On the Evaluation Metric

Howard Lasnik
University of Maryland

LSLT already had acquisition as major concern. The first page of Chapter I says:

> We are antecedently interested in developing a theory that will shed some light on such facts as the following:
>
> 1 A speaker of a language has observed a certain limited set of utterances in his language. On the basis of this finite linguistic experience he can produce an indefinite number of new utterances which are immediately acceptable to other members of his speech community. He can also distinguish a certain set of "grammatical" utterances, among utterances that he has never heard and might never produce. He thus projects his past linguistic experience to include certain new strings while excluding others.
>
> *61*

Syntactic Structures discusses a possibly related task: How the linguist faced with linguistic data formulates a grammar. Chomsky there discusses three conceivable requirements on the relation between a theory of linguistic structure and particular grammars. The weakest, and the only one SS regards as even remotely feasible is that there be an evaluation procedure. That is, "given a corpus and given two proposed grammars G_1 and G_2, the theory must tell us which is the better grammar of the language from which the corpus is drawn." p. 51

[Interestingly, in SS there is no mention of acquisition, not even in Ch. 6 'On the Goals of Linguistic Theory']

One example of a possible evaluation procedure:

"Suppose, for example, that we were to evaluate grammars by measuring some such simple property as length. Then it would be correct to say that we have a practical evaluation procedure for grammars, since we could count the number of symbols they contain" p.53

This suggested evaluation procedure of the linguist in SS becomes in Aspects a suggested evaluation procedure for the language learner.

EM is PART of a linguistic theory. [See SS pp. 53-54 for this point wrt the task outlined in that book; and see Aspects p. 37 for EM as empirical wrt acquisition.] Hence it can't be used to choose between linguistic theories. Aspects p.38 bottom.

What the EM is is entirely an empirical question. One empirical proposal (or, really, speculation) was based on symbol counting.

Notational devices are part of the EM. ( ) doesn't follow from math, aesthetics, or anything else external to psychology. Rather, it makes an empirical claim about what's a significant generalization, what will be easier to learn

Early generative work assumed EM would be needed because it was assumed that the theory of possible grammars, though imposing many restrictions, didn't impose enough, so that there might be multiple allowable grammars compatible with PLD (but diverging somewhere beyond it). And conversely, even with an EM, we still need a restrictive theory of grammar. (Aspects, p.35 top)

There was opposition to EM, notably by generative semanticists. But the logic of their position was somewhat bizarre. If they had proposed a much more restrictive theory of grammars

than Chomsky's, then, conceivably, they could have made the case that the problematic situation would never arise, so EM wouldn't be needed. (As discussed in Aspects, p.36 bottom). BUT they explicitly argued for LESS restrictive theories than Chomsky did, so the EM would be even MORE necessary.

Was the EM computation difficult in these early theories? Was that why EM was abandoned? I don't think so. As SS pointed out, one likely candidate was 'practical'. What happened, I think, is that the format for grammars changed so much that symbol counting seemed irrelevant..

[[In the early GB era, where learnability. of course, was a/the central concern, there was much discussion of the need for a theory of possible parameters (replacing the earlier theory of possible transformations, as explicitly discussed in Aspects). I thought, and still think, that made no sense. You needed, e.g., a theory of possible Ts so the learner would have a fighting chance of picking out the right ones. But there is no analogy. The learner doesn't pick out the parameters. Rather, s/he picks out the VALUES of the parameters. So, at most, we need a theory of possible parameter values. As for the parameters, we need a theory of the ACTUAL ones

Parameters vs. rules. Does the former per se solve the problem of explanatory adequacy? Early GB era talk made it sound like yes. But that can't be right. Take an LSLT or Aspects style theory, call every possible T a parameter. Is learning now magically easier?]]

In the GB era, there was little talk about the EM, but much talk about markedness. Parameters were often argued to have an unmarked setting. Typically, there was just a diverse list. Assuming a very small set of (actual) parameters, that's good enough. The 'EM' now just ranks the unmarked settings more highly. As far as I know, with one possible exception, there were no big generalizations to be found here comparable to those of the early work (e.g., symbol counting). Janet Fodor will discuss this point tomorrow.

The one possible exception: Dell's 1981 LI Grammars (generating languages) in an inclusion relationship (better known as Berwick's Subset Principle).

"Whenever there are two competing grammars generating languages of which one is a proper subset of the other, the learning strategy of the child is to select the less inclusive one."

On Dell's interpretation, this is definitely a general metric, part of the EM. As Dell notes, this involves "evaluating two competing grammars by comparing, not these grammars themselves, but the extension of the languages that these grammars generate." This is quite different from Chomsky's EM. Among several consequences Dell deduces from GIR is that a rule is obligatory in the unmarked case. This consequences, along with all the others Dell discusses, involve the **grammars**. But Dell claims that these consequences are but theorems of the actual principle: GIR.

Berwick (1985) p.237 notes that the Subset Principle would be difficult to utilize as a learning procedure in the general case since for context-free languages and beyond determination of whether one language is a subset of another is undecidable. He offers two possible solutions to this problem. The first proposes limiting UG in such a way that at most degree 2 data are needed for successful acquisition (as proposed by Wexler and Culicover (1980)). Then the subset computation is decidable. Alternatively, the opposite of Dell's tack, it might be that the Subset Principle is actually just a descriptive generalization with the specific unmarked cases "fixed over time by an outside oracle, e.g., natural selection." This second possibility again brings to mind one of the topics Janet Fodor will discuss tomorrow: that learning implementations of current theories don't seem to provide "broadly principled metrics contemplated as possible components of EM in *Aspects*."