*Anjum Saleemi,National Chi Nan University and Teresa Satterfield, University of Michigan*

# Mind the Gap: Epistemology and Development of Natural Language

In this paper we suggest a way to bridge the inevitable gap between certain related epistemological conceptions of natural language and the empirical evidence from language development. One such conception, enshrined in the 'poverty of the stimulus' hypothesis (Chomsky 1965), argues for substantial innate predetermination of linguistic knowledge (virtually equating it with a Language Acquisition Device), and on that basis makes the claim that a minimal theory of language learning should be sufficient to explain the uncontroversial fact of language acquisition, a claim whose explicit formalizations have led to rather unexpected conclusions (Wexler and Culicover 1980). The second related conception, grounded in the theory of parameters (Chomsky 1981), assumes without argument the notion of innate structures, which is reconceptualized as Universal Grammar (UG), but incorporates a novel view of language variation that is extended to the hypothesis that the mechanisms of language variation and learning are one and the same, that children actually do set parameters, and consequently that language learning is indeed, as claimed by its predecessor, nothing but a trivial endeavour. Parametric theory fails firstly because it incorrectly assumes that language learning mechanisms must be impoverished just because those of language variation are (Clark 1992, Gibson and Wexler 1994; also the contributions to Bertolo 2001), and secondly because the view of language variation adopted in it is grossly simplified in the first place.

Instead of positing a minimal and uniform language learning theory that would require either nearly rigid pre-specification, or none at all, with the difference that it would not work in the latter case, we add another dimension to this ongoing debate by arguing that (contrary to Pullum and Scholz 1992, and some others) the notion of a richly deductive UG (Chomsky 1981) can be maintained independently of the familiar assumptions about the degeneracy of input. We also claim that under our modified UG hypothesis the language learning system need not to be undifferentiated and impoverished, as is commonly assumed, since language is a complex cognitive domain permitting a degree of variation that is by no means insignificant.

We outline a view of syntactic variation that is not parametric, except in a diffuse or 'distributed' sense. Essentially along the lines of the Minimalist theory (Chomsky 1995 and subsequent related work), we hypothesize that linguistic variation is a cumulative outcome of the association of the tenets of UG with classes of lexical items and, by extension, with the classes of the derivative linguistic constructions. Given such a conception of variation, a grammar need not, perhaps normally cannot, be a completely unified system. We show that adult grammars often contain rules which may yield syntactic systems that are partially 'split' or internally inconsistent (Saleemi 2002, Satterfield 1999). To cite only a few examples, V-raising in English applies only to Aux and not to lexical verbs, giving rise to two different but co-existing patterns; the Case and agreement system of Hindi-Urdu is split between ergative and nominative patterns depending on a combination of aspectual, lexical and structural factors (Saleemi 2002, 2003); and V2 in German is not applied across the board but excludes some verbs. (Needless to say that the 'subgrammars' involved must have sufficient structural and lexical overlap for them to hang together to some extent.)

In line with the view of variation outlined above, we hypothesize that the human language learning system comprises a collective of domain-specific and domain-general learning mechanisms (see Culicover 1999, Saleemi 2002, Satterfield and Pérez-Bazán, in preparation, Yang 1999, 2002, for related, but by no means identical views). Domain-specific learning (DSL) consists of two subtypes which we will label in rather general terms for now. The first, *category*

*learning* (CL), is designed to acquire the relatively arbitrary and less systematic knowledge of a target grammar and to assign categorial features (e.g., V, N, Infl) to lexical items from a highly structured matrix of universal categories. The second, termed *structural learning* (SL), handles the increasingly systematic aspects of the grammar, for example the placement of heads in maximal projections, V-raising, movement of V to I, WH-movement, scrambling, etc. As already mentioned, we also posit a statistically sensitive system of domain general learning (DGL). DGL mediates between the complex surface configurations of the primary input data and the domain-specific mechanisms. For example, it might feed partially pre-processed data to CL or operate with the SL on a pool of competing grammars to select the one most compatible with the target data (Yang 2002). Thus the various mechanisms work in tandem with each other, feeding information back and forth among them dynamically (Culicover 1999, Satterfield 1999). Further, DGL, in combination with some maturational factors (Borer and Wexler 1987, Wexler 1999), will help explain why child languages develop incrementally the way they do, and why they usually lack erratic developmental shifts. Furthermore, our analysis seeks to demonstrate that none of these mechanisms can function independently of the other, and that none is causally efficient (in the classical Aristotelian sense) on its own. For instance, SL cannot carry out lexical category assignment which must be treated by CL, and without the output of the lexical projections provided by CL, SL cannot make a decision about operations like V-raising.

With regard to DGL, what might be less obvious from the preceding discussion is that it has to presuppose both SL and CL. Logically, it should not require anything explicit prior to being implemented, but being an abstraction over specific processes (and arguably a 'late-comer' in phylogenetic terms), one would assume that there has to be something *prior* for the DGL to operate upon. This observation underscores a point often absent from the relevant discussions: any general mechanisms of cognition do not dispense with the specific ones; instead, they crucially presuppose them. When dealing with the myriad complexities of linguistic structure, Skinner's box cannot afford to be empty: it must at least contain a *random* assortment of various specific entities. In other words, induction without some prior history of deduction is a non-starter; however, once induction does become available, deduction simply cannot work effectively without it in any complex domain.

Assuming, as we claim, that DGL and DSL are indeed interdependent, it is not unreasonable to conclude that DGL (and indirectly DSL) may be sensitive to some variety of non-positive information; for example, Satterfield (1999) and Yang (1999, 2000) propose punish/penalty probabilities among competing grammars (see also Saleemi 1992). In fact, if the current proposal is indeed plausible, the distinction between positive and (indirect) negative evidence, the stock-in-trade of much traditional learnability reasoning, at best represents two extremes of the same spectrum of variously usable input.

This view of language acquisition also has profound implications for early bilingualism (simultaneous acquisition of two languages from birth), which we assert is merely a more complex and dynamic case of what is customarily and somewhat naively set apart as monolingual acquisition (see the remarks in Chomsky 2000).