Minimalist Behaviorism: the role of the individual in explaining language universals

T.G.Bever
Dept of Linguistics
304 B Communication
University of Arizona
tgb@u.arizona.edu

in:

Christiansen, M., Collins, C., Edelman, S., *Language Universals,* Oxford, (2009). pp. 99-126

1. Background – the evolution of grammars

Since the late 1950s, it has been accepted that many universal properties of attested languages are due to behavioral constraints on language performance: these included the boundedness of sentence and word length, the limits on the number of words in the lexicon, restrictions on the amount of ambiguity in sentences, and so on. Excluding such performance-based universals isolates the computational processes involved in describing linguistic knowledge. The early phases of transformational syntax focused on a computational model of that knowledge. This model posited several enduring ideas about the computational basis for language:

1) a) Sentences involve derivations, from primitive/partial structures to surface forms.
   b) Syntactic operations are cyclic, i.e. recursive, applying successively to their own output.

This set the role of behavioral theory and research within the Language Sciences. Initially, the question was about "the psychological reality of linguistic structures". That is, did the computational model involving transformations and derivations correctly describe what speakers deploy when they use language, not just descriptively, but in terms of the actual computational operations. This question dominated the first phase of modern psycholinguistics, lead by George Miller and his students (Miller, 1962). Indeed it appeared initially that the transformational/ derivational model was psychologically valid, sentences with more transformations are harder to process, sentences with shared underlying structures are perceived as related, and so on.

This initial success was overtaken by a rapid evolution of syntactic theory both with respect to its technical operations, and its related ontology. The major settled stages of generative syntactic theory between 1955-70 and 1980-90 were:

(2)

a) Specific phrase structure generates underlying structures: many specific movement rules correspond to construction types (e.g., "passive, raising, question," etc.). Derivations are 'guaranteed' to be correct by virtue of the application of the sets of rules.

b) 1980-90) X-bar theoretic phrase structure is projected from lexical categories. Movement occurs where it can or is required by filtering constraints on derivations ("case theory", "binding principles"….). A set of parameters describe a small set of options how these filters and processes operate: each language has its own setting on each of these parameters).

At a technical level, derivations became less specific, less dependent on being triggered by specific structural indices, and more automatic: the corresponding ontological shift was from viewing syntax as describing the language that people know, to describing the internal processes that result in language. (For recent reviews of the development of generative syntax see Townsend and Bever, (2001), Chapter 3; Chomsky, (2004); Freidin and Vergnaud, (2001); Lasnik, (2003); Hornstein, Nunes and Grohman (2005); Martin and Uriagereka, (2000); Vergnaud, (1985); Boeckx, (2006)) The – perhaps ultimate – development in the trend to simplify syntactic theory is today's 'minimalist program'. This approach explicitly gives up the claim to be a particular theory of knowledge or processes of any kind. Rather, it is a framework for isolating the minimally required processes that could result in language. The approach takes as a boundary condition on possible languages the principle that language is an interface between humans' conceptual organization and motor systems. This idea was presaged in the early stages of generative grammar: for example, Miller and Chomsky (1963) noted that the existence of transformations and constraints on them that preserve recoverability of their source structures, followed from the fact that language maps complex propositional structures, onto a linear sequence. Today's minimalist program makes this feature of language the central constraint. The goal then is to discern the essential operations that meet that constraint as perfectly as possible (Chomsky, (1995; 2005); Lasnik, (2003)).

This has started a re-named, if not totally new kind of approach to the study of language universals, "biolinguistics". In this enterprise, the attempt is to show how a small number of minimally required processes can account for the essential computational operations in language. This is the focus of the "faculty of language". All other properties of attested languages are to be interpreted as a function of biological, psychological, even social constraints. (Hauser, et al, (2002)).

The main operation is cyclic "merge" of trees to form constituent hierarchies and derivations: if the head of a phrase is the same kind of constituent as the daughter of another phrase, the head can be merged with the daughter to form a more complex phrase. The process is cyclic in the sense that trees are constructed by iteration of the merge process from the most to least embedded phrase of a derivation. The process is recursive in the sense that it is possible for the daughter of one phrase to be the head of another. It is striking, that these major computational features of the faculty of language now proposed within the minimalist program are essentially those in (1) above, except for X-bar theory, which is now itself rendered by cyclic merge. But the basic notion of cyclic operations has been the enduring properties of every variant of generative syntax for fifty years.

The simplification of grammatical processes has heightened the puzzle of language acquisition and its potential genetic basis. Traditional learning theory – essentially induction – can not account for the discovery of most abstract features involved in sentence derivations. A new theory of learning is needed, and the default is "parameter setting"; that is, the child has a set of innate language-constraining parameters, and simply needs to pick up enough data from the environment to trigger one or another setting of each parameter. (Pinker, (1984); Fodor, (2001); Fodor and Sakas, (2004), Yang, (2002)).

How is the role of psychology in the explanation of linguistic universals changed by the developments within linguistic theory? In particular, has "biolinguistics" gone past "psycholinguistics", explaining everything of interest about language as a function of a few biological constraints and boundary conditions on the faculty of language, along with innate parameters that facilitate learning? If it has not yet done so, could it ever do so?

In what follows, I suggest that the answer to both questions is no. Formal approaches to explaining universals via abstracted biological constraints on the function of language, or by examining the data required to set parameters in an ideal learner, are limited to clarifying the boundary conditions on individuals learning and using language. Yet it is a collection of individuals that learn and use language, and we can profit by examining how they do this. The extent to which they do so by way of mechanisms not specific to language, clarifies what we need to keep looking for as the essential universals of language.

2. Some kinds of psychological constraints on language universals

We can distinguish three categories of language universals that may have explanations outside of the faculty of language:

(3)

a) constraints from language use
b) constraints from language acquisition
a) constraints from the neurological substrate

2.1 Constraints from language use.

Historically, many constraints on attested language universals have been based on models of sentence comprehension. An early, classic instance of this was Chomsky and Miller's (1963) discussion of center embedding in English: they noted that one choice was to include a "recursion counter" as part of syntactic theory, limited to 2 in the case of center embeddings, but with no limit for right (or left) branching recursions. This would have increased the options available to grammars in many undesirable and unutilized ways. So the preferred alternative interpretation was that a center-embedded sentence is hard to understand because it requires extra processing memory to compute the output of an operation on an item that is itself incompletely computed by the same operation. While I think this analysis is not entirely correct, it exemplifies several principles at work together in deciding if a Universal, Ui should be attributed to syntax or some other source.

a) Ui would require an otherwise unmotivated computational process i.e. the ability to count recursions
b) Ui can be explained as a special case of a process motivated by some extra-linguistic-structure system

General ideas on how comprehension works have been alleged to account for a number of syntactic universals and constraints. A simple example is heavy XP shifts, in which there is preference (sometimes a requirement) that a complex argument phrase be moved from its base (or semantically local) position, to a position towards the end of a sentence.

5a) That Bill is in charge is likely - > it is likely that bill is in charge

5b) For Bill to be in charge is likely – (it) is likely for Bill to be in charge -> Bill is likely to be in charge

In these cases, there is a clear preference (for some speakers a requirement) that the complex phrase be placed at the end of the sentence. This exemplifies a simple principle, "save the hardest for the last" (Bever, 1970), on the assumption that phrase tree structures are assigned as part of comprehending a sentence. If a complex sub-tree appears first, it must be held in memory while assigning a later simpler sub-tree, until the entire sentence tree structure is computed. But if the complex tree appears at the end of the sentence, its structure can be assigned as part of completing the entire sentence structure, and hence does not have to be held in memory. This constraint explains a number of cases of preferred and sometimes required constructions.

Hawkins (this volume) has argued that heavy phrase shifts actually vary according to the head/modifier order characteristic of each language: English is head-initial with a basic pattern of "Head + modifier". But Japanese is a head-final language with the opposite order: and Hawkins notes Japanese constructions are often preferred with complex phrases prior to simpler ones. If true, this may show that the original "save the hardest for the last" principle is not based on the surface seriality of sentence input: rather it is sensitive to the sequence of steps in phrase assignment, in which the head is first posited and then modifiers are attached to it. Hawkins proposes a more general constraint on head-modifier distance to explain the facts: which view is correct awaits further research. But both views share the concept that the order constraint follows from aspects of how sentences are processed in comprehension.

I will not spend further time here on such well-explored examples of how surface processes can constrain sentence constructions. Interested readers can consult Bever (1970), Bever and Langendoen, (1971), Bever Carrol and Hurtig (1976), for early examples of such ideas, and Hawkins (this volume) for other instances. The basic moral is that sentence constructions that cannot be understood or are hard to understand because of constraints on a serial comprehension process, will not be learned. In some cases, such as center-embedded sentences, this may block certain instances of otherwise grammatical sentences. In other cases, such constraints may be argued to limit the possibility of certain kinds of syntactic operations. This leads us to consider models of how sentence comprehension actually works.

2.1.1 – The integration of associations and rules in sentence comprehension

Current researchers on language comprehension generally start with one of two powerful and appealing ideas:

> 6a) Meaning is extracted via associatively learned patterns
> 6b) Meaning is extracted from syntactic structures

The associationist view dominated psycholinguistics (such as it was) for many years, until the Cognitive Revolution of the 1960s: it has been given new life in the form of computationally enriched connectionism (old fashioned associationism, plus various schemes for multi-level training, yielding a computationally tractable mediation SR theory – at least more tractable than the earlier ideas of Clark Hull, Charles Osgood and colleagues). The idea is roughly that insofar as syntactic structures play a role in comprehension, they do so via the application of learned pattern-templates. Problems with this view abound and are well understood. Major ones include: The proper domain problem (what is the proper domain of an associative template?), how are overlapping and competitive templates organized, the recursive nature of syntactic structures. (See the articles in Pinker and Mehler, (1988); also Steedman (this volume)). Nonetheless, while connectionist models are not adequate for the structure of language, they do capture an important property of behavior in ways more sophisticated than prior implementations of associationism – i.e., that much of behavior depends on habits.

Realization of the importance of syntactic structures underlies many variants of syntax-first models of comprehension: on these models, syntax is assigned (somehow), and then meaning, context, knowledge, and other modalities of information are integrated with the assigned syntactic structure. These models characteristically give no initial causal role to statistical information, including eccentricities in the frequency of particular syntactic structures. Rather, syntax is first assigned based on structure building principles, and then statistical information and knowledge of all kinds can play a role in determining the final representation of meaning. The difficulty with these models is the persistent undeniable fact that statistical properties of sentences and meaning do appear to play some kind of immediate role in comprehension. (See Townsend and Bever, 2001, Chapter 4).

The inadequacy of each kind of comprehension model alone reflects a duality in sentence processing: There are two well-known facts about language comprehension which require explanation by any adequate theory:

7)

a) it is very fast: words in sentences are extra clear acoustically
b) syntactic derivational histories are assigned as part of comprehension

These two kinds of facts reflect a conundrum:

8)

a) Sentences are horizontal, from "left" to "right"
b) Derivations are "vertical", (i.e., if they were serial they would sometimes span entire sequences, often "right" to "left" as in 'raising' operations)

It is interesting and significant that each of these facts is explained by a corresponding view on processing: Associative templates are excellent at rapid pattern completion and apply immediately, going from signal to meaning. Syntactic structures as assigned build their derivations.

Townsend and Bever (2001) rehabilitated an Analysis by Synthesis model that embraces both kinds of information (Halle and Stevens, (1962, 1964)). On this model, sentences are initially assigned a functional structure and meaning, based on statistically dominant patterns; they are separately assigned a derivational history. In the ordinary run, the latter follows the former, perhaps by 100 milliseconds – that is, the model assigns correct syntax last, not first. This model has several features that are surprising in light of the goals of linguistics and presumptions about behavior:

- it is inelegant, simply gluing together the two kinds of processes
- it involves understanding every sentence at least twice
- it involves assigning a correct structure last
- initial meaning representations can be based on initially incorrect structures

We adduced a full range of existing and new facts to support the model, indeed to support several of the surprising features. The reader is invited to consult the book for a full description. Here I focus on one case study, the comprehension of syntactic passives. Consider (9-12)

(9a) Athens was attacked
(9b) Athens was ruined

Classically, the passive form of verbs can be differentiated into 'syntactic' vs. 'lexical' passives. The latter distribute in the same way as normal (stative) adjectives, motivating their categorization as lexically coded stative-like adjective forms.

10a) *Athens was quite attacked
10b) *Athens looked attacked
10c) Athens was being attacked

11a) Athens was quite ruined
11b) Athens looked ruined
11c) *?Athens was being ruined

The corresponding surface forms from a derivation in a theory that includes traces from movement looks schematically like the following:

12a) Athens was attacked [t-Athens]
12b) Athens was ruined

Various studies have shown that there is some evidence that the trace is actually present in the mental representation of sentences with syntactic passives, and not present in sentences with lexical passives. Typical studies show that shortly after the trace, the antecedent of the trace is more salient, e.g., in a word probe paradigm. At the same time, the evidence suggests that the trace does not acquire its force in the representation immediately, but only after about a tenth of a second (McElree and Bever (1989), Bever and Sanz, (1997).

These facts are given a handy explanation in the analysis by synthesis model. On that model, both kinds of "passives" are initially understood via a variant of the canonical sentence schema for English:

(13) N V (N) => agent/experiencer action/predicate

That schema initially mis-assigns “attacked” as an adjective, part of a predicate phrase. That analysis, while syntactically incorrect, is sufficient to access a form of semantic information – modeled on the semantic interpretation schema for lexical passive adjectives. Thus, an initial comprehension of the sentence can be based on a syntactic misanalysis, which is eventually corrected by accessing the correct derivation. This sequence also explains the fact that the evidence for the trace appears only after a short time has passed.

The psycholinguistic experimental literature of the last two decades is rife with controversy over how quickly and effectively statistically reliable information is assigned during comprehension. Much of this controversy has been presented under the rubric of proving or disproving that connectionist associative models can account for language behavior without recourse to linguistic derivational rules. While not a lot of light has come out of this controversy, it has documented that comprehenders are indeed sensitive to a wide range of statistically grounded information early in their comprehension. At the same time, experiments like the preceding also demonstrate that derivational structures are assigned as part of the comprehension process. Thus, the ‘inelegance’ of the analysis by synthesis model in postulating two kinds of overlapping computational operations, captures an evident fact that this is how people do it.

Aside from time consuming and often inconclusive experimental investigations, this model explains a number of simple and well known facts. Consider the following examples.

14a) The horse raced past the barn fell
14b) More people have visited Paris than I have

Our intuitions about each of these cases exemplifies a different aspect of the analysis by synthesis model. The first reflects the power of the canonical form strategy in English, which initially treats the first six words as a separate sentence (Bever, 1970). This sentence is often judged ungrammatical by native speakers until they see some parallel sentences of the same formal structure or related to it:

15a) The horse ridden past the barn fell
15b) The horse that was raced past the barn fell
15c) The horse racing past the barn fell

The example is pernicious in part because of the canonical form constraint, but also because recovering from the mis-analysis is itself complex: the correct analysis in fact includes the proposition that “the horse raced” (i.e., was caused to race). Thus, as the comprehender re-works the initial mis-parse, the correct analysis reinforces the incorrect surface analysis on which ‘the horse’ is taken to be the subject of the embedded verb. This seduces the comprehender back into the mis-parse.

The second example (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 in fact it does not have a correct syntactic analysis. The initial perceptual organization assigns it a schema based on a general comparative frame of two canonical sentence forms – ‘more X than Y’, reinforced by the apparent parallel structure in X and Y (….have gone to Paris….I have).

On the analysis by synthesis model, this superficial initial analysis gains entry to the derivational parse system, which then ultimately blocks any coherent interpretation.

I do not expect to have convinced the reader of our model via such simplified examples alone: in our book, we organize a range of experimental and neurological facts in support of the general idea that an early stage of comprehension rests on frequent statistically-valid patterns, followed by a more structurally complete assignment of a syntactic derivation.

An important consequence of the model for linguistics is that it requires certain universal features of actual languages in order to work. Most important is the otherwise surprising fact that actual languages have a characteristic set of statistically grounded structural patterns at each level of representation. It further requires that complex constructions with intricate derivations 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. In the next sections, I develop the implications of this for language learning and linguistic universals and relate it to cognitive science in general.

2.2 Constraints based on acquisition

For centuries, the two ideas about comprehension mentioned above (1a,b, repeated below), have alternatively dominated the entire Science of the Mind:

> 16a) Everything (real) is based on habits, prototypes, associations
> 16b) Everything (important) is based on rules, categories, computations

The repeated scientific mistake has been to make 16a or 16b, the only principle and to force it to account for everything. This has lead to alternating prescriptive rejection of such superficially obvious facts as:

> 17a) Languages involve categorical operations
> 17b) Languages are learned

That is, attempts by each school to account for everything mental have lead to some correspondingly stark mottos:

> 18a) Language cannot be learned by any "general" learning process
> 18b) Language cannot involve "ungrounded" symbolic computations

Something has to be wrong here. Clearly, we need a theory that both explains what language is, and how individual languages are learned by individual children. Any theory that requires language to be non-symbolic is wrong: any theory that requires language to be "unlearned" is wrong. What we need to do is develop a learning theory that could work, and then see if it does work that way.

Today's theory, popular amongst structural linguists is parameter setting – the child throws switches on universal language dimensions this way and that, based on impoverished evidence. A parameter setting model sets constraints on what such an acquisition theory must include: but it has almost the entire structure of attested and non-attested languages built in or autonomously constructable at the start (a tabula plena) (Recent attempts to explore how parameter setting might work more fully, still include prior knowledge of the set of grammars to try out (Yang, 2002), or even a scaffolding structure of proto-grammars designed to maximize efficient convergence on the correct grammar (Fodor and Sakes, 2004). While the brute force of this model may overcome the "poverty of the stimulus", in itself, it explains little about how the learning process unfolds in each individual, about interactions with emergent cognition, about the role of individual motivations and introspections. At the very least, we need a theory of acquisition "performance" to understand the individual mechanics and dynamics of setting parameters.

The analysis by synthesis model we developed for comprehension, can embrace parameter setting constraints, while also explaining other constraints on grammars in the framework of a general learning theory. We made some preliminary suggestions about this in our book, and I will expand on them a bit here. Basically, the idea is that the child alternates (logically, not always temporally) between developing statistically grounded syntactic/semantic patterns and providing structural derivations for sentences which fit those patterns. Indeed, the patterns that the child develops based on statistics in what s/he hears are just those patterns which adults rely on for comprehension; the derivations which the child tries out to compute the patterns it has acquired must converge on the same derivations as used by adults. There is a range of data suggesting that this model describes what individual children do during language acquisition. I just briefly sketch some salient facts here.

First, perceptual and production strategies emerge from experience: young children compute initial stages of syntax based on initial structural assignments independent of structural statistics or semantic probability and then acquire statistical patterns.

There is considerable evidence that children under age 2 already have a basic grasp of the notion of agent, and in English have some sensitivity to word order. (Golinkoff, (1975); Hirsch-Pasek and Gollinkoff, (1996)). We can show how this early capacity develops by systematically looking at the development of comprehension patterns in which they use puppets to act out short sentences. Two and four year olds show the following pattern of correctly acting out sentences (this is a recent replication of a study by me, Virginia Valian and Jacques Mehler, first reported in Bever, (1970)):

	Age 2	Age 4
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%
The dog ate the cookie	92%	96%
The cookie ate the dog	73%	45%
The cookie got eaten by the dog	55%	85%

Table 1. Percentage correct acting out interpretations of simple sentences

The early capacity of the two year old suggests an available schema of the form NV=Agent (action). Most striking is the fact that the object cleft construction is correctly interpreted, while semantic constraints have a relatively small effect on the two year old. This is consistent with the idea that the child has acquired a local categorical sequence syntactic strategy, but no general dependence on semantic information. The systematic performance at age 4 suggests that the child has now developed the more general sequence strategy, based on the canonical form of English mentioned above, on which the sequence initial noun is generally taken to be the agent (see (11) above):

19) #NV(N) = Agent, action (patient).

At the same time, the child now shows much more sensitivity to semantic constraints. In other words, the four year old has acquired some statistical patterns for comprehension not yet available to the two year old.

Another important fact is that children know about the difference between how they talk and how they should talk grammatically. There are numerous anecdotes reporting this awareness. For example, after one three year old child was repeatedly teased by a (linguist) father for using the incorrect weak past tense for go (i.e., “goed”) he finally said:

20) Daddy, I talk that way, you don’t.

This simple rejoinder shows that the child was aware of the distinction between the correct sentence form and his own dependence on forming the past tense with the regular ending. Awareness of this kind is consistent with the view here that the child develops statistical patterns as part of the overall acquisition process, of which s/he himself can be aware.

I also note that the model comports well with recent research showing that the infant is indeed a good extractor of certain kinds of patterns, and that Motherese actually has many statistically grounded properties which lead towards (but not all the way TO) correct syntactic analyses. (e.g., Curtin et al, 2003 (for segmentation); Golinkoff et all, (2005); Mintz, (2002, 2003, 2006); Redington et al, (1998); Brent, (1996); Cartwright and Brent, (1997); Gerken, (1996)). At the same time, there is now considerable research showing that infants are quite good at statistical inference from the presentation of serial strings with various kinds of structure. (Saffran et al, (1996); Saffran, (2001, 2003); Gomez and Gerken, (1997); Marcus et al, (1999); older children also show statistical sensitivity in developing grammatical and lexical ability (Bates and MacWhinney, (1987); Gillette et al, (1999); Moerk (2000), Yang (2006), Naigles and Hoff-Ginsburg (1998)).

The idea that the child acquires knowledge of syntax by way of compiling statistical generalizations and then analyzing them with its available syntactic capacities reflects a claim about how learning works in general that has been proposed in various forms for many years. 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. A different way of expressing this is in the classic work by Karmilov-Smith and Inhelder (1973) – 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 – 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 this expresses itself for example as the superficial identity of passives and active constructions: the resolution of the problem is to find a derivational structure for the problem that shows how actives and passives are both differentiated and related derivationally. (In this sense, Piaget's attempts to explain language acquisition were well directed – albeit incomplete). Hence, not only is language learning hereby interpreted in the context of a general learning device, it is also a special instance of a general problem solver. Language remains special because of its unique characteristics and role in human life: but it is no longer "special" because it is "unlearned".

The model also resolves some classical puzzles about acquisition. Notable is the problem of how children understand sentences before they have a grammatical analysis for them? (Valian, 1999). The idea that the child maintains a list of grammatically unresolved sentences is unsatisfactory because any given list is heterogenous unless it is given some kind of prior ordering and structure. The AxS model suggests that they rely on statistical patterns and occasional false analyses to generate an internal bank of meaning/form pairs which present a constant puzzle for coherent derivational analysis.

The example-generating role of such internalized patterns cannot be overemphasized. To some extent it mitigates the "poverty of the stimulus" the fact that the child receives sporadic, errorful and limited input to work with. It allows the child to generate new exemplars of acquired patterns, thereby expanding its internal data bank of slightly different meaning/form pairs to be analyzed syntactically. This partially resolves, or at least clarifies the problem of how children access positive and negative feedback as guides to their emerging syntactic abilities, even if they treat each sentence initially as a unique item. 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. (Choinard and Clark, (2003); Dale and Christiansen, (2004); Golinkoff et al, (2005); Lieven, (1994); Moerk, (2000); Morgan et al, (1995); Saxton, (1997, 2000); Valian, (1988, 1999))

Other phenomena of acquisition fall out of this description. The most often ignored fact exemplified by the anecdote above, is that children are aware of the difference between how they understand/say certain meanings and how they should say them. How can this be? It is actually a different expression of the puzzle of how children can understand sentences for which they do not have a complete grammatical analysis: in this case, the child can show that s/he is explicitly aware of the distinction – making overt the fact that s/he has an analysis available, while recognizing that it is not correct.

Note that there is no comfort here for empiricist associationists or any other model which attempts to show how computational structures are causally “grounded” or internalized from explicit patterns. The model assumes and requires that the child have the computational equipment to represent statistical patterns with some form of structure and to try out computational derivations that ‘explain’ how one gets from the form to the meaning. However, the explicit role of statistical generalizations as a dynamic factor in the process of discovering structural rules may also lay the groundwork for a solution to an outstanding problem in parameter setting theory – what in fact, are the appropriate and sufficient data to induce a child to set a syntactic parameter? (Pinker, (1984); Lightfoot, (1991); Fodor, (1998); Fodor and Sakas, (2004), Yang, 2002). A frequent discussion has involved the “subset principle”, the idea that the parameters are set to a default, which can only be undone by exposure to a particular example that triggers the exceptional value of the parameter (Berwick, (1985); Pinker, (1984)). This depends on languages being organized so that actual sentences exhibit the default values of a parameter, with the exceptional value being a “subset” of the default. The model of learning which includes the development of statistical patterns can explain this kind of apparent parameter setting without assuming that the default is specified in the infant’s mind ahead of time. It is typically the case that the default value of the parameter also describes the more frequent kind of syntactic construction. Exposure to the more frequent constructions (sentences *with* apparent subjects) and the corresponding incorporation of a template lays the statistical groundwork for recognition of sentences with the default parameter setting.

This way of thinking about how parameters are set during acquisition, may reduce the requirement that all parameters are innately specified with a default value actually included. That in turn, leads to a dynamic interpretation of language learning, with the formation of patterns, and a subsequent (logically) assignment of structural analyses of the patterns, and of their exceptions. Again, the reader may come away with the false impression that this view is simply one on which statistical features and patterns explain what is learned (as in, e.g., Bates and MacWhinney, (1987), and their later writings). On our theory, this is only one quarter of the story: the other three quarters are: 1) the (apparently automatic) mental isolation of the relevant dimensions of the parameters; 2) the emergence of structural processes which provide integrated derivations for sentences exhibiting the learnable patterns: 3) the availability of the analysis by synthesis model which integrates the statistical templates and derivational processes dynamically.

It is also useful to note that this structure resolves some of the limitations of each of the major approaches taken alone: access to statistical patterns mitigates the limitations of the poverty of the stimulus input: access to parametric dimensions and derivational processes partially defines the domains over which statistical generalizations can be formed.

The analysis by synthesis model is of course, also only a model. Its major failing for many is that just as in its application to sentence comprehension, it is inelegant – a kluge that can embrace many other theories in a dynamic framework. I too would prefer a more elegant system. But the behavioral facts do not allow that luxury, and no one, including the most devoted evolutionists have argued that all evolutionary developments move towards the most elegant and simple organizations. Most important for cognitive science in general is the fact that the model describes a special case of more general processes. This does not mean that language is not “special”, drawing on numerous special capacities unique to it. But at least, we see a way to recover the idea that language is learned in a dynamic manner, typical of all human learning, creativity and aesthetic judgment.

2.2.1. Implications for linguistic universals – the Canonical Form Constraint

We now turn to the question of how acquisition processes might constrain languages to exhibit certain kinds of universal properties. There have been various approaches to the question of how acquisition constrains linguistic structures. Most ambitious have been attempts to show that formal learnability of derivational relations constrain those relations to be of certain types (e.g., Osherson and Weinstein, 1982; Wexler and Culicover, 1980). The enduring results of these investigations is a set of abstract constraints on possible kinds of derivational processes that guarantee recovery of inner from outer forms of sentences. Stabler (this volume) and Steedman (this volume) also note that the interrelation of semantic structures and learning syntax may account for certain universals. These investigations propose boundary conditions on the architecture of grammars. But they tell us little about the dynamics of actual acquisition processes. More recently, Gleitman et al (2005) and Papafragou, et al.(in press) have outlined how cognitive and informational constraints on conceptual learning interact with syntactic structures to shape lexical structures, and through them, certain aspects of syntactic structures.

The analysis by synthesis model of language acquisition requires that actual attested languages have a number of properties not explained by linearity constraints, nor by the usual array of computational linguistic universals or parameters. The most significant involve the existence of levels of representation and their interrelation. The model of requires two features to start it off: it must have access to an initial syntactic vocabulary to characterize input sequences in some formal language; the input to the child must exhibit a standard statistically dominant form – without this “canonical form constraint” (CFC) the language learning device cannot develop statistically valid generalizations. The CFC has at least two consequences for attested languages: mapping systems between levels of representation “conspire” to make sure there is a canonical form at each level of representation; mappings between forms of representation are unidirectional, from the more inner/abstract to the more superficial;

The notion of derivational conspiracies is not novel, be it in syntax or phonology (cf. Ross, 1972;1973). In the case of English, the vast majority of sentences and clauses have a canonical form in which there appears to be a subject preceding a verb:

19a) The boy hit the ball
19b) The ball was hit by the boy
19c) It is the boy who hit the ball
19d) The boy was happy
19e) The boy seemed happy
19f) The boy was easy to see
19g) It was easy to see the boy
19h) Who saw the boy?
19i) Who did the boy see?
19j) Visiting relatives can be a nuisance.

The coincidence of the surface forms reflects a constraint on derivations such that they almost all end up with similar phrase structure and surface case relations. This is despite the fact that the architecture of many grammatical theories – including all variants of generative grammar - would allow languages in which each underlying form is expressed in a uniquely distinct surface phrase organization and sequence. On our interpretation, such computationally possible languages will not be learned because they make it hard for the language learning child to develop an early statistically based pattern that it can internalize and manipulate for further stages of acquisition.

There are numerous levels of representation mediating between the propositional form of sentences and the linear string. Syllabic, morphological, declension/conjugational, phrase, sentence. Attested languages have a canonical form at each level – canonical syllable, morphology, declension, conjugation, phrase (right vs. left branching), sentence syntax. The canonical form can constrain the computational relations between levels to ensure its maintenance (conspiracies).

If there is a canonical form at each level, then for effability, it must embrace more than one derivational relation to input levels – this means that the derivation from one level to another will generally not be completely discoverable via operationalist principles, based on surface forms. Rather some form of problem solving, via hypothesis generation and testing is required. This explains an otherwise puzzling fact about languages – why is each level distinct from the others, not subject to operational discovery procedures. Languages with transparent subset relations between levels could have the same expressive power as existing ones: in that case, each level of representation would offer direct evidence of its relation to the derivationally prior level. If every derivation had such a distinct surface form, it would fulfill the role of syntax as mapping each hierarchical propositional structure onto a linear sequence. But then there would not be a canonical form, hence, the language would not be learnable.

Accordingly, on this view, actually attested languages have mapping relations between levels which are unidirectional because they are many-to-one from inner to outer forms. Put in the terms of the modern 'biolinguistic metaphor', syntax defines operationally opaque derivations to relate meaning to sound. The existence of derivations is a structural universal of the language faculty in the narrow sense – but the opacity of the derivations, and hence their directionality, is a universal of the language faculty in the broad sense – a function of what makes languages learnable by a general learning process. (Hauser et al, (2002)).

2.2.2 Special implication for EPP – a non-syntactic principle after all?

The preceding discussion offers an explanation for why sentence forms tend to converge on a common superficial form. This may offer an explanation for one of the more problematic syntactic constraints, which is still an anomaly within the minimalist framework. This is the so called "extended projection principle" (EPP). This principle was first proposed to account for the appearance of subject-like phrases in sentences, so called expletives - basically a principle that all sentences have to have (surface) subjects. For example consider the following sentences:

20a) John seemed foolish
20b) Foolish is what John seemed (to be)
20c) What John seemed (to be) was foolish

Each of these has a constituent in the subject position, but with a different semantic role. This suggests that the subject position must always have an overt phrase. In some cases, the subject position is filled with a word that has no semantic role in the sentence ("it"), further demonstrating the force of the EPP.

20d) It seemed that John was foolish

The EPP has been initially proposed as a fixed syntactic universal, part of the set of syntactic constraints which all languages must respect. While more or less correct for English, the EPP has been further studied and a number of troubling facts have emerged (see discussions in e.g., Lasnik, (2001), Epstein and Seeley, (2002), and papers in Svenonius, (2002)):

- it may not be universal (e.g. Irish as analyzed by McCloskey, (1996, 2001)
- It can have a variety of expressions in different languages, e.g., in a standard relation to focus as opposed to subject, in intonation patterns and so on.
- It generally corresponds to the statistically dominant form in each language
- It has not found a formal derivation within syntactic theories or the minimalist program – that is, it simply must be stipulated as a "syntactic" constraint on derivations

The conclusion generally appears that the EPP may be a "configurational" constraint on derivations – it requires that sentences all conform to some basic surface pattern. Epstein and Seeley (2002, p. 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 that has to be stipulated in the minimalist framework, and is one that at the same time violates some of that framework's basic structural principles and simplicity.

We can now treat the EPP as a Ui in the manner discussed above. 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, basically an expression of the canonical form constraint – sentences have to "sound" like they are sentences of the language to afford the child a statistical entree into acquiring the language.

How can we decide between these two explanations? We can apply the same logic as Miller and Chomsky applied to restrictions on center embedding. Namely, the EPP adds a stipulated constraint to grammars, which would otherwise be formally simpler. Second the EPP appears to be a heterogenous constraint, with different kinds of expressions in different languages, not always strictly syntactic. Third, the canonical form constraint is independently attested and motivated: it explains statistical properties of language, stages of acquisition and significant facts about adult language processing. Thus, we can conclude that the phenomena that motivated the EPP are actually expressions of the Canonical Form Constraint (CFC). That is, languages could violate the EPP like phenomena so far as structural potential is concerned, but in so doing would violate the CFC and therefore not be learnable.

Syntacticians may object that this line of reasoning can appear to be circular. That is, many specific syntactic processes in individual languages appear to be explained by the EPP, governing not just the acceptability of sentences, but indeed their grammaticality. For example, in the examples at the beginning of this paper on heavy phrase shift, it is clear that some phonological phrase must appear in surface subject position to maintain grammaticality of the sentences with displaced phrasal-complements. Thus, the EPP constraint does not merely exert ‘stylistic’ preferences on sentence constructions, it (can appear to) dictate syntactic requirements on derivations involving movement or other constraints.

What is at issue is the source of the constraint that results in processes that appear to conform to the EPP. On the view presented here, the child learns sentence constructions that conform to the canonical form constraint, and tends not to learn constructions that do not. But on our interpretation, the notion of “learn” can be glossed as “discovers derivations for, using his or her available repertoire of structural devices….”. This allows for the discovery by the child of a wide range of structural derivational processes that maintain the canonical form constraint. Thus, we can accept that for individual languages there are specific derivational processes that conform descriptively to the EPP. But we argue that the EPP itself is merely a descriptive generalization which reflects acquisition constraints as its true cause. Note that the canonical form constraint can be different in different languages. For example, in inflected languages, the canonical form may involve particular suffixes, and not constituent order. Slobin and Bever (1982) found that the early canonical form that children arrive at in Serbo-Croation or Turkish conform to canonical properties in those languages. In Serbo-croation this involves a mixture of order and inflectional properties: in Turkish, with largely free constituent order, the canonical form appears to be two arguments and a verb, in which one of the arguments has a clear object suffix.

2.3 Constraints from neurological substrates

Consider now a speculative case study of computational constraints on the Faculty of Language based on neurological considerations. While it has some logical force and a small amount of supporting data, my main purpose is to flesh out the range of ways we can consider where universals of language may come from. The claims are strong and require much more empirical demonstration than there is space for here.

One of the enduring properties of syntactic operations has been an upward moving cycle. That is, in every generative model, operations iterate from the most to the least embedded part of a sentence. This has often appeared as a constraint on movement to be “upward”, from a more to a less embedded constituent. Attempts to define top-down computation (Richards, (2001, 2002);

Phillips, (1995)) characteristically involve ‘look-ahead’ templates or feature-codings, which simulate “upward” constraints, and involve computational demands that increase geometrically with string length.

The priority of upward movement has recently found an explanation (Chomsky, (2004, 2005)). Merge as a successive operation that forms trees, has two possible expressions at each iteration: “internal” merge results in new membership of X and prior Y within the same constituent; “external” merge results in an X outside of a prior X and Y, in effect a “copy” of the now more embedded X. Since the latter is by definition higher in the emerging hierarchy, upward “movement” is rendered if movement is interpreted as “copying by a higher constituent and deletion of the lower identical one”). Thus, “movement” is now represented as copying, with default constraints on expressing only the least embedded copy of a constituent. That is, upward “movement” is a natural result of recursion, which itself has been proposed as the essential computational operation critical to language evolution. (Hauser et al, 2002).

One possible route for this evolution is that the capacity for recursive merge appeared as a single (set of) mutations, which immediately lead to more powerful minds. But another possibility is that recursive merge was “exapted” from other modalities with a long prior evolutionary background. An obvious candidate is from vision, because the brain areas devoted to language in humans are largely homologous to areas evolved for vision in our close biological relatives. The question arises: what computational devices – if any - might have evolved to solve visual problems which could later transfer as the neurological basis for recursion in language?

Consider first some basic facts about motion perception. If a 0 moves from the left below, to the right, under temporal constraints consistent with motion (say a fifth of a second), what we see is a dot in motion arriving at the right side:

0……………………………………………………………………..…………….>**0**

The motion may appear continuous. This is the basis for the well known “phi phenomenon”, in which motion is deduced from discrete presentations at a distance from each other. (Wertheimer, (1912); Kolers, (1972); Rock, (1983); Breitmeyer, (1984)). The organization of motion from one point to another is not only deductive, later positions can ‘absorb’ earlier ones. For example, in well known cases of “metacontrast” the initial representation can become invisible as a separate entity (Ramachandran and Gregory (1991); Shepard and Judd, (1976)). In such cases of metacontrast, we perceive the final location of the 0, and the sensation that it moved to get there, but the representation of the original 0 is itself subsumed under the final representation: the simplest computational representation of this classic phenomenon in fact involves a form of merge, in which the final representation is organized as a function of its precursor(s).

0…………………………………………………………………………….………..>**0** (0)

This analysis, along with the assumption that only the highest representation in the structure is perceived, can explain the fact that the sequence of prior representations can be seen as object-in-motion, arriving at the final point. Each of the “snapshots” that the eye captures is subsumed by the next, resulting in recursive iteration of the entire motion sequence.

This process of the perception of integrated motion is a potential computational foundation for the exaptation of recursive merge into language functions. Of course, we cannot test this directly, but we can consider some empirical predictions of our computational-recursive interpretation of simple motion. In particular, the metaphor of “movement” as “recursive copying” from less to more embedded structures, suggests the hypothesis that real visual movement is easier to perceive from more to less embedded parts of a visual scene.

We have run a series of investigations of this (Lachter, Weidenbacher and Bever, in preparation), showing that indeed movement is easier to perceive from an embedded part of a scene than into the embedded part. One way we test this is by studying the effectiveness of perceived motion between a small square and a larger square that contains the smaller square (there are various appropriate controls for attention and other factors). The results confirm the general hypothesis.

Suppose we take this to support the idea that “upward” movement is computationally pre-figured in the visual system, so that it provided the computational basis for exaptation into linguistic computations. In no way does this explain away the faculty of language as the result of “general” cognitive or visual processes. Rather, it would be an example of appropriation of structures and processes evolved for one purpose, by another function. This is not an unusual idea in evolutionary theory.

But we owe ourselves an argument as to why “upward” movement took on priority in the visual system in general – many computational processes other than hierarchical merge might have been adaptive for motion perception. A full discussion of this involves consideration of the visual perception of biological movement. There is emerging evidence for special visual mechanisms that recognize biological motion as a function of hierarchically structured input, which requires upward integration of sub-vectors of motion into an overall vector (Pomplun et al 2002; Tsotsos et al, 2005). This could be the basis for a generalization of hierarchical processing. A simpler, and more classically Darwinian argument is that the perception of objects that appear from behind other objects is more salient and more important in survival of any species, than perception of objects that move behind another object. Things jumping out at you are more likely to be efficiently nutritious or effectively dangerous than things jumping away and hiding.

Of course, our results and speculation are only part of a much larger empirical and theoretical story that we must support before we can come to the strong conclusion that upward movement (or its correspondent, downward control) in language structures is the result of corresponding constraints developed for vision. I mention it here as an example of the kind of investigation that can be pursued as part of delimiting and understanding the boundary conditions on syntactic operations and their evolution.

3. Conclusion.

3.1 Minimalist Behaviorism?

Language may have design features that are responsive to its role as an interface between thought and behavior. It may also rest on specific genetically structured filters or computational capacities, which may explain other universal properties. But it is also learned and performed by individual children, with individual motives and learning principles, all of which may be more general than language itself. I have explored a few of such cases in outline form, not to convince you that we know the answers, but to outline the potential range of constraints on universals that extend well beyond the idealized function of language itself.

In a sense, this represents a return to a classic notion of the "ideal speaker/hearer/learner" (cf. Chomsky, 1965). Today's minimalist syntax is a framework for the exploration of language as the optimal solution for the interface between thought and speech – the idea in part is to define the minimal assumptions about the universal faculty of language just sufficient to account for its structure. The goal is to explain language as it is, not humans' knowledge of it. The ideas raised in this paper can be taken as examples of the complementary study – what is minimally required to account for the idealized knowledge and use of language.

3.2 Recommended readings.

The topics touched on in this paper include among others, the history of syntax, current minimalism, current theories of language behavior, theories of language acquisition, the Extended Projection Principle as well as general theories of learning and vision. That is a lot of territory to cover in any limited set of readings. The reader is invited to select from the bibliography, those that intrigue him or her. The main technical theme, however, is the integration of statistical and structural kinds of knowledge, in behaviour and learning. I have attempted to sketch a theory of learning which is more general than just for language, but which has implications for language learning and structure. The main goal is to capture the two enduring generalizations about behavior, that it depends on habits, and that it depends on symbolic computations.

References

Bates E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp.157-193). Hillsdale, NJ: Erlbaum.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 279-362). New York: Wiley.

Bever, T.G., & Langendoen, T. (1971). A dynamic model of the evolution of language. *Linguistic Inquiry,* 2.

Bever, T.G., Carroll, J.M., & Hurtig, R. (1976). Analogy or ungrammatical sequences that are utterable and comprehensible are the origins of new grammars in language acquisition and linguistic evolution. In T.G. Bever, J.J. Katz & D.T. Langendoen (Eds.), *An integrated theory of linguistic ability* (pp. 149-182). New York: T.Y. Crowell Press

Bever, T., & Sanz, M. (1997). Empty categories access their antecedents during comprehension. *Linguistic Inquiry, 28,* 68-91.

Boeckx, C. (2006) *Linguistic Minimalism: Origins, Concepts, methods, and Aim*. Oxford University Press

Borer, H. (1994). *The Projection of Arguments*. *UMOP14*, University of Arizona.

Breitmeyer, B. G. Visual Masking (Oxford Univ. Press, New York, 1984).

Brent, M. R. (1996). *Computational approaches to language acquisition*. Cambridge, MA: MIT Press.

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63*(2), 121-170.

Chomsky, N. (1995). *The minimalist program.* Cambridge, MA: MIT Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In Belletti, A. (Ed.), *Structures and beyond – The cartography of syntactic structures, Vol 3* (pp. 104-131), Oxford, UK: Oxford University Press.

Chomsky, N. (2005). Three Factors in the Design of language. *Linguistic Inquiry, 36,* 1-22.

Chomsky, N. & Miller, G. (1963). Introduction to the formal analysis of natural languages. In D. Luce, R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 269-321). New York: Wiley.

Chouinard, M.M. & Clark, E.V. (2003). Adult reformulations of child errors as negative

Dale, R. & Christiansen, M.H. (2004). Active and passive statistical learning: Exploring the role of feedback in artificial grammar learning and language. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 262- 267). Mahwah, NJ: Lawrence Erlbaum.

Epstein, S., & Seely, D. (2002). *Derivation and explanation in the minimalist program.* Malden, MA: Blackwell Publishing.

Fodor, J. (1998). Unambiguous triggers. *Linguistic Inquiry, 29,* 1-36.

Fodor, J. D. (2001). Setting syntactic parameters. In M. Baltin and C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 730-738). Oxford, UK: Blackwell Publishers.

Fodor, J. D., & Sakas, W. G. (2004). Evaluating models of parameter setting. In A. Brugos, L. Micciulla & C. E. Smith (Eds.), *BUCLD 28: Proceedings of the 28th Annual Boston University Conference on Language Development* (pp.1-27).

Somerville, MA: Cascadilla Press.
Freidin, R., & Vergnaud, J. R. (2001). Exquisite connections: Some remarks on the evolution of linguistic theory. *Lingua, 111,* 639-666.
Gerken, L. A. (1996). Phonological and distributional cues to syntax acquisition. In J. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition (*pp. 411-426). Mahwah, NJ: Erlbaum.
Gillette, J., Gleitman, L., Gleitman, H., & Lederer, A. (1999). Human simulation of vocabulary learning. *Cognition, 73,* 35-176.
Gleitman, L., Cassidy, K., Nappa, R., Papafragou A., Trueswell, John. (2005). Hard Words. *Language Learning and Development 1*(1), (23-64).
Golinkoff, R. M. (1975). Semantic development in infants: The concepts of agent and recipient. *Merrill-Palmer Quarterly, 21*, 181-193.
Golinkoff, R., Pence, K., Hirsh-Pasek, K., & Brand, R. (2005). When actions can't speak for themselves: Infant-directed speech and action may influence verb learning. In D. Massaro, R. Calfee, J. Sabatini, & T. Trabasso (Eds.), *From orthography to pedagogy: Essays in honor of Richard L. Venezky* (pp. 63-79). Mahwah, NJ: Erlbaum.
Gómez, R.L., & Gerken, L. A. (1999). Artificial grammar learning in one-year-olds: Leads to specific and abstract knowledge. *Cognition, 70,* 109-135.
Halle, M. and Stevens, K.N. (1962) Speech Recognition: A model and a program for research. RLE Reports: Reprinted in J.A.Fodor and J.J. Katz, (Eds.) *The Structure of Language: Readings in the philosophy of language.* Englewood Cliffs, NJ; Prentice-Hall
Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569-1579.
Hirsh-Pasek, K., & Golinkoff, R. M. (1993). Skeletal supports for grammatical learning: What the infant brings to the language learning task. In C. K. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research, Vol 8* (pp. 299-338). Norwood, NJ: Ablex.
Hornstein, N., Nunes, J. Grohman, K. (2005). *Understanding Minimalism.* Cambridge, UK. Cambridge University Press.
Karmiloff-Smith, A., & Inhelder, B. (1973). If you want to get ahead, get a theory *Cognition, 3,* 195-212.
Kolers, P. A. (1972) Aspects of Motion Perception *Memory & Cognition, 30*(5), 678-686.
Lasnik, H. (2001). A note on the EPP. *Linguistic Inquiry*, *32,* 356-362.
Lasnik, H. (2003). *Minimalist investigations in linguistic theory.* New York: Routledge Ltd.
Lasnik, H. and Depiante, M. (2000) *Syntactic structures revisited.* Cambridge, MA, MIT press
Lieven, E. (1994). Crosslinguistic and crosscultural aspects of language addressed to children, In C. Galloway and C. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56-74). New York: Cambridge University Press.
Lightfoot, D. (1991). *How to set parameters*. Cambridge, MA: MIT Press.
Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science, 283,* 77-80.
Martin, R, & Uriagereka, J. (2000). Some possible foundations of the minimalist program. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays in honor of Howard Lasnik* (pp. 1-24). Cambridge, MA: MIT Press.

McCloskey, J. (1996). Subjects and subject positions in Irish. In R. Borsley and I. Roberts (Eds.), *The syntax of the Celtic languages* (pp. 241-283). Cambridge, UK: Cambridge University Press.

McCloskey, J. (2001). The distribution of subject properties in Irish. In W. Davies & S. Dubinsky (Eds.), *Objects and other subjects* (pp. 157-192). Dordrecht: Kluwer Academic Publishers.

McElree B., & Bever, T. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research, 18,* 21-35.

McGinnis, M. & Richards, N. (Eds.). (in press). *Proceedings of the EPP/Phase Workshop,* Cambridge, MA: MIT Working Papers in Linguistics.

Miller, G. (1962). Some psychological studies of grammar. *American Psychologist, 17,* 748-762.

Miller, G., Galanter, E. & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt.

Miller, G., & N. Chomsky (1963). Finitary models of language users. In R.D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. II* (pp. 419-492). New York: Wiley.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language.
grammatical categories in speech to young children. *Cognitive Science, 26*(4), 393-424.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90(1),* 91-117.

Mintz, T. H. (2006). Finding the verbs: Distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 31-63). New York: Oxford University Press.

Moerk, E. (2000). *The guided acquisition of first-language skills.* Westport, CT: Ablex.

Morgan, J. L., Bonamo, K. M., & Travis, L. L. (1995). Negative evidence on negative evidence. *Developmental Psychology, 31,* 180-197.

Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language, 25*(1), 95-120.

Osherson, D. and Weinstein, S. (1982). Criteria of Language Learning. *Information and control, 52*(2), 123-138.

Papafragou, A., Cassidy, K., Gleitman, L. (in press). When we think about *thinking*:

The acquisition of belief verbs. *Cognition*

Pinker, S. (1984). *Language learnability and language development.* Cambridge, Mass: Harvard University Press.

Pinker, S. and J. Mehler (1988). *Connections and symbols.* Cambridge, MA: MIT Press.

Phillips, C. (1995). Right association in parsing and grammar. In C. Schütze, J. Ganger & K. Broihier (Eds.), *Papers on language processing and acquisition. MITWPL #26,* (pp.37-93).

Phillips, C. (1996). *Order and structure*. Unpublished doctoral dissertation, Massachuchetts Institute of Technology, Cambridge.

Piatelli-Palmarini, M. and Uriagereka, J. (2004) The immune syntax: the evolution of the

language virus. In Lyle Jenkins, editor, Variation and Universals in Biolinguistics, pages 341--377. Oxford: Elsevier.
Piattelli-Palmarini, M., Uriagereka J., Salaburu, P., (Eds) (In preparation). *Of Minds and Language: The Basque Country Encounter with Noam Chomsky*. Oxford University Press
Pomplun, M. Liu, Y., Trujillo, J., Simine, E., Tsotsos, J. A neurally-inspired model for Detecting and Localizing simple motion patterns in image sequences (2002). Workshop on Dynamic Perception, Bochum, Germany, Nov 14-15.
Ramachandran, V. S. & Gregory, R. L. (1991) Perceptual filling in of artificially induced scotomas in human vision. Nature 350, 699−702
Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13*(2-3), 129-191.
Richards, N. (2002). Lowering and cyclicity: Attraction by X from Spec XP. In M. Hirotani (Ed.), *Proceedings of NELS, 32.* Amherst, MA: GLSA.
Richards, N. (2001). *Movement in language.* Oxford, UK: Oxford University Press.
Richards, N. (2003). Why there is an EPP. *Gengo Kenkyu, 123,* 221-256.
Rock, I. *The Logic of Perception* MIT Press, Cambridge, Massachusetts, 1983.
Ross, J. (1972). The category squish: Endstation Hauptwort. In Paul Perntau et al (Eds.), *Chicago Linguistic Society, 8,* 316-328.
Ross, J. (1973). A fake NP squish. In Bailey & Shuy (Eds.), *New ways of analyzing variation in English* (pp. 96-140). Washington: Georgetown University Press.
Ross, J. (1973). Nouniness. In O. Fujimura (Ed.). *Three dimensions of linguistic theory* (pp. 137-257). Tokyo: TEC Corporation.
Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12,* 110-114.
Saffran, J.R., Aslin, R.N. & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.
Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81(2),* 149-169.
Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language, 24,* 139-161.
Saxton, M. (2000). Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language, 20,* 221-252.
Shepard, R. N. & Judd, S. A. (1976) Perceptual illusion of rotation of three-dimensional objects. Science 191, 952−954|
Slobin, D.I., & Bever, T.G. (1982). Children use canonical sentence schemas: A crosslinguistic

study of word order and inflections. *Cognition*, 12, pp. 229-265.

Svenonius, P. (2002). *Subjects, expletives and the EPP.* New York: Oxford University Press.
Tsotsos, J., Liu, Y., Trujillo, J., Pomplun, M. Simine, E., Kunhao, Z., (2005) Attending to visual motion. *Computer vision and image understanding. (100)*(1-2). 3-40
Townsend, D., & Bever, T.G. (2001). *Sentence comprehension.* Cambridge: MIT Press.
Uriagereka, J. (1999). Multiple spell out. In Epstein, Samuel David & Norbert Hornstein (Eds.), *Working minimalism* (pp. 251-282). Cambridge, MA: MIT Press.

Valian, V. (1999). Input and language acquisition. In W. C. R. T. K. Bhatia (Ed.), *Handbook of child language acquisition* (pp. 497-530). New York: Academic Press.
Vergnaud, J. R. (1985). Dépendance et niveaux de représentation en syntaxe. *Lingvisticae Investigationes Supplementa*, *13*, 351-371.
Wertheimer, M. (1912) Experimentelle studien über das sehen von bewegung. Z. Psychol. 61, 161−265
Wertheimer, M. (1925). *Drei abhandlungen zur gestalttheorie.* Enke, DE: Erlangen.
Wertheimer, M. (1945). *Productive thinking. New York: Harper.*
Wexler, K., & Culicover, P. (1980). Formal principles of language acquisition. Cambridge, MA: MIT Press.
Yang, C. (2002). *Knowledge and learning in natural language.* New York: Oxford
Yang, C. (2006). *The infinite gift.* New York: Scribner.

AKNOWLEDGEMENTS
This paper has benefited from discussions with at least the following: Andrew Carnie, Noam Chomsky, Heidi Harley, David Medeiros, Steven Ott, Massimo Piatelli-Palmarini.