In Piatelli et al, M. *Of minds and language: the Basque country encounter with Noam Chomsky.* Oxford University Press. 279-295

CHAPTER 18

Remarks on the Individual Basis for Linguistic Structures*

Thomas G. Bever

This paper reviews an approach to the enterprise of paring away universals of attested languages to reveal the essential universals that require their own explanation. An example, discussed at this conference, is the long-standing puzzle presented by the Extended Projection Principle (EPP, Chomsky 1981). I am suggesting an explanation for the EPP based on the learner's need for constructions to have a common superficial form, with common thematic relations, the hallmark of EPP. If one treats EPP phenomena as the result of normal processes of language acquisition, the phenomena not only receive an independently motivated explanation, they also no longer constitute a structural anomaly in syntactic theory.[1]

18.1 EPP and its implications for structural universals

EPP was initially proposed as the structural/configurational requirement that sentences must always have a subject NP, even without semantic content (cf. Chomsky 1981, Lasnik 2001, Epstein and Seely 2002, Richards 2002; see Svenonious 2002, McGinnis and Richards, in press, for general reviews). This principle was first proposed to account for subject-like phrases in sentences, so called expletives (e.g., "it"):

* These remarks are based on what I planned to present at this conference. What follows is influenced by extensive discussions with Noam and the editors. Of course, mistakes and infelicities are all mine.

[1] See discussion by Noam and Massimo of this, pp. 55–57.

(1) a. "it" is raining
b. "there" are three men in the room
c. "it" surprised us that john left
d. "es" geht mir gut
e. "il" pleut

The EPP was initially proposed as a universal syntactic constraint that all languages must respect. While roughly correct for English, a number of troubling facts have emerged:

(2) a. EPP may not be universal (e.g., Irish as analyzed by McCloskey 1996, 2001).
b. Different languages express it differently: e.g., via focus as opposed to subject, in intonation patterns, with different and inconsistent agreement patterns.
c. It generally corresponds to the statistically dominant form in each language.
d. It has not found a formal derivation within current syntactic theory – it must be stipulated.

Accordingly, the EPP may be a "configurational" constraint on derivations – it requires that sentences all conform to *some* typical surface pattern. Epstein and Seeley (2002: 82) note the problem this poses for the minimalist program:

> If (as many argue) EPP is in fact "configurational," then it seems to us to undermine the entire Minimalist theory of movement based on feature interpretability at the interfaces. More generally, "configurational" requirements represent a retreat to the stipulation of molecular tree properties . . . It amounts to the reincorporation of . . . principles of GB . . . that gave rise to the quest for Minimalist explanation . . .

In other words, the EPP is a structural constraint stipulated in the minimalist framework (as well as others), which violates its structural principles and simplicity. Yet EPP-like phenomena exist.

Below I outline a language acquisition model which requires that languages exhibit a canonical form, the Canonical Form Constraint (CFC) – which renders EPP phenomena in attested languages. Thus, there are two potential explanations of EPP phenomena. Either it is indeed a syntactic constraint, part of universal syntax in the narrow faculty of language; or it is a constraint on learnable languages: Sentences have to conform to the CFC – they must sound like they are sentences of the language to afford the individual child a statistical entrée into acquiring it. How can we decide between these two explanations? First, the EPP adds a stipulated constraint to grammars, undercutting their simplicity. Second, the EPP is a heterogeneous constraint, with different kinds of expressions in different languages. Third, the CFC, as we will see, is independently motivated: it explains statistical properties of language, stages of acquisition, and significant facts about adult language processing. Thus, I argue that the phenomena that motivated the EPP are actually expressions of the Canonical Form Constraint (CFC).

Syntacticians may object that this line of reasoning is circular. In many languages, the EPP constraint does not merely exert "stylistic" preferences on sentence constructions, it dictates syntactic requirements on grammatical derivations. But the issue is the source of the constraint that results in processes that conform to the EPP. On my view, the child tends to learn sentence constructions that conform to the canonical form constraint, and not other constructions. The notion of "learn" can be glossed as "discovers derivations for statistically frequent meaning/form pairs, using its available repertoire of structural devices." Thus, in individual languages the child accesses and learns specific derivational processes that conform descriptively to the EPP. But the EPP itself is merely a descriptive generalization reflecting acquisition constraints as its true cause. In the sense of Boeckx (this volume), EPP-like phenomena are among the set of E-universals (corresponding to E-language), not I-universals (corresponding to I-language). In the sense of Hauser et al. (2002), it is a property of the interface between the narrow faculty of language and the acquisition interface.

The following discussion will serve as an outline of how a simplified model of what individuals do during language acquisition, based on a general model of human learning, can explain universal properties of attested languages, such as the EPP. My argumentation strategy here is the following:

(a) a general method of paring down universals, with some non-syntactic examples
(b) a comprehension model showing how the linguistic structures are implemented in an analysis-by-synthesis comprehension model
(c) an application of the analysis-by-synthesis model as a model of acquisition
(d) implications for the Canonical Form Constraint (CFC) as a language universal
(e) implications of the CFC for a correct interpretation of EPP phenomena
(f) implications of this model in general (a potential solution to constraining the abduction of generalizations, and learning grammar as intrinsically motivated problem solving)

This line of argument follows a general research program of isolating true linguistic universals.

> The concept of "language" is like those of . . . "organ", as used in biological science . . . grammatical structure "is" the language only given the child's intellectual environment . . . and the processes of physiological and cognitive development . . . Our first task in the study of a particular [linguistic] structure in adult language behavior is to ascertain its source rather than immediately assuming that it is *grammatically* relevant . . . Many an aspect of adult . . . linguistic structure is itself partially determined by the learning and behavioral processes that are involved in acquiring and implementing that structure . . . Thus, some formally possible structures will never appear in any language because no child can use [or learn] them. (Bever 1970: 279–280)[2]

Here I focus on the dynamic role of the individual language learner in shaping properties of attested languages (aka E-languages). Certain linguistic universals that seem to be structural are in fact emergent properties of the interaction of genetic endowment, social context, and individual learning dynamics. My argument is this: Language acquisition recruits general mechanisms of growth, learning, and behavior in individual children: only those languages that comport with these mechanisms will be learned. I first review some non-syntactic universals, to outline relatively clear examples of the role of development, as background for the main focus of this paper.

18.2 Neurological foundations of language: the enduring case of cerebral asymmetries

The left hemisphere is the dominant neurological substrate for much of language – true for everyone, including the vast majority of left-handers (Khedr et al. 2002). This leads directly to *post hoc propter hoc* reasoning about the biological basis for language: the unique linguistic role of the left hemisphere reflects some unique biological property, which itself makes language possible. This argument has been further buttressed by claims that certain primates have left-hemisphere asymmetries for species specific calls (Weiss et al. 2002), claims that infants process language more actively in the left hemisphere (Mehler et al. 2000), demonstrations that artificial language learning selectively activates the left hemisphere (Musso et al. 2003; Friederici 2004, this volume). However plausible, this argument overstates the empirical case. First, we and others demonstrated that asymmetries involve differences in computational "style" ("propositional" in the left, "associative" in the right; Bever 1975, Bever and Chiarello 1974). In nonlinguistic mammals, the asymmetries may nonetheless parallel those for humans: for example, we have shown that rats learn mazes relying on serial ordering in the left hemisphere, and specific locations in the right (Lamendola and Bever 1997), a difference with the computational flavor

[2] See Cedric Boeckx's quote of Noam's recent reformulation of this approach, Chapter 3 above.

of the human difference. Second, the facts about asymmetries for language could follow from a simple principle: the left hemisphere is slightly more powerful computationally than the right (Bever 1980). Even the simplest sentence involves many separate computations, which during acquisition compound a small incremental computational superiority into a large categorical superiority and apparent specialization. Thus the left hemisphere's unique relation to language function accumulates from a very small quantitative difference in the individual learner.

18.3 Heritable variation in the neurological representation of language

Loss of linguistic ability results from damage to specific areas of the left neocortex. The fact that normal language depends on (rather small) specific areas suggests that it may be critically "caused" by those areas. However, certain aspects of language may have considerable latitude in their neurological representation. For example, Luria and colleagues noted that right-handed patients with left-handed relatives ("FLH+ ") recover faster from left-hemisphere aphasia, and show a higher incidence of right-hemisphere aphasia than those without familial left-handers (FLH−) (Hutton et al. 1977). They speculated that FLH + right handers have a genetic disposition towards bilateral representation for language, which often surfaces in their families as explicit left-handedness. We have found a consistent behavioral difference between the two familial groups in how language is processed, which may explain Luria's observation. Normal FLH+ people comprehend language initially via individual words, while FLH− people give greater attention to syntactic organization (a simple demonstration is that FLH+ people read sentences faster and understand them better in a visual word-by-word paradigm than a clause-by-clause paradigm; the opposite pattern occurs for FLH− people). The bilateral representation of language in FLH+ people may be specific to lexical knowledge, since acquiring that is less demanding computationally than syntactic structures, and hence more likely to find representation in the right hemisphere. On this view, FLH+ people have a more widespread representation of individual lexical items, and hence can access each word more readily and distinctly from syntactic processing than FLH− people (Bever et al. 1987, 1989a; Townsend et al. 2001).

This leads to a prediction: lexical processing is more bilateral in FLH+ right-handers than FLH− right-handers, but syntactic processing is left-hemisphered for all right-handers. Recently, we tested this using fMRI brain imaging of subjects while they are reordering word sequences according to syntactic constraints or according to lexico-semantic relations between the words. We found that the lexical tasks activated the language areas bilaterally in FLH+ right-handers, but activated only the left hemisphere areas in the FLH− right handers: all subjects showed strong left-hemisphere dominance in the syntactic tasks (Chan et al. in preparation). This confirms our prediction, and supports our explanation for Luria's original clinical observations. It also demonstrates that there is considerable lability in the neurological representation of important aspects of language.

18.4 The critical period: differentiation and segregation of behaviors

The ostensible critical period for learning language is another lynchpin in arguments that language writ broadly (aka E-language) is (interestingly) innate. The stages of acquisition and importance of exposure to language at characteristic ages is often likened to stages of learning birdsong – a paradigmatic example of an innate capacity with many surface similarities to language (Michel and Tyler 2005). However, certain facts may indicate a somewhat less biologically rigid explanation. First, it seems to be the case that adult mastery of semantic structures in a second language is much less restricted than mastery of syntax, which in turn is less restricted than mastery of phonology (Oyama 1976). This decalage invites the interpretation that the critical period is actually a layering of different systems and corresponding learning sequences. Phonological learning involves both tuning perceptual systems and forming motor patterns, which is ordinarily accomplished very early: linguistically unique semantic knowledge may be acquired relatively late, draws on universals of thought, and hence shows relatively little sensitivity to age of acquisition.

Noam suggested (in email) a non-maturational interpretation of this decalage, based on the specificity of the stimulus that the child receives, and the corresponding amount which must be innately available, and hence not due to different mechanisms of learning with different time courses. The semantic world is vast: much of semantics must be universally available innately, and hence a critical period for semantic acquisition is largely irrelevant. In contrast, all the phonological information needed for learning it is available to the child, and can be learned completely in early childhood.

The notorious case is syntactic knowledge of an explicit language, which is neither determined by sensory/motor learning nor related directly to universals of thought. I have argued that the critical period for syntax learning is a natural result of the functional role that syntax plays in learning language – namely, it assigns consistent computational representations that solidify perceptual and productive behavioral systems, and reconciles differences in how those systems pair forms with meanings (Bever 1975, 1981). On this view, the syntactic derivational system for sentences is a bilateral filter on emerging perceptual and productive capacities: once those capacities are complete and in register with each other, further acquisition of syntax no longer has a functional role, and the syntax acquisition mechanisms decouple from disuse, not because of a biological or maturationally mechanistic change. (See Bever and Hansen 1988 for a demonstration of the hypothesis that grammars act as cognitive mediators between production and perception in adult artificial language learning).

This interpretation is consistent with our recent finding that the age of the critical period differs as a function of familial handedness: FLH+ deaf children show a younger critical age for mastery of English syntax than FLH− children (Ross and Bever 2004). This follows from the fact that FLH+ people access the lexical structure of language more readily, and access syntactic organization less readily than FLH− people: FLH+ children are acquiring their knowledge of language with greater emphasis on lexically coded structures, and hence depend more on the period during which vocabulary grows most rapidly (between 5 and 10 years: itself possibly the result of changes in social exposure, and emergence into early teenage). Consistent with my general theme, it attests to the role of general mechanisms of learning and individual neurological specialization in shaping how language is learned.

18.5 Language learning as hypothesis testing and the EPP

Of course, how language learning works computationally is the usual determinative argument that the capacity for language is innate and independent from individual mechanisms of learning or development. Typically cited problems for a general inductive experience-based empiricist learning theory are:

(3) a. The poverty of the stimulus. How do children go beyond the stimulus given?
  b. The frame problem: how do children treat different instances as similar?
  c. The motivational problem: e.g., what propels a 4-year-old to go beyond his already developed prodigious communicative competence?
  d. The universals problem: how do all languages have the same universals?

Parameter-setting theory is a powerful schematic answer to all four questions at the same time. On this theory, a taxonomy of structural choices differentiates possible languages. For example, phrases are left- or right-branching; subjects can be unexpressed or not; wh-constructions move the questioned constituent or it remains in situ. The language-learning child has innate access to these parameterized choices. Metaphorically, the child has a bank of dimensionalized "switches" and "learning" consists of recognizing the critical data setting the position of each switch: the motivation to learn is moot, since the switches are thrown automatically when the appropriate data are encountered. This is a powerful scheme which technically can aspire to be explanatory in a formal sense and has made enormous contributions in defining the minimally required data (Lightfoot 1991; Pinker 1984; J. D. Fodor 1998, 2001; Fodor and Sakas 2004; Fodor, this volume): but it is also very far removed from the motivational and daily dynamics of individual children. We are left with an abstract schema and no understanding of what the individual child might be doing, why it might be doing it, and how that activity might itself constrain possible choices of parameters, and hence, attested linguistic universals.

My hypothesis, and that of a few others who accept the idea that children in fact acquire generative grammar (e.g., Gillette et al.1999; Gleitman 1990, this volume; Papafragou et al. 2007) is that neither a parameter-setting scheme, nor inductive learning alone is adequate to the facts. On this view, acquisition involves both formation of statistical generalizations available to the child and the availability of structures to rationalize violations of those generalizations. A traditional view of this kind is "hypothesis testing," which allows for hypotheses to be inductively generated and deductively tested, and conversely.

Now to the central thesis of this discussion: there is a model of acquisition that integrates inductive and deductive processes; such a model requires the existence of canonical forms in languages; this motivates the facts underlying the Extended Projection Principle, which requires that (almost) every sentence construction maintain a basic configurational property of its language. The exposition starts with a narrowly focused discussion of how inductive and deductive processes can be combined in a model of comprehension – itself experimentally testable and tested with adults. Then I suggest that this kind of model can be generalized to a model of acquisition, with corresponding empirical predictions – at least a few of which are confirmed.

18.6 Integrating derivations into a comprehension model

The first question is, do speakers actually use a psychological representation of generative grammar – a "psychogrammar" – of the particular form claimed in derivational models, or only a simulation of it? If adult speakers do not actually use the computational structures posited in generative grammars as part of their language behavior, we do not have to worry about how children might learn it. In fact, fifty years of research and intuition have established the following facts about adult language behavior (4):

(4) a. Syntactic processes in generative models are "psychologically real": derivational representations are used in language comprehension and production (see Townsend and Bever 2001).
b. Syntactic processes are recursive and derivational: they range over entire sentences in a "vertical" fashion (as opposed to serial) with successive reapplications of computations to their own output. These properties have been true of every architectural variant of generative grammar, from Syntactic Structures (Chomsky 1957), to the minimalist program (Chomsky 1995).
c. Sentence behavior is instant and "horizontal" – speakers believe that they comprehend and produce meaningful sentences simultaneously with their serial input or output. Comprehension does not start only at the end of each sentence: production does not wait until a sentence is entirely formulated.

These three observations set a conundrum:

(5) a. Sentence processing involves computation of syntax with whole sentences as domain – it is *vertical*.
b. Language behavior proceeds serially and incrementally – it is *horizontal*.

Recently, Dave Townsend and I rehabilitated the classic comprehension model of Analysis by Synthesis (AxS) that provides a solution to the conundrum (following Halle and Stevens 1962, Townsend and Bever 2001). On this view, people understand everything twice: once based on the perceptual templates; once by the assignment of syntactic derivations. In the AxS architecture the two processes are almost simultaneous. First, the perceptual templates assign likely interpretations to sentences, using a pattern completion system in which initial parts of a serial string automatically trigger a complete template. Typical templates of this kind are:

(6) a. Det . . . X → np[Det . . . N]np
b. NP V(agreeing with NP) (optional NP) → Agent/Experiencer Predicate (object/adjunct)

Second, the initially assigned potential meaning triggers (and constrains) a syntactic derivation. The two ways of accessing meaning and structure converge, roughly at the ends of major syntactic units. That is, as we put it, *we understand everything twice*. The model has several unusual features (Townsend and Bever 2001). First, the model assigns a complete correct syntax *after* accessing an initial meaning representation. Second, that meaning is sometimes developed from an incorrect syntactic analysis. For example, syntactic passives (7a) are initially understood via the variant of the canonical sentence template in (6b) that applies correctly to lexical passives (7b); raising constructions (7c) are understood initially via the same kind of misanalysis.

(7) a. Syntactic passive: Bill was hit
    b. Lexical passive: Bill was hurt
    c. Raising: Bill seemed happy
    d. Control: Bill became happy

The schema in (6b) initially misassigns "hit" as an adjective within a predicate phrase. That analysis is sufficient to access semantic information modeled on the interpretation template for lexical passive adjectives – a syntactic misanalysis. This analysis is then corrected by accessing the correct derivation. This sequence of operations also explains the fact that the experimental evidence for the trace appears in syntactic passives and raising constructions only after a short time has passed (Bever and McElree 1988, Bever et al. 1990, Bever and Sanz 1997). This model also explains a number of simple and well-known facts. Consider the following examples:

(8) a. The horse raced past the barn fell
    b. More people have been to Russia than I have

Each of these cases exemplifies a different aspect of the AxS model. The first reflects the power of the canonical form strategy in English (6b), which initially treats the first six words as a separate sentence (Bever 1970). Native speakers judge this sentence as ungrammatical, often even after they see parallel sentences with transparent structure:

(9) a. The horse ridden past the barn fell
    b. The horse that was raced past the barn fell
    c. The horse racing past the barn fell

The example is pernicious because recovering from the misanalysis is itself vexed: the correct analysis includes the garden-pathing proposition that "the horse raced" (i.e., was caused to race): thus, every time the comprehender arrives at the correct interpretation she is led back up the garden path.

Example (8b) (due to Mario Montalbetti) is the obverse of the first example. The comprehender thinks at first that the sentence is coherent and meaningful, and then realizes that it does not have a correct syntactic analysis. The initial perceptual organization assigns it a schema based on a general comparative template of two canonical sentence forms – "more X than Y," reinforced by the apparent parallel Verb Phrase structure in X and Y ("…have been to

Russia . . . I have"). On the AxS model, this superficial initial analysis triggers the derivational parse system, which ultimately fails to generate a derivation.

I do not expect to have convinced the reader of our model via such simplified examples: in our book, we organize a range of often surprising experimental and neurological facts supporting an early stage of comprehension resting on frequent statistically valid templates, followed by a structurally correct syntactic derivation (Townsend and Bever 2001, Chapters 5–8; see Friederici, this volume, for imaging data consistent with this bi-phasic model of comprehension).

This model requires languages to have certain universal features. Most important is the otherwise unmotivated fact that actual languages have a characteristic set of statistically grounded structural patterns at each level of representation (phonological, morpho-lexical, syntactic). It further requires that complex constructions be functionally homonymous with simpler constructions in ways that allow the simpler constructional analysis to convey the more complex meaning at an initial pre-derivational stage of processing. The model is inelegant in that it solves the conundrum (5) by fiat – sentence processing is both fast and complex because it is simultaneously handled by two systems, one fast and sometimes wrong, one slower but ultimately correct. This is an inelegant solution to the conundrum, but shows that humans may solve it, albeit inelegantly.

18.7 AxS in language acquisition – the Canonical Form Constraint

Two historically competing principles about the mind have alternately dominated the cognitive sciences:

(10) a. Everything we do is based on habits.
b. Everything (important) we do is based on creative rules.

The AxS model architecture shows how the two insights might be integrated together in adult behavior. A corresponding model holds for the acquisition of language. On that model, the child alternates (logically) between formulating statistical generalizations about the language, and assembling derivational operations that account for those generalizations. Many researchers are demonstrating that child-directed speech in fact has statistical regularities that might guide the infant and child towards language (e.g., Curtin et al. 2005, for segmentation, Golinkoff 1975; Brent 1997; Cartwright and Brent 1997; Gerken 1996; Golinkoff et al. 2005; Mintz 2002, 2003, 2006; Redington et al. 1998). At the same time, infants are quite good at detecting statistical patterns from serial strings with specific kinds of structure (Gomez and Gerken 1999;

Marcus et al. 1999; Saffran 2001, 2003; Saffran et al. 1996); older children also show statistical sensitivity in developing grammatical and lexical ability (Bates and MacWhinney 1987, Gillette et al. 1999, Moerk 2000, Naigles and Hoff-Ginsburg 1998, Yang 2006). If one component of syntax acquisition is the compilation of relevant generalizations, this model requires that the child be presented with some statistical regularities in the language he hears. This requirement explains several computationally eccentric facts about attested languages:

(11) a. Each language has a canonical surface form: in English this is schematically as presented in the left side of the expression in (6b).
b. Statistically, the canonical form has a dominant assignment of semantic relations: in English this is the template we found explanatory for much adult language behavior (6b).
c. The canonical semantic interpretation is violated in a set of minority constructions: in English, this includes passives, raising, unaccusatives, middle constructions.
d. The minority constructions that violate the form can nonetheless be approximately correctly interpreted by application of the canonical form interpretation. (This is exemplified in the initial stages of comprehending syntactic passives and raising constructions, discussed above in examples (7).)

None of these properties follows from the computational architecture of any of the last fifty years of generative grammar. Yet they are characteristic of attested languages. In English, the first property has been noted as the result of rule "conspiracies," which guarantee that sentences have the same surface form regardless of their thematic relations and derivation. The vast majority of sentences and clauses have a canonical form with a subject preceding a correspondingly inflected verb:

(12) a. The boy hits the ball
b. The ball was hit by the boy
c. It is the boy who hits the ball
d. The boy was happy
e. The boy seems happy
f. The boy was eager to push
g. The boy was easy to push
h. It was easy to push the boy
i. The boy pushes easily
j. Who pushed the boy?
k. Who does the boy push?

The notion of such conspiracies is not novel, be it in syntax or phonology (cf. Ross 1972, 1973a,b). In traditional derivational terms, it reflects constraints on derivations such that they have the same general surface form regardless of differences in logical form or semantic relations. This is despite the fact that each underlying form could be reflected in a unique surface sequence or signaled by a specific marker. On our interpretation, such computationally possible languages would be allowed by generative architectures, but are not learnable: they would make it hard for the language-learning child to develop a statistically based pattern that it can internalize and use for further stages of acquisition.

The canonical form (11a) facilitates the discovery of a surface template based on statistical dominance of the pattern. The semantic schema (11b) above relates the canonical form to a standard interpretation – although a majority of individual constructions may not conform to that schema, the vast majority of actual utterances in corpora do so – another fact about languages unexplained by generative architectures. The third fact (11c) – that some cases violate the canonical semantic interpretation of the canonical form – is particularly important if the child is eventually to discover that there are actual derivations in which a given surface form expresses different patterns of thematic relations. Finally, the interpretability of the schema-violating constructions via misanalysis and homonymy with simpler constructions (11d) – contributes to the child's ability to interpret sentence types for which it does not yet have a syntactic analysis. I summarize the set of these conditions as the "Canonical Form Constraint (CFC)."

The CFC suggests a way in which the child can transcend the "poverty of the stimulus." First, the child can create and then analyze his own set of form/meaning pairs going beyond the actual sentences he hears, based on these generalizations. Second, this solves an important problem for any learning scheme – how do children remember and understand sentences for which they do not yet have a correct syntactic analysis? (Valian 1999). It would not work for the child to maintain a list of grammatically unresolved sentences: any given list is heterogenous without prior structural ordering. The AxS model suggests that children can rely on statistical patterns and occasional false analyses to generate an internal bank of meaning/form pairs and maintain an internalized data bank to evaluate candidate derivational analyses. This reduces the need for children to access positive and negative feedback as guides to their emerging syntactic abilities. On this view, the child can attempt derivation of a construction based on a subset of sentences of a given general pattern, and then "test" the derivational structure on other sentences of a similar pattern. (For related ideas, see Chouinard and Clark 2003, Dale and Christiansen 2004, Golinkoff et al. 2005, Lieven 1994, Moerk 2000, Morgan et al. 1995, Saxton 1997, Valian 1999). These facts and considerations offer an explanation for the CFC – peculiar in the sense that the computational architecture of syntax does not in itself require the CFC. It is reflected in attested languages because it makes them learnable, using a hypothesis formation procedure.

If this picture is correct, children should show evidence of actually learning perceptual strategies, based on statistical frequency of preponderant features of their surrounding language. We and others have found evidence supporting this (Bever 1970, Maratsos 1974, Slobin and Bever 1982). The original finding was based on having children act out simple sentences with puppets. (Typical data are summarized in Table 18.1). 2-year-old children use a simple strategy that focuses primarily on the exact sequence NounPhrase + Verb, interpreting that as Agent + Verb. Thus, at age 2, children interpret declarative and object cleft sentences, along with semantically unlikely sentences, above chance: in these constructions, the noun immediately before the verb is the agent. By age 3–4, they rely both on a more elaborated analysis of word order and semantic strategies:

(13) a. #N . . . = Agent
b. Animate nouns are agents, inanimate nouns are patients

(13a) represents a shift from assigning the noun immediately before the verb as agent, to assigning the first noun in the overall sequence as agent. This produces correct performance on simple declarative sentences, but a decrease in performance on sentence types in which the first noun phrase is not the agent (object clefts and passives).

The emergence of the two kinds of strategies accounts for the decrease in performance on semantically reversible sentences that violate the CFC

Table 18.1 Percentage correct interpretations of simple sentences by children[a]

	Age 2	Age 4
SEMANTICALLY REVERSIBLE		
The dog bit the giraffe –	90%	98%
It's the giraffe that the dog bit –	87%	43%
The giraffe got bit by the dog	52%	27%
SEMANTICALLY IRREVERSIBLE		
The dog ate the cookie	92%	96%
The cookie ate the dog	73%	45%
The cookie got eaten by the dog	55%	85%

[a]Children make small puppets act out short sentences. The primary measure is which noun is the agent and which the patient: chance performance is 50%.

(the giraffe was kicked by the dog). The emergence of reliance on semantic information accounts for the increase in performance on sensible sentences (the dog ate the cookie), and the decrease in performance on semantically odd sentences (the cookie ate the dog). The reliance on semantic factors at age 4 also can override the word-order strategy, leading to correct performance on irreversible passives (the cookie got eaten by the dog).

The perceptual strategies differ from language to language: we found that by age 4, children acquire processing strategies adaptive to the statistical regularities in the structure of their own language (Slobin and Bever 1982). Thus, in English what develops is sensitivity to word order, in Turkish, sensitivity to patient/object inflectional markers, in Italian and Serbo-Croatian, sensitivity to a mixture of the two kinds of linguistic signals. This reflects the fact that each language has its own CFC, which children learn.

18.8 Coda: Some broader implications of the AxS acquisition model

The following points are in large part the result of email discussions with Noam.

18.8.1 Language acquisition as enjoyable problem solving

The idea that the child acquires knowledge of syntax by way of compiling statistical generalizations and then analyzing them with its available syntactic capacities is but another instance of learning by hypothesis-testing. For example, it is technically an expansion on the TOTE model proposed by Miller et al. (1960). An initial condition (statistically grounded pattern) triggers a TEST meaning, and an OPERATION (derivation) which triggers a new TEST meaning and then EXIT. Karmilov-Smith and Inhelder (1973) advanced a different version – cognition advances in spurts, triggered by exposure to critical instances which violate an otherwise supported generalization.

The dual nature of the acquisition process is also related to classical theories of problem solving (e.g., Wertheimer 1925, 1945). On such models, the initial stage of problem organization involves noting a conceptual conflict – for example, "find a solution that includes both X and Y: if the answer is X then Y is impossible, but if Y then X is impossible": characteristically the solution involves accessing a different form of representation which expresses the relation between X and Y in more abstract terms. In language the initial conflict expresses itself as the superficial identity of all the constructions in (12) which exhibit the canonical form constraint, while assigning different semantic relations; the resolution is to find a derivational structure for the set that shows how the different surface constructions are both differentiated and related derivationally. Hence, not only is language-learning hereby interpreted in the context of a general set of learning principles, it is also interpreted as a special instance of a general problem solver. This also explains why language learning is fun, and hence intrinsically motivating: the gestalt-based model suggests that language-learning children can enjoy the "aha" insight experience, an intrinsically enjoyable sensation which may provide critical motivation to learn the derivational intricacies of language (cf. Weir's 1962 demonstration that children play with their language paradigms when they are alone).

Note that the terms "motivation" and "fun" are technical terms based in aesthetic theory, not the everyday notion of conscious desire, nor any notion of "reinforcement." Elsewhere, I have developed analyses of what makes objects and activities intrinsically enjoyable (Bever 1987). The analysis draws on the classic aesthetic definition: stimulation of a representational conflict which is then resolved by accessing a different form or level of knowledge. The formal similarity of this definition to the gestalt model of learning affords an explanation of why aesthetic objects are enjoyable: they are mini-"problems" involving conflicting representational solutions, resolved by accessing a level which creates a productive relation between those solutions, thereby eliciting a subconscious "aha." This kind of analysis is ordinarily applied to serial arts such as drama or music, in which the representational conflict and its resolution can be made explicit over time. But the analysis works for static objects, explaining the preference, for example, for the golden mean rectangle. In language, one kind of conflict is elicited by the thematic heterogeneity of superficially identical surface phrase structures: the child's resolution of that conflict requires access to an inner form of the sentences, via distinct derivational histories – a resolution which involves accessing a distinct level of representation. Thus, learning the structure of a language elicits a series of mini-ahas in the child, making it an activity which is intrinsically attractive.

The model also offers a partial answer to the frame problem (see Ford and Hayes 1993), the problem of how statistical generalizations are chosen out of the multiple possibilities afforded by any particular set of experiences. This problem was classically addressed by Peirce (1957) as the problem of abduction, who argued that there must be constraints on all kinds of hypotheses, even those ostensibly based on compilation of observations (cf. Chomsky 1959c, on the corresponding problem in S-R associative theory, and this volume). But the problem is also a moving target for the language-learning child. At any given age, the generalizations that are relevant to progress in learning are different: if the child has mastered simple declarative constructions, or some subpart of her language's inflectional system, this changes the import of further exposure to the language. Thus, we must not only address constraints on the initial state of the child (see Mehler and Bever 1968 for discussion), we must address how constraints apply to each current state of knowledge, as the child matures and acquires more structural knowledge. That is, the abductive constraints themselves have a developmental course. By what process and dynamic? Another way of putting this is, what filters (aka "frames") possible generalizations and how does the filter itself change as a function of current knowledge?

In the AxS scheme, there are two kinds of processes which filter generalizations. First is the set of salient regularities among elements that are available to the input: at a phonological level, infants have available perceptual categories that provide an initial organization of the input; this affords an innate categorization of sound sequences, available for formal derivational analysis. The other side of the filtering process is the set of computational devices available to provide a derivation. That is, those generalizations about sound sequences that endure are just those that can be explained by a set of possible computational phonological rules. Such rules must have natural domains (presumably innately determined) such as segmental features, syllabic structures, lexical templates. At the syntactic level, the corresponding problem is to isolate a natural segmentation of the potential compositional input. To put it in terms of the example we are focusing on, how does the system isolate "NP V (NP)" as a relevant kind of sequence over which to form a generalization? In the model proposed, the solution lies in the fact that the derivational component has its own natural units, namely clause-level computations. The result is that the derivational discovery component acts as a filter on the multiple possible statistical generalizations supported by any finite data set, picking out those that fit the derivational templates. Most important is that the properties of the derivational filter change as the knowledge base increases in refinement.

18.9 Finale: Biolinguistics and the individual

Recent discussions, and this conference, have clarified current linguistics as "biolinguistics," the isolation and study of genetic endowment and boundary conditions on the faculty of language. The formal approaches to isolating and explaining universals via abstracted biological constraints on what language is, or by examining the data required to set parameters in an ideal learner, clarify the relevant abstract conditions on individuals learning language. Yet it is a collection of concrete individuals that learn and use language. Thus, these boundary conditions may profit from inclusion of the motivations and actions of individual learners. I have given various kinds of examples of linguistic universals, showing how we can benefit by examining the dynamics of language learning in individuals. The extent to which individuals learn language by way of mechanisms not specific to language alone clarifies what we should take as the essential universals of language. The discussion in this paper of EPP is an example of this kind of argument.