

# MIND THE GAP: EPISTEMOLOGY AND DEVELOPMENT OF NATURAL LANGUAGE

Teresa Satterfield, *University of Michigan*

Anjum Saleemi, *National Chi Nan University*

## I. INTRODUCTION

This paper is part of a newly-initiated research program, as such it is an effort at resolving a longstanding tension between certain related epistemological conceptions of natural language and the real-time evidence from L1 language acquisition. We hope to bridge this gap by exploring a unique account of syntactic development that departs in critical ways from standard conceptions of child grammatical development. Specifically, in the most prominent generative-grammar based acquisition studies Universal Grammar (UG) is viewed as an overarching innate linguistic system constituting a unit of universal principles, some which contain parameters (or *p-parameters*, subsets of grammatical principles, such as pro-drop as described in Hermon (1990)) as choice points which can be fixed in one of a limited number of ways. A particular grammar is assumed to be immediately derived from UG by ‘fixing’ the parameters in a certain way on the basis of experience: Spanish, Russian, etc., are direct expressions of UG under particular, and distinct, sets of parametric values. No theory of language learning is explicitly postulated; instead, most final state grammatical structures are assumed to be directly attainable by UG principles, under the parametric options. Under this conception acquiring a language becomes merely a trivial endeavor: the initially novel assumption of language variation adopted in the Principles and Parameters approach (Chomsky 1981, etc.) now is subsumed in the hypothesis that the mechanisms of language variation and learning are identical—one and the same—and that children really do ‘set’ parameters, in the sense of selecting, among the options generated by the mind, those which match experience, and somehow discarding all others.

In this paper, we entertain the hypothesis that the parameters associated with UG principles, as characterized in the standard acquisition literature are not conceptually desirable. These assumptions remain dominant in the literature because, yes, parameter theory has provided an entirely new way to think about language acquisition; and yes, in the process it has yielded certain interesting ideas about language uniformity. However, as we will argue, the presupposition of such a type of parameter leads to a number of difficulties. Notably, when attempts are made to explicitly formalize the standard claims using mathematically-based models, failures arise because parameter theory incorrectly assumes that language learning mechanisms must be impoverished just because those of language variation appear to be (Clark 1992; Gibson and Wexler 1994; also contributions to Bertolo 2001). Secondly, because the standard view of language variation adopted is grossly oversimplified in the first place (e.g., [+/- value]), methodological and interpretive difficulties emerge in the most basic analyses (Culicover 2000). For instance, there exist an empirically significant number of adult grammars whose syntactic systems are partially ‘split’ or internally ambiguous with respect to the presumably mutually exclusive values of parameters represented by the ‘core’ phenomena (Satterfield 1999, 2003; Saleemi 2002). Take examples in movement, such as wh-questions in French, or scrambling in West Germanic; or in word order, as in the SVO/VSO alternation and asymmetrical agreement in Standard Arabic, where often neither variation is considered more ‘basic’. Standard Arabic manifests two different and co-existing word orders in addition to alternate inflectional patterns. SVO order elicits verb agreement in person, number and gender with its DP

subject, whereas VSO order yields verb agreement of person and gender only with the subject. It is important to note that these ‘mixed paradigm’ facts are not epiphenomenal, but rather are systematic variations that present several problems for standard parameter theory, and in turn for learnability theory, in their current conceptions.<sup>1</sup> Using the ‘mixed’ paradigm data as support, this paper addresses the issues of determining the relevant characteristics of syntactic variation and distinguishing universal principles of grammar from optional features in order to begin to provide a plausible account of the acquisition of L1 syntax that reflects real-time data. We begin with two working hypotheses, stated, below:

(1) DISTRIBUTED UG HYPOTHESIS: UG is made up of a multi-level, separable system of learning mechanisms. Within UG, the L1 learning system need not be an undifferentiated “blackbox” which is uniformly impoverished, as is commonly assumed. Language is a complex cognitive domain permitting a degree of variation that is by no means insignificant. Given our conception of variation below, a grammar is not, perhaps normally cannot, be a completely unified (in Jackendoff’s 2002 terms, a ‘monolithic’) system.

(2) VARIATION WITHOUT PARAMETERS HYPOTHESIS (VWP): Syntactic variation is not parametric. We contend that linguistic variation is a cumulative outcome, the interplay between the highly restricted tenets of UG with lexical items and, by extension, with the classes of derivative linguistic constructions.

In the remainder of this paper we present some in-depth arguments to support the Distributed UG and the Impoverished Parametric Variation Hypotheses. We will then further argue against parameter theory in its standard formulations as a mechanism of grammatical acquisition; that is, we will demonstrate that certain “contradictory” options are maintained even after the parameter has presumably been fixed for the alternate value, using cases across various languages as examples. We then turn to the alternative learning architecture itself, hypothesizing that within the language faculty (FL) the human language learning system comprises a collective of both domain-specific and domain-general (as opposed to domain-neutral) learning mechanisms (see Culicover 1999, Saleemi 2002, Yang 1999, and Satterfield and Pérez-Bazán (in preparation) for related but by no means identical views).<sup>2</sup> In the proposed account, the learner is endowed with two types of UG principles, both generic and ‘systematic’ but there are none of the

---

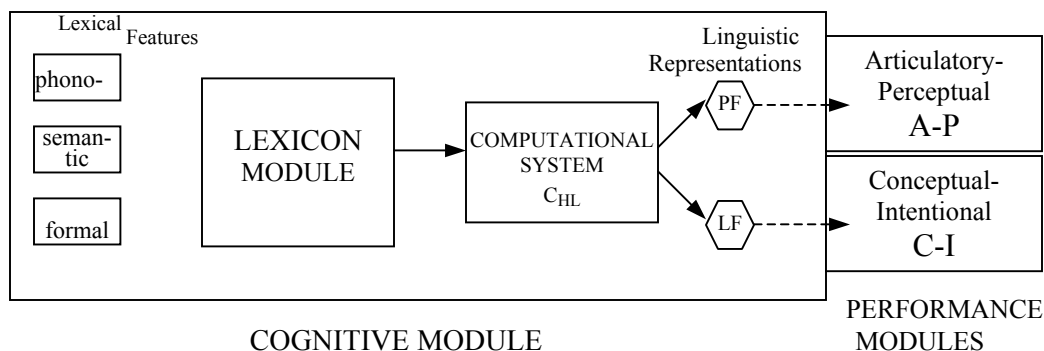
<sup>1</sup> And while we agree with the research strategy of limiting the data, somehow extracting from the wide-range diversity found across human languages to provide the linguist with an idealized decontextualized system within which hypotheses can be tested and manipulated, the object of inquiry cannot be simplified to the point of not representing the true structure of the grammar. To what extent is one justified in extrapolating from the analysis of one language to another: related languages typically contain subsets which are deceptively identical in some (many?) respects, and a closer look is usually likely to reveal significant differences in the underlying grammars. We think in some of the current linguistic theory the practice of generalizing from one language to another is indulged in rather too quickly.

<sup>2</sup>As will be seen, in order to develop an account of a domain-specific learning mechanism, we need to do two things. First, we need to come up with a principled way to individuate domains. Second, we need to say what makes a learning mechanism specific to a domain (or not).

standardly associated parameters. The task of the learner is to ultimately determine which lexical items are linked with the particular systematic principles, as mediated via the maturationally-sensitive learning mechanisms. What the learner acquires from experience is the lexicon---that is, the particular morphological and lexical structures of his language. As the learner accumulates particular lexical items and their associated categorial properties, the grammar will ‘evolve’ over time to reflect the increasing interaction of these structures with the various UG principles.<sup>3</sup> Thus, instead of positing a minimal and uniform language learning and variation theory that would require either no pre-specified structure or nearly rigid pre-specification, with the difference that the former case would be untenable, we explore another dimension by arguing that the notion of richly deductive UG (Chomsky 1981) made up of ‘adaptive’ components can be maintained independently of the familiar assumptions of degeneracy of input (poverty of the stimulus). We see syntactic options as emerging as a matter of degree into the grammars as the different principles crystallize into grammatical structures. In short, we are claiming that the need for the strong [+/-] conception of parameterization can be obviated by better articulating syntactic principles and by making use of the powerful learning mechanisms that can be argued to already reside internal to the cognitive domain of language. Insofar as UG is distributed, we claim that language learning proceeds according to distinct mechanisms and maturational timetables.

## II. BACKGROUND

Returning to our working hypotheses, (1) (DUG) and (2) (VWP) are meant to be intrinsically connected, but not redundant, if as we claim, language variation and language learning cannot be construed as a single unified property. We hypothesize that linguistic variation is a cumulative outcome, the interplay between the highly restricted tenets of UG with classes of lexical items and, by extension, with the classes of derivative linguistic constructions. The product of this interaction is the derivation, following along the lines of the Minimalist Program (Chomsky 1995 and subsequent related works), which at the Logical Form (LF) and Phonological Form (PF) interfaces of the language faculty provides specific instructions for the conceptual-intentional (C-I) and the articulatory-perceptual (A-P) systems, respectively.



<sup>3</sup> See Hermon (1990) for early thought-provoking arguments regarding a non-parametric analysis of binding facts under GB.

Figure 1: Human Language Faculty and Interface levels

As a point of departure, we assume the following characterizations in line with Chomsky (2002:85): part of the human biological endowment consists of a specialized “language organ,” the faculty of language (FL). The initial state of FL,  $S_0$ , is Universal Grammar, an expression of the genes that delimits the possible structure of human language, and displays the biologically essential universals of linguistic knowledge. These generic aspects of human language, which may include pronominal reference (in the form *c-command*), structure-dependence, Merge, fixed word order, Economy, etc., are universal not only because they are pre-specified by our genetically-endowed language faculty, a component of the biological make-up of the species, but, we would argue that they are the most abstract parts of language, the ones least amenable to semantic and pragmatic support.<sup>4</sup> On Chomsky’s view, UG operates via the triggering of a set of parameters, such that it can be regarded as a device that maps experience into state  $S_L$  attained: “a language acquisition device” (LAD) (Chomsky 2002:86). One of the central theses of our current research program, however, is that language learning is based on a collective of specialized mechanisms or “guided learning” mechanisms (cf. Jenkins: 2000:164) that operate on input that has been mapped to the FL initial state. These learning mechanisms should be thought of as specialized language organs within the brain, each respectively performing its specific kind of computation. A plausible conceptualization of these mechanisms will be taken up in detail shortly. For the moment, our hypothesis is that there could be several initial knowledge states, some of which possess domain-specific mechanisms which we term category learning devices, designed to acquire relatively arbitrary and less systematic knowledge of the target grammar, such as lexical information. States that seem to stabilize over several progressive stages can then emerge, with additional domain-specific learning mechanisms, which we call structural learning devices, that handle increasingly more systematic principles of the grammar such as *wh-movement* or scrambling and there are also domain-general learning mechanisms that interact with the domain-specific components. Each state in turn is shaped by the myriad effects of learning, or more appropriately, of growth, which we view as the combination of experience and internally-determined processes such as maturation, learning mechanisms, and computations of the FL.

The foundation for the DUG and NPV hypotheses naturally begins with the ‘logical problem of language acquisition.’ Two important dimensions of this conception must be addressed, both related to language *learnability*. One is the general problem of how to delimit the hypothesis space so as to ensure that the learner can in principle only converge to a possible natural language. We can now begin to consider this question from the perspective of a very recent thesis formulated within the Minimalist Program (Chomsky 1995, 2000a, 2002), known as the “No Dead End Assumption.” The “No Dead End” idea, in its most radical interpretation, makes the claim that every possible human language fits minimalist standards, such that the condition of infinite legibility must be satisfied, and in an optimal manner, at the interfaces of linguistic representation, PF and LF. And if it is assumed that the FL consists only of states, which are all (real?) languages—whether it is the initial state, intermediate states, or the stable states that learners ultimately arrive at—any state that the FL can attain is a

---

<sup>4</sup>In crosslinguistic terms, UG is also seen as a “theory of linguistic invariance (cf. Belletti and Rizzi 2002:9)”, insofar as it expresses the universal properties of natural language.

language, yielding an infinite number of interpretable expressions. That is, if the strong ‘No Dead End Condition’ obtains, then the task of determining the possible languages is mitigated, as minimalist assumptions hold that you cannot have linguistic variation in such a way that you get a system that will fail to have an infinite satisfaction of the interface conditions, including in the initial state. The question of convergence to a possible human language is then redirected to the question of how exactly the learner converges to the ambient variations.

In addition, there is a more specific dimension of the learnability problem, which is how to delimit the hypothesis space such that it can be efficiently searched by means of a memory-less enumeration function with access to solely positive input. Models within the standard Principles and Parameters approach (Chomsky 1981; Chomsky and Lasnik 1993) support the notion of a deductive learner who, driven by input data from the environment, searches the finite parameter space as defined by UG to select the appropriate binary-valued parametric option (and associated cluster properties) as suitable for the target grammar. At any point in the process, a single grammar is made available to the L1 learner. Recall that for Chomsky, parameters are set merely based on a preponderance of evidence, with no implicit learning mechanisms in place. Saleemi (2002) has pointed out that this conception is fraught with learnability problems. For example, even by constraining choices to only yes or no options, a linguistically simple 10-parameter space still yields a daunting number of grammar combinations to be considered by the learner ( $2^{10} = 1024$  possible options). Moreover, in the worst case, if certain parameters are linked, setting a sequence of parameters in the face of independent cluster properties would soon become an insurmountable task. Several well-known accounts maintain that a triggering model can allay these difficulties; however Gibson and Wexler’s (1994) analysis appears to inherit all of the conceptual problems of the account proposed by Chomsky, in that the GW model similarly operates on the learner moving through a [+finite] hypothesis space to select a grammar based on highly idealized input, based on a class of Comprehensive Triggering Algorithms. Moreover, to the degree that this model reflects real-time acquisition, questions arise regarding GW’s rationale for implementing certain default parameter values for the potential success of V2 acquisition in a three-parameter learning space.

Even granted the viability of standard triggering models to provide a reasonable solution to the learnability problem, a related difficulty still concerns ambiguous input, in the form of hypothesis sentences that correspond to several possible grammars. The logic follows that if the binary-valued parameters are fixed via very limited linguistic evidence (Chomsky 2000), then an undesirable outcome is that a triggering model is able to drastically alter the learner’s grammar based on one input sentence. Thus, as a result of ambiguous incoming data, the all-or-nothing constraint imposed on triggering could easily lead the learner to converge to local maxima, and inevitably to a grammar that did not correspond to the target. If conditions of learnability could actually be met under trigger-parameter theory, another important question that comes up is the possibility of inducing unambiguous triggers. Fodor (1998) not only demonstrates clear-cut evidence of unambiguity for each of the grammars in the three-parameter domain, as proposed by GW (1994), she then formulates an alternate triggering model that implements only unambiguous input. On a more trivial note, parameters are presumably fixed based solely on their correct “match” with the ambient data. At a very fundamental level then, parameter-setting, especially when viewed as a fine-tuning of the invariant grammar endowed in UG, is merely an exercise in associative learning. Since the nature of the information that the learner extracts from the input data and how it is represented is as yet unknown to us, how can we have a significant discussion of “triggers” at this point? In light of the important projects recently initiated (Fodor and Sakas, (pc)), hopefully we can begin to address notions of triggers, by

exploring probabilistic learning mechanisms with natural language input that provide both an empirical basis for assessing how innate constraints interact with information derived from the environment, and a source of hypotheses for experimental tests.

Nevertheless, the obstacle remaining for a one-grammar trigger theory of parameter-setting is how it might include the often attested presence of mixed-paradigm input. As mentioned, mixing occurs with non-marginal structures not associated with the particular target grammar in the literature, even though they actually appear frequently within native speech patterns. Briefly, there are the well-known cases used to substantiate triggering models such as V2 in German; however, the verb-second rule applies not across the board, but instead excludes certain verbs. In English, V-raising applies only to Aux and not to lexical verbs, giving rise to two distinct but co-existing patterns; the Case and agreement system of Hindi-Urdu is split between ergative-absolutive and nominative-accusative patterns depending on a combination of aspectual, lexical, and structural factors (Saleemi 2002, 2003). The above arguments can be repeated, for many different languages and many different structures.

To begin to address the conspicuous failure of many parametric accounts to accommodate mixed-paradigms found naturally across languages, it is necessary to subject the assumed properties of UG to greater scrutiny. As noted by Jackendoff (2002), perhaps the main snag lies in the conception of UG---both by its proponents and its detractors---in terms of a “grammar box cut off from the rest of the mind. This is clearly an oversimplification (cf. Jackendoff 2002: 79).” We therefore concur, with the spirit of Tomasello’s (1995) pronouncement that we should be judicious about how much linguistic structure is ascribed exclusively to the initial state UG. As Chomsky (1965: 207) concedes: “In fact, it is an important problem for psychology to determine to what extent other aspects of cognition share properties of language acquisition and language use, and to attempt, in this way, to develop a richer and more comprehensive theory of mind.” This said, we are by no means advocating general mechanisms alone to provide the kind of coverage to explain the unique architecture of language and the repertoire of rule types that governs it. In our view, any general mechanisms of cognition do not dispense with the specific ones; instead, they crucially presuppose them. When dealing with the countless 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, if UG is to be the biologically-endowed basis from which L1 development commences, it had better be minimally equipped to supply the learner with a specific knowledge of fixed word order, grammatical functions, and Case systems, etc., just in case the child encounters these properties in his primary linguistic data.

If instead of viewing UG as a single, unified “grammar box,” we follow Jackendoff’s (2002) lead to envisage it instead as a “toolbox,” then beyond the most universal “no-frills” minimum (perhaps this being syntax, the equivalent of Merged constituents), the learner must utilize many resources of his “toolbox.” In this light, languages cannot be constrained to simply activating a ‘yes or no’ value of a parameter to obtain the maximal advantages of the grammar. They pick which apparatuses they use, more specialized or less-specialized tools, and to what degree in a flexible and adaptive manner, but only in case UG is not conceived of as an indivisible black box. Keeping this postulation in mind, consider the critical period in L1 acquisition as further support for the DUG Hypothesis. It has been successfully argued that not all aspects of language display critical period effects. Specifically, the acquisition of lexical items, the concatenation of words, and the nuts and bolts of semantically-based principles of word order seem immune to age-related “atrophy” of language acquisition (Hudson and Newport 1999). However, the capacity to acquire other parts of language in late language learning, such as

the inflectional system or real intricacies in phrase structure, appear to be largely diminished. As we claim, these common L1 versus L2 conditions become at once isolatable by viewing UG as a multi-layered construction with distinct learning mechanisms.

### III. ALTERNATIVE PROPOSAL

We now put forward an alternative account as a preliminary attempt to avoid the bottlenecks of parameter theory, and to provide a more conceptually attractive story for the acquisition data to be discussed. The current analysis is drawn up with an eye to the non-parametric adaptive learning theory hypothesized in Culicover (1995). In our approach, the FL consists minimally of a Domain-general learning (DGL) device that serves as the interface between other brain systems that interface with FL and Domain-specific learning mechanisms (DSL) internal to the FL. Since we claim that maturational constraints also come to bear on the FL, it is reasonable to assume that no learning mechanism is fully operative in the initial state, nor are all mechanisms necessarily functioning in all knowledge states.<sup>5</sup> We will also assume without further discussion that a mechanism is domain specific if it is used only to learn about the things within its domain. Thus, if a learning mechanism M were used just for learning about a grammatical property, it would be a domain-specific language learning mechanism. A learning mechanism is more domain neutral, the more things outside its domain it is used to learn about. Thus, if M were used both for learning about statistical regularities and language when its domain was just language, it would, to that extent, be domain general.<sup>6</sup>

---

<sup>5</sup> Please note previous discussion (page 4) regarding Chomsky's "No Dead Ends Assumption."

<sup>6</sup> Clearly, our views about what the domain of a learning process is must be revisable in the light of what is discovered about how that process actually works and interacts with other processes.

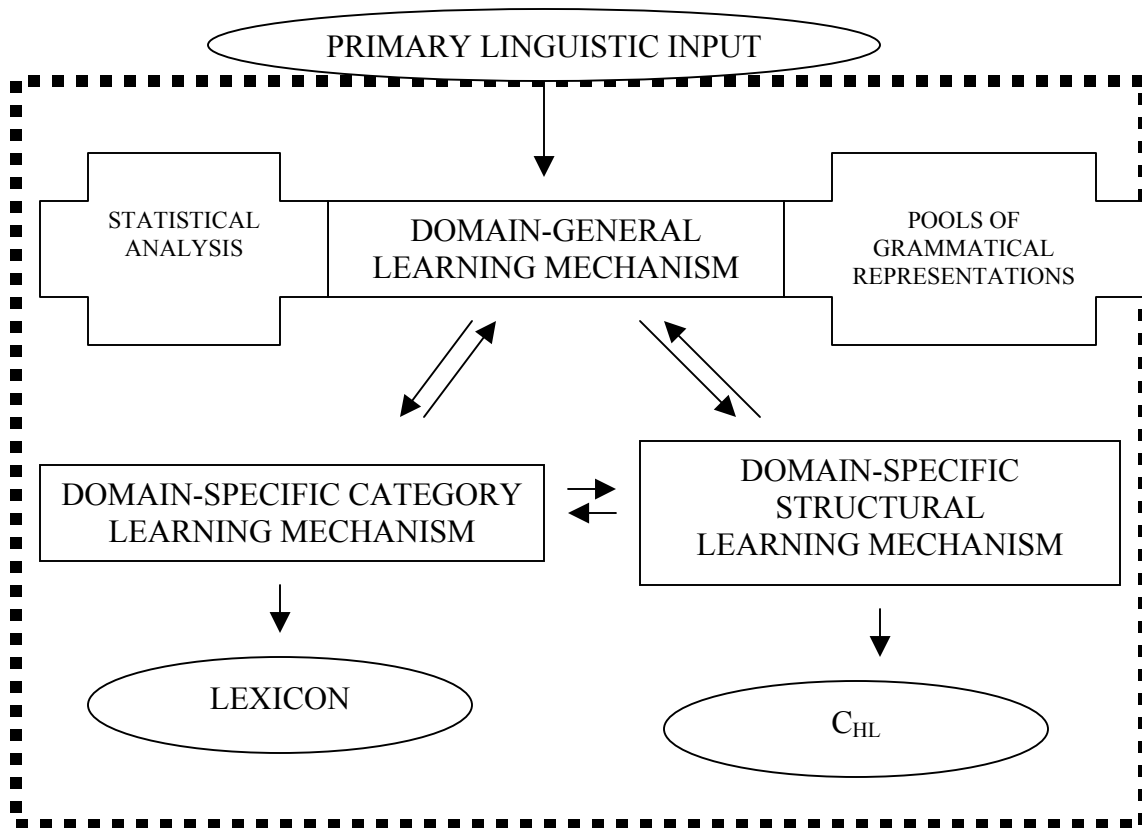


Figure 2. Proposed Learning System

In the initial state,  $S_0$ , an input sentence is analyzed by the DGL, which given our characterization above, is a relatively general learning mechanism within the domain of language. It serves several purposes, as it works as both a general parser and as a data organizer that is statistically sensitive. Its function as a type of universal parser permits the DGL to process the surface configurations of the primary linguistic data on the basis of the innate UG principles expressed. In other words, the DGL is a generic ‘tool’ in the FL toolbox. The DGL mechanism initially tracks arbitrary lexical elements as it deduces the basic UG features from the incoming data into as its output produces general structural descriptions (GSDs). The GSDs will be converted, over time, to the appropriate grammatical representations that are maintained by the learner. Why do we need a general statistical component? As recent computational research on natural language corpora reveals, relatively simple statistical learning mechanisms could make an important contribution to certain aspects of language acquisition (for example, statistical methods can provide valuable cues to the acquisition of inflectional morphology, syntactic classes, and aspects of word meaning). In each case, these cues are partial, and must be integrated with additional

information, whether from other environmental cues or innate knowledge, to provide a complete solution to the acquisition problem.<sup>7</sup>

As we claim, L1 learners have access to the principles of UG from the onset of language acquisition; however, they cannot completely process complex input strings or construct full grammatical representations because they have insufficient knowledge of the lexical and morphological properties of the language, as illustrated in the grammars that can be attributed to child language learners.<sup>8</sup> Given maturation and experience, the DGL becomes increasingly capable of fully parsing the input from the specific linguistic environment. Grammatical representations are incrementally constructed as the DGL mediates between the complex surface configurations of the primary input data and additional learning mechanisms that are domain-specific. After being encoded and organized by the DGL, the partially pre-processed input is mapped to a domain-specific learning device within each grammatical module. We consider learning mechanisms that perform one type of specialized operation on the data to be domain-specific. In the case of the acquisition of syntax, we posit a specialized learning mechanism that we will term, the Category Learning Mechanism (CLM). The CLM, as a subcomponent of the FL, is the most specialized “tool” as it resides over all things “non-systematic,” which primarily translates to material in the lexicon. The CLM is designed to analyze lexical items, to assign categorial features from a highly constrained matrix of universal categories, to compile an inventory of idiosyncrasies, idioms and, basically to organize and store the relatively arbitrary and less systematic knowledge of the target language. Again, as a function of maturation, the CLM will become more efficient over time. We posit that the CLM comes “online” and interacts with the DGL at the earliest point of language learning. Why should this be the case? Take word classes, for example: even assuming that the child innately possesses a universal grammar and syntactic categories, identifying the category of particular words must primarily be a matter of *learning*. Universal grammatical features can only be mapped on to the specific surface appearance of a particular natural language once the identification of words with syntactic categories has been made. Although once some identifications have been made, it may be possible to use prior grammatical knowledge to facilitate further identifications, the contribution of innate knowledge to initial linguistic categories must be relatively slight.<sup>9</sup>

---

<sup>7</sup> The success of these methods with real natural language corpora demonstrates their feasibility as part of the language acquisition mechanism, where much previous research has been limited to highly idealized artificial input.

<sup>8</sup> In recent formulations (Crain and Thornton 1998, Wexler 1998, Hyams 1996, Poeppel and Wexler 1993, etc.) the maturational view of grammatical development, with the possibility that functional categories mature later and upon doing so, make new grammatical options available has been widely challenged. However, studies with production data covering the earliest stages of acquisition exhibit broad ranges in terms of results, so it seems reasonable to not set aside the maturational hypothesis yet.

<sup>9</sup> One possible approach which has a long history, even though often dismissed on a priori grounds within the language learning literature, is the “distributional test.” It is based on

Along with the CLM, there is another DSL mechanism termed the Structural Learning Mechanism (SLM). SLM represents the moderately “specialized” tools in the FL toolbox, in terms of the systematic UG principles that are implemented in the structures, and translate mainly to the computational system in the FL. SLM mediates between the information in the CLM and the general structural descriptions mapped from the DGL. It may be the case that a sufficient threshold of lexical items must be accumulated in the CLM and thus, then when the GSDs are mapped to SLM, they can be re-analyzed for very specific syntactic operations such as movement operations in terms of wh-move, or scrambling, or V-to-I movement, etc. Once a GSD has been analyzed by the DSL mechanisms, it is again mapped to the DGL where it is maintained now as structural description, (SD) in the child’s grammar. Various data sequences of particular structures found in the linguistic environment can be stored by the DGL and organized into “like” pools of grammar which increasingly correspond to the target L1. As these pools of SDs are motivated by the classes of SDs characteristic of the target both in terms of systematic and “arbitrary” structures, over time those SDs of high frequency are stored as grammatical representations in the mind/brain of the L1 speaker. Given this formulation, the learner generalizes many aspects of the developing grammar in a gradual manner, moving from generic DGL to systematic SLM as linguistic evidence presented to him supports this step.

It is important to recognize that while each FL learning mechanism is specialized, none of them can function independently of the other, and none is causally efficient (in the classical Aristotelian sense) on its own. For instance, SLM cannot carry out lexical category assignment which can only be treated by CLM; and without the output of the lexical projections provided by CLM, SLM cannot make a decision about how to analyze V-raising in this particular knowledge state. This interdependence of the UG learning mechanisms in the absence of parameters seems to be the right strategy. Many problems in language acquisition are difficult because no single feature of the input correlates with the relevant aspects of language structure. Although it is a natural starting point for computational and empirical research to study input cues in isolation, it may be that the problem of acquisition is easier when multiple information sources are taken into account. Figure 3 below shows abstractly how three constraints A, B and C, represented by regions of the hypothesis space, are insufficient to identify the correct hypothesis when considered in isolation. It is only by combining these input cues that the hypothesis space can be substantially narrowed down. Thus, as the quantity of cues that learner considers increases, the difficulty of the learning problem

---

the observation that if all occurrences of word A can be replaced by word B, without loss of syntactic well-formedness, then they share the same syntactic category. For example, *dog* can be substituted freely for **boy**, in phrases such as *the boy sat on the mat, nine out of ten boys prefer ...*, indicating that these items have the same category. The distributional test is not a foolproof method of grouping words by their syntactic category, because distribution is a function of many factors other than syntactic category (e.g., word meaning). Thus, for example, *boy* and *thunder* might appear in very different contexts in some corpora, although they have the same word class. Nevertheless, it may be possible to exploit the general principle underlying the distributional test to obtain useful information about word classes. One common method records the contexts in which the words to be classified appear in a corpus of language, and then groups together words with similar distributions of contexts.

may decrease. This suggests that the cognitive system may aim to exploit as many sources as possible in FL:

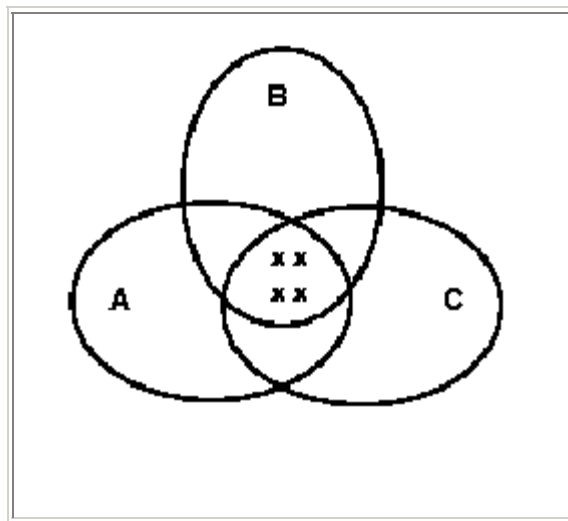


Figure 3. A conceptual illustration of 3 hypothesis spaces given the information provided by input A,B, and C (x' s correspond to hypotheses consistent with all 3 inputs)

Moreover, it is possible that input cues only be useful when considered together. For example, in the sequences in Figure 4 below, each cue X and Y seems completely random with respect to the target Z; but when considered together X and Y determine Z perfectly (specifically, Z has value 1 exactly when just one of X and Y have the value 1). Considering input information in isolation implicitly assumes that there is a simple additive relation between cues:

<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px 2px 10px;">X: 1 0 0 1 1 1 0 1 0 1 1 0 0 0 1 1 0 1 0 1 0</td> </tr> <tr> <td style="padding: 2px 10px 2px 10px;">1 Y: 0 1 1 0 1 0 0 1 1 0 1 1 0 0 0 0 1 1 1 0 0 1</td> </tr> <tr> <td style="padding: 2px 10px 2px 10px;">0 Z: 1 1 1 1 0 1 0 0 1 1 0 1 0 0 1 1 1 0 1 1 0</td> </tr> </table>	X: 1 0 0 1 1 1 0 1 0 1 1 0 0 0 1 1 0 1 0 1 0	1 Y: 0 1 1 0 1 0 0 1 1 0 1 1 0 0 0 0 1 1 1 0 0 1	0 Z: 1 1 1 1 0 1 0 0 1 1 0 1 0 0 1 1 1 0 1 1 0
X: 1 0 0 1 1 1 0 1 0 1 1 0 0 0 1 1 0 1 0 1 0			
1 Y: 0 1 1 0 1 0 0 1 1 0 1 1 0 0 0 0 1 1 1 0 0 1			
0 Z: 1 1 1 1 0 1 0 0 1 1 0 1 0 0 1 1 1 0 1 1 0			
Figure 4. A sequence of cues.			

#### IV. INFORMAL ANALYSIS

Let us walk through a well-known example using some Null Subject facts. There is actually a striking difference between the types of languages that

are classified as so-called Null Subject Languages, since they are observed to license null subjects, according to the conditions of the standard NSP analyses. Languages such as Standard Italian, Castilian Spanish, and European Portuguese (EP) are held to represent the non-mixed paradigm. Neapolitan (Italian), Caribbean Spanish, and Brazilian Portuguese (BP) represent the mixed paradigm which can allow the same surface distribution and often same semantic functions to co-exist between overt and null subjects.

#### Inventory of Pronominals

- 1) ENGLISH: S/he ate pizza:  
     EUROPEAN PORTUGUESE (EP)                      BRAZILIAN PORTUGUESE (BP)  
     Pro comeu pizza. = [+referential pro]                      \*Pro comeu pizza = Topic delete
  
- 2) EP AND BP: Expletives and arbitrary subjects, same pattern for EP and BP
  - a. Pro<sub>arb</sub> Esquiam muito bem na Suíça = They<sub>impersonal</sub> ski very well in Switzerland.  
     ARBITRARY
  - b. Ontem \*ele/\*aquele/pro fez muito frio. = It was very cold yesterday.  
     QUASI ARGUMENT
  - c. \*Ele/\*Aquele/ Pro é certo que (ele/ela/pro) fala bem = It is true that s/he speaks well  
     EXPLETIVE
  
- 3) OPC FACTS (Montalbetti 1984: overt pronouns cannot link to formal variables)
  - a. EP: Todo aluno<sub>i</sub> achou que pro<sub>i</sub>/ele<sub>\*i/j</sub> ia a passar de ano.
  - b. BP: Todo aluno<sub>i</sub> achou que pro<sub>i</sub>/ele<sub>i?</sub> ia a passar de ano. (Negrão 1997)  
     “Every student thought that he was going to pass.”
  - c. EP: O João<sub>i</sub> disse que \*ele<sub>i</sub>/ pro<sub>i</sub> vai trazer uma garafa de vinho.
  - d. BP: O João<sub>i</sub> disse que ele<sub>i</sub>/ pro<sub>i</sub> vai trazer uma garafa de vinho.  
     “João said that he is going to bring a bottle of wine.”
  - e. EP: Alguns alunos<sub>i</sub> disseram que pro<sub>i</sub>/\*eles<sub>i</sub> acham que eles<sub>i</sub> são inteligentes.  
     “Some students said that they think that they are intelligent.”
  - f. BP: \*Quem<sub>i</sub> acha que pro<sub>i</sub> disse que ele<sub>i</sub> é inteligente? (Negrão 1997)  
     “Who thinks that (pro) said that he is intelligent?”

## 4) GENERALIZATIONS

*EP OVERT SUBJECT PRONOUNS*

Not inherently contrastive

Not robust in topic/non-topic

Interpretations

Licensed in OPC when linked to pro

*BP OVERT SUBJECT PRONOUNS*

Not inherently contrastive

Robust in topic/non-topic

interpretations

Not licensed in OPC when linked to pro

*EP NULL SUBJECT PRONOUNS*

Function robustly in the grammar

Pro in non-contrastive contexts not bound by variables

*BP NULL SUBJECT PRONOUNS*

Function robustly in the grammar, BUT in specialized distributions (Negrão &amp; Müller 1996)

Pro, even in non-contrastive contexts bound by variables (obtains bound variable interpretation)

To account for this variation of overt and null subject alternation in the languages in question, analyses employing standard P&P theory have proposed that: 1) BP is not a null subject language, with the implication being that BP is a non-NSL, presumably similar to French or English (XXXX); or 2) BP is a null subject language, but sets the parameter differently, along the lines of speakers of NSLs who have topic-drop settings such as in Chinese (Modesto 1999). Neither type of analysis corresponds to the actual nature of BP as a mixed paradigm of co-existing, sometimes interchangeable, non-null and null-subject elements. In either case, the predictions falling out from standard parameter-based accounts in terms of L1 acquisition are that the learner will not be able to set the parameter correctly, since the amount of ambiguous data will be excessive. That is, given the mixed nature of the input, the implications are that the child will not have the appropriate cues from the primary linguistic data to determine whether the target grammar is truly a [+null subject] language or not, and will not arrive at any setting.<sup>10</sup> Nevertheless, it is apparent that children in these environments do acquire this aspect of the L1 with relative ease. As previously mentioned, it may be possible to formulate certain micro-parameters; however to do so, one would need to adopt a number of these option points in order to accommodate the free variation of subject elements, as well as the restricted distribution that occurs in certain contexts. Certainly this tactic contradicts the spirit of Principles and Parameters. In sum, given any of these scenarios, the parameterized approach falls short on the grounds of descriptive adequacy, learnability, and elegance of exposition.

Returning to our alternative proposal, the data reviewed above does not contradict our proposed formulation of a non-parametric account of variation. While the full syntactic analysis will not be discussed here (see Satterfield

---

<sup>10</sup> We will not consider the possibility of the child eventually determining the correct null subject setting for his language based on the Subset Principle, since the null subject versus non-null subject contrasts cannot be expressed in a subset relation. It is not the case, for example, that all properties of English subject pronouns are properly contained within all the possible properties of Spanish subject pronouns.

2003)<sup>11</sup>, the approach we advance views the occurrence of [+referential] null subject arguments as the confluence of independently motivated generic and systematic UG principles such as Economy (both of fewest derivational steps and projection), Agree operation (Lasnik 2000) as a way of licensing and identifying null elements, and Move, in addition to the learner's knowledge of morphology. In brief, based on this analysis speakers do not categorically "Avoid Pronoun" whenever it is possible, as has been previously claimed. Indeed, under certain syntactic conditions it is less "costly" for speakers to generate overt subjects rather than null ones. There is no reference made to a parametric option, but the outcome allows us to explain the strong presence of overt subjects in languages such as BP, without needing to discount the robust occurrence of pro in particular constructions, in order to make a clean "fit" into the [--NSL] class for BP. The task of the learner in our proposed framework is to acquire the lexicon (in this case, pronominals and morphological affixes) as presented in the input and then to identify and associate these items with the various constraining UG principles, as mediated by the specialized learning mechanisms. Unlike the parameter-based models, a learner in this analysis would not face conflicting evidence, as the robust constructions of co-existing subject elements can all be stored as SDs, since they reflect in increasing degree the characteristics of the target language. From this pool of SDs, the most complete constructions can be selected as grammatical representations for the learner's L1 grammar.

### III. FURTHER IMPLICATIONS AND CONCLUSION

In this paper we have laid the foundations for a research program that diverges in key ways from the present trajectory of Principles and Parameters approach, and as we claim, points to solutions for learnability matters that have eluded P&P analyses in the past. We have argued that the conception of language acquisition as motivated by parameter-setting fails for diverse reasons. We have also attempted to make a case for the notion of a distributed system of specialized learning mechanisms that operate within the FL, dispensing with the need for rigid parameters of variation. As with any promising inquiry, this program opens the door for many more important questions which could ultimately offer greater explanations for the common variations found in (child) language data. This view of language acquisition has profound implications for future research, now that an abstract path has been put into place. It is necessary to elaborate our hypotheses by formalizing the learning mechanisms as algorithms to more thoroughly evaluate their viability. For example, to the extent that the learner analyzes and accumulates data over time, it begs the questions of the resources of memory and storage that would be required for the very young child to house various candidates for grammatical representations (SDs). Secondly, empirical evidence is needed, perhaps in the form of psycholinguistic experiments in order to assess if our hypotheses align with natural language patterns. For instance, information is a potentially valuable cue for many aspects of language acquisition. Does the child use these sources of information? As with all theories of language acquisition, empirical evidence regarding distributional methods is difficult to obtain and interpret. It is encouraging that recent experimental evidence in both children and adults shows that the cognitive system is sensitive to features of the input (e.g., co-occurrence statistics) which underlie the DGL mechanisms described here (Saffran, Aslin, and Newport 1996). It

---

<sup>11</sup> Also see Roeper and Rohrbacher (1995) for a non-parametric approach to early Null Subjects.

seems a reasonable working assumption that, given the immense difficulty of the language acquisition problem, the cognitive system is likely to exploit such simple and useful sources of information, and worth exploring. Another possible project would be to determine whether certain surface configurations of syntactic development can appear to resemble very early parameter-setting (VEPS in the sense of Wexler 1998), due to intermittent, albeit observable, adult-like features which appear at certain points in the child's speech production. However it may be the case that the child's grammar may still not be in a final (steady) state in terms of the interplay of all the deeper derivative information necessary for the truly stable emergence of a given linguistic property. This is an interesting question to pursue. Perhaps the more intriguing aspect of this program is to determine how it would fare in the case of truly multilingual speakers, who must organize two or more distinct grammars from disparate inputs. Lastly, a problem with this account is that while domain specificity is well defined, domain generality comes in degrees. That is, we can say that a mechanism is domain specific, period, so long as it is only used for learning within its domain, that is, so long as it only does what it evolved to do. But we can only say that it is *more or less* domain general: a mechanism is the more domain general the more its operations generalize to other learning tasks. Now, this might not be a problem in principle - lots of perfectly good distinctions, after all, involve matters of degree. A computational simulation must be designed, since it will oblige us to make explicit such inner-workings of the model (and of the learner) that we are assuming.

#### *REFERENCES*

- Belletti, A. & Rizzi, L., (eds.). (2002). *Noam Chomsky: On Nature and Language*. Cambridge: Cambridge University Press.
- Bertolo, S. (ed.). (2001). *Language Acquisition and Learnability*. Cambridge: Cambridge University Press.
- Bosch, L. & Sebastian-Galles, N. (1997). Native language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition* 65, 33-69.
- Chomsky, N. (2000a). Minimalist inquiries: the framework. In R. Martin, D. Michaels, & J. Uriagereka, (eds.), *Step by step-Essays in minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: MIT Press.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, N. & Lasnik, H. (1993). The theory of principles and parameter. In J. Jacobs, (ed.), *Syntax: An International Handbook of Contemporary Research*. Berlin: Walter de Gruyter.
- Chomsky, N. (1988). *Language and problems of knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.
- Chomsky, N. (1986). *Knowledge of language: its nature, origin, and use*. New York: Praeger.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Clark, R. (1992). Selection of syntactic knowledge. *Language Acquisition* 2(2):83-149.
- Crain, S. & Wexler, K. (1999). A modular approach to methodology. In W. Ritchie & T. Bhatia (eds.), *Handbook of Child Language Acquisition*, pp. 387-425. London: Academic Press.
- Crain, S. & Thornton, R. (1998). *Investigations in Universal Grammar*. Cambridge, MA: MIT Press.
- Crocker, M. (1996). *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*. Dordrecht: Kluwer.

- Culicover, P. (1995). *Adaptive Learning and concrete minimalism*, Ms. Ohio State University.
- de Boysson-Bardies, B. & Vihman, M. (1991). Adaptation to language: Evidence from babbling and first words in two languages. *Language* 67, 297-319.
- Felix, S. (1992). Language acquisition as a maturational process. In J. Weissenborn, H. Goodluck, & T. Roeper (eds.), *Theoretical Issues in Language Acquisition: continuity and change in development*, pp.25-52. Hillsdale, NJ: Lawrence Erlbaum.
- Fodor, J.D.. (1998). Unambiguous triggers. *Linguistic Inquiry* 29, 1-36.
- Gibson, E. & Wexler, K. (1994). Triggers. *Linguistic Inquiry* 25, 407-454.
- Hermon, G. (1990). Syntactic theory and language acquisition. *Studies in the Linguistic Sciences* 20:2, 139-163.
- Hudson, C. & Newport, E. (1999). Creolization: Could adults really have done it all? In A. Greenhill, A. Littlefield, & C. Tano, (eds.), *Proceedings of the Boston University Conference on Language Development* 23(1), 265-276. Somerville, MA: Cascadilla Press.
- Hallé, P.A. & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavioral Development* 19, 463-481.
- Hyams, N. (1986). *Language acquisition and the theory of parameters*. Dordrecht: Reidel.
- Jackendoff, R. (2002). *Foundations of Language*. Oxford: Oxford University Press.
- Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1992a). *Languages of the mind*. Cambridge, MA: MIT Press.
- Jenkins, L. *Biolinguistics*. Cambridge: Cambridge University Press.
- Jusczyk, P.W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.
- Lasnik, H. (2000). *Syntactic structures revisited*. Cambridge, MA: MIT Press.
- Lust, B, S. Flynn, C. Foley, & Y. Chien. (1999). How Do We Know What Children Know? Problems and Advances in Establishing Scientific Methods for the Study of Language Acquisition and Linguistic Theory. In W. Ritchie & T. Bhatia (eds.), *Handbook of Child Language Acquisition*, pp. 427-456. London: Academic Press.
- Montalbetti, M. (1984). After binding: on the interpretation of pronouns. PhD thesis, MIT.
- Moon, C., Cooper, R.P., Fifeer, W.P. (1993). Two-day old infants prefer native language. *Infant Behavioral Development*. 16, 495-500.
- Negrão, E. (1997). Asymmetries in the distribution of overt pronouns and empty categories in Brazilian Portuguese. In J. Black & V. Montapanyane, (eds.), *Current Issues in Linguistic Theory* 140. Amsterdam: John Benjamins.
- Negrão, E. & Müller, A. (1996). As mudanças no sistema pronominal do Português Brasileiro. *DELTA* 12(1):125-152.
- Newport, E. (1990). Maturational Constraints on Language Learning. *Cognitive Science* 14, 11-28.
- Newport, E. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences* 10, 147-172.
- Pinker, S. (1995). Why the Child Holded the Baby Rabbits: A Case Study in Language Acquisition. In L. Gleitman & M. Liberman (eds.), *An Invitation to Cognitive Science, Volume One: Language*, pp. 199-241. Cambridge, MA: MIT Press.
- Poepfel, D. & Wexler, K. (1993). The Full Competence Hypothesis of clause structure in Early German. *Language* 69, 1-33.

- Roeper, T. & Rohrbacher, B. (1995). Null subjects in early child English and the theory of Economy of Projection. *Penn Working Papers in Linguistics*, 2, 83-117.
- Saffran, J., Aslin, R., & Newport, E. (1996). Statistical Learning by 8-month-old Infants. *Science* 274 (13 December): 1926-1928.
- Sakas, W.G. (2000). Modeling the Effect of Cross-Language Ambiguity on Human Syntax Acquisition. In *Proceedings of CoNLL-2000 and LLL-2000*, pp. 61-66.
- Saleemi, A. (2002). Syntax Learnability: the problem that won't go away. In R. Singh, (ed.), *The Yearbook of South Asian Languages and Linguistics*
- Saleemi, A. (1992). *Universal grammar and language learnability*. Cambridge: Cambridge University Press.
- Satterfield, T. & Pérez-Bazán. (in preparation). The dynamic emergence of early bilingualism.
- Satterfield, T. (2003). Economy of interpretation: patterns of pronoun selection in transitional bilinguals. In V. Cook, (ed.), *Effects of L1 on L2*. Clevedon: Multilingual Matters.
- Satterfield, T. (1999). *Bilingual Selection of Syntactic Knowledge: Extending the Principles and Parameters Approach*. Dordrecht: Kluwer.
- Werker, J. F. & Tees, R. C. 1999. Influences on infant speech processing: toward a new synthesis. *Annual Review of Psychology*. 50, 509-535.
- Wexler, K. (1999). Maturation and Growth of Grammar. In W. Ritchie & T. Bhatia (eds.), *Handbook of Child Language Acquisition*, pp. 55-110. London: Academic Press.
- Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106, 23-79.
- Wexler, K. (1990a). Innateness and maturation in linguistic development. *Developmental Psychobiology* 23(3):645-660.
- Yang, C.D. (1999). A selectionist theory of language acquisition. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*.