

   $1 +_r 2$

9. [+ tense] $\begin{bmatrix} +v \\ -n \\ +aux \end{bmatrix}$ $\begin{bmatrix} +v \\ -n \\ +aux \end{bmatrix}$ 2/1

   (i.e. replace 1 by 2)

10. $\begin{bmatrix} +v \\ -n \\ +aux \end{bmatrix}$ do/1

B. assumes Kiparsky's 'elsewhere condition' as a principle of rule application. This replaces a complex set of traffic rules.

11) If T_1 and T_2 are both applicable to P, and the potential inputs to T_1 are properly included in the potential inputs to T_2, then T_2 (the more general rule) may not apply to P.

Some relevant data

12) John is cautious 13) John is not cautious 14) *John does not be cautious

15) Be cautious 16) *Be not cautious 17) Do not be cautious