Bever, T.G., (2008) The canonical form constraint: language acquisition via a general theory of learning. In Guo et al, *Cross-linguistic approaches to the psychology of language.* (Oxford University Press) pp.475-492

23

The Canonical Form Constraint

Language Acquisition Via a General Theory of Learning

Thomas G. Bever

"Poppa, Will you Stop That!"...."Sure"

Dan I. Slobin (1978, p.52)

Background

In the early 1970s, Dan Slobin acquired substantial funding to test the development of language comprehension strategies in several languages – English, Italian, Serbo-croation, Turkish. This represented the intersection of our interests: his in developing a comprehensive theory of language acquisition: mine in the idea that early stages of acquisition depend on the creation of statistically supported behavioral strategies that short-cut syntactic knowledge. My laboratory's original finding was in English, based on a set of studies that had children act out simple sentences with puppets (a collaborative study with Jacques Mehler and Virginia Valian). Typical performance data showed that at age 2, children use a simple strategy that focuses primarily on the exact sequence Noun Phrase+Verb, interpreting that as Agent+Verb. By age 3-4, they rely both on a more elaborated analysis of word order and semantic meaning. 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 in fact the agent. By age 4 their performance depends on two strategies (1):

(1) a. NV(N) = Agent, predicate (patient)
b. Animate nouns are agents, inanimate nouns are patients.

(1a) 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 maintains 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). Typical data are summarized in Table 23.1:

The emergence of the two kinds of strategies accounts for the decrease in performance on semantically reversible sentences that violate the normal order of English (1a). 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 four also can override the word order strategy, leading to correct performance on irreversible passives *(the cookie got eaten by the dog)*.

Although I was an interloper in the language acquisition enterprise, Dan embraced the idea of early childhood language processing strategies, and immediately asked the following questions: are the strategies universal? If so, do they differ in importance for each lan-

TABLE 23.1 Percentage Correct Interpretations of Simple Sentences by Children (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%).

	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%

guage? It is important (and amusing) to remember that before Dan focused his major research program on language acquisition, he was a charter member of George Miller's original psycholinguistics platoon, devoted to demonstrating the 'psychological reality' of transformational grammar (in its 'syntactic structures' manifestation of the day): a typical result was that passive sentences are harder to process than actives, presumably due to the additional, 'passive' transformation (Miller, 1962). In that role, Dan showed that in fact passive sentences are not harder than actives if they are semantically irreversible, i.e., adults follow the childhood strategy (1b) above (Slobin, 1966). This finding of Dan's had played a role in prompting me to think about processing strategies as distinct from rule-based syntactic computations (Bever, 1970). My idea was that children learn these strategies, based on their 'cue validity' in child directed speech. I never went as far over the statistical cliff into an agrammatical abyss, as emerged in the later writings of Liz Bates and Brian MacWhinney, but the idea of separate-but-equal processing and rule systems became a constant theme in my language science program, even persisting today. Dan himself wrote about the processing issue in his foundational, short and brilliant introduction to psycholinguistics – based on the notes he wrote to help his students in the first post-transformational grammar course in modern psychology of language.

We all know that Dan Thinks Big. He was intrigued by the possibility that semantic and syntactic strategies are themselves learned, and continue to override syntactic processing, even in adulthood. So, Big Thinking in this case led to a month-long 'working' conference to develop the cross linguistic materials and experimental methods to be used in the study. The conference brought together several professors (me, him, Sue Ervin-Tripp, Eve Clark) and graduate students (including a surprise visit from Liz Bates). There was nothing really special about a working conference, except that it was for a full month and its location was set in the dramatic hills just above Dubrovnik. This was at a time when Yugoslavia was still behind the Plastic Curtain. Enough said.

The resulting project involved long and complicated management, which Dan and students slogged through; and this is what we found in the end (Slobin and Bever, 1982):

(2) a. Children acquire processing strategies adaptive to the statistical regularities in the structure of their own language. Thus, in English what develops is sensitivity to word order, in Turkish, sensitivity to patient-recipient inflectional markers, and in Italian and Serbo-croation, a mixture of the two kinds of linguistic signals.
b. There is *some* acquired tendency for initial nouns to be taken as agents when the stimuli violate general syntactic properties of the language.

Let me now turn to an important and lasting theoretical implication of these findings: the necessity of a canonical form to make languages learnable, and how that constraint filters computationally possible languages to actually attested languages.

Do Adults Use A Derivational And Recursive Grammar?

The basic puzzle of acquisition is set by the undeniable fact of 'the poverty of the stimulus', the fragmentary and often ungrammatical language, which children hear from their caretakers. No doubt there are many subtle cues, even subtle feedback when the child makes a linguistic error, but only a highly pre-tuned and hypothesis forming system could make use of them to develop knowledge of syntactic rules and architectures. Of course, the first question is, *do speakers actually develop a generative grammar, or only a simulation of it?* This is the original question about psychological reality of grammar – if adult speakers do not actually use grammatical knowledge as part of their language behavior, we do not have to worry about how children might learn it. Fifty years of linguistic and psycholinguistic research and intuition have established the following facts about adult language behavior (3):

(3) a. Syntactic processes are indeed 'psychologically real', that is, sentence level syntax plays a causal role in representations used in language comprehension and production.
b. Syntactic processes are recursive and derivational: they range over entire sentences in a 'vertical' fashion (as opposed to serial) with successive re-applications of processes to their output. These properties have been true of every 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 certainly does not wait until the end of each sentence: production of a sentence does not wait until it is entirely formulated.

These three observations set a conundrum (4):

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

Standard processing models deal with one or the other side of the conundrum. Models, which assign syntactic derivations as an initial stage of comprehension and production account for (4a) but not so well for (4b). Models, which assign meaning via pattern recognition and completion, account for (4b) but not for (4a) at all well.

Recently, Dave Townsend and I rehabilitated the classic comprehension model of analysis by synthesis to provide at least a logical solution to the conundrum. (Halle and Stevens (1962), Townsend and Bever, 2001). On this view, people understand everything twice, once by way of initial perceptual strategies such as (1a) and (1b); once by the assignment of a syntactic derivation. The analysis by synthesis architecture lays out the two processes as proceeding in time. First the perceptual patterns assign likely interpretations to sentences, using something like a pattern completion system in which initial parts of a serial string automatically trigger a complete template. Second, the initially assigned potential meaning triggers (and constrains) a syntactic derivation. Another way of stating this is that 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.*

In our book, we adduced a wide range of intuitive and experimental evidence supporting this view of comprehension, reviewing more than 500 references. Some examples include the following:

(5)*a. The NVN template in English is extremely powerful and ineluctable. It explains the unacceptability of reduced relative sentences (e.g., 'the horse raced past the barn fell'; compare with 'the horse ridden past the barn fell').

* NVN = Noun Verb Noun; NP=Noun phrase

b. We can recognize sequences as forming sentences using the initial perceptual patterns, only to subsequently realize that they are nonsense. (e.g., 'More people went to Russia last year than I have')
c. Processing is unevenly distributed across sentences, with the greatest load just at the end of major units. (First demonstrated in Bever, 1970, later re-minted by Carpenter and Just, 1975 as 'end of clause wrap up').
d. Empty categories are assigned to sentences in the temporal order that they are assigned derivationally: WH-trace, NP-trace, PRO.
e. Full derivations that specify minor details of phrase structure are not available until a sentence is just completed.
f. Semantic information can become available before syntactic information in a sentence.
g. Discourse context processes can influence early semantic interpretation, but not syntactic assignments.
h. Broca's aphasics can discriminate the grammaticality of sentences they cannot understand.
i. Brain imaging studies show an early and late process, the first associated with assignment of surface patterns, the second with derivational syntactic assignments.

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 (Townsend and Bever, 2001). Here I focus on one case study, the comprehension of syntactic passives (see 6).

(6) a. Athens was attacked.
b. Athens was ruined.

Classically, the passive form of verbs can be differentiated into 'syntactic' *vs.* 'lexical' passives. Lexical passive participles (e.g. ruined) distribute in the same way as normal (stative) adjectives, motivating their categorization as lexically coded, stative-like adjective forms (see examples in 8). This motivates the decision that lexical passives are coded in the lexicon as separate lexical items. Thus the derivation of a sentence with a lexical passive participle (e.g. 8a), is like that of a corresponding adjectival sentence (8 d).

(7) a. *Athens was quite attacked.
b. *Athens looked attacked.
c. Athens was being attacked.

(8) a. Athens was quite ruined.
b. Athens looked ruined.
c. *?Athens was being ruined.
d. Athens was quite large.

Syntactic passive participles (e.g. 'attacked' in 7c) do not pattern as adjectives (cf. 7 a and 7b), but are analyzed as part of a true passive construction with a derivation. The corresponding surface forms from a derivation of syntactic passives in a theory that includes traces from movement looks schematically like that in (9a), in contrast with lexical passive constructions that do not include a trace.

(9) a. Athens was attacked. [t-Athens]
b. 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. That is, there is a classic kind of psychological demonstration of the 'reality' of syntactic passive trace. 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:

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

That schema initially misassigns '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 couched under the rubric of proving or disproving that connectionist associative models can account for language behavior without recourse to linguistic derivational rules. Often the researchers argue past each other or about different examples. While not a lot of light has come out of these efforts, they have 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 ones on trace 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.

(11) a. The horse raced past the barn fell.
b. More people have gone to Russia in the last decade than I have.

Our intuitions about each of these cases exemplify a different aspect of the analysis by synthesis model. The first (11a) 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:

(12) 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 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 above (11b, due to Mario Montalbetti), is the obverse of the first example. In the second 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 Russia….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 will develop the implications of this for language learning and linguistic universals and relate it to cognitive science in general.

Analysis by Synthesis in Language Acquisition

The model is inelegant in the sense that it solves the conundrum by fiat – sentence processing is both complex and fast because it is simultaneously handled by two systems, one fast and sometimes wrong, one slow but ultimately correct. Our achievement was not to create an elegant solution to the conundrum, but rather to show that humans solve it inelegantly. This model unites two historically competing observations in the cognitive sciences about the mind (13):

(13) a. Everything we do is based on habits.
b. Everything (important) we do is based on symbolic computations.

These two insights have alternately dominated scientific dogma about the mind for two centuries. The analysis by synthesis model is

an architecture that shows how the two insights might be integrated together in adult behavior. A corresponding model holds for the acquisition of complex behaviors, such as language. On that model, the child alternates (logically) between formulating statistical generalizations about the language, and formal derivational processes that account for those generalizations. The evidence from Slobin and Bever (1982) supports the first idea, while the child's ultimate mastery of the structure of language supports the second idea. In today's zeitgeist, many researchers are exploring the extent to which child-directed speech offers helpful statistical regularities that might guide the infant and child towards language, and also the extent to which infant pattern learning strategies might presage and facilitate language learning. Such recent research shows that the infant is indeed a good extractor of certain kinds of patterns, and that child-directed speech actually has many statistically grounded properties, which lead towards (but not all the way to) correct syntactic analyses. (e.g., Curtin et al, 2005, for segmentation, Golinkoff (1975), Golinkoff et al, 2005; Mintz, 2002, 2003, 2006; Redington and Chater, 1998, 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 drawing 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).

At the same time, there are various schemes proposed as to how a richly endowed innate structural scheme might arrive at a correct grammar with very impoverished input data—the dominant acquisition scheme within the framework of generative grammar is 'parameter setting': this is the idea that all languages can be described by a set of parameters (e.g., right vs. left branching, Np V Np vs. Np Np V, prodrop vs.no prodrop)—by hypothesis, the child has the set innately, each with a default setting: the child's task in learning his/her language is to learn which parameters in his/her language have a non-default setting (Lightfoot ,1991, Pinker, 1984, Fodor, 1998, 2001, Fodor and Sakas, 2004).

My hypothesis, and that of a few others who accept the idea that children in fact acquire generative grammar (e.g., Gilette et al.,1999, Gleitman 1990, Papafragou et al., in press) is that neither kind of scheme alone is adequate to the facts. On this view, acquisition is as inelegant as adult language behavior—it involves both formation of readily

available statistical generalizations and the availability of structures to rationalize violations of those generalizations. This view converges onto a traditional view of learning—hypothesis formation and rejection/refinement of the hypothesis based on further data (Karmiloff-Smith and Inhelder,1973).

This view requires that the child be presented with overwhelming statistical regularities in the language s/he hears, in order to facilitate the formation of statistical generalizations. This requirement explains several computationally eccentric facts about attested languages (14):

(14) a. Each language has a canonical surface form: In English this is essentially NP V (NP) (It can differ by language and is not always serial, as in Slobin and Bever, 1982).
b. Statistically, the canonical form has a dominant interpretation in semantic relations: In English this is NP V (NP) = Agent/experiencer, predicate, patient.
c. The canonical form 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 passives, discussed above in examples (6) – (9)).

None of these properties follows directly from the computational architecture of grammar. Yet they are characteristic of languages. In English, the first property (14a) 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. 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:

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

The notion of such conspiracies is not novel, be it in syntax or phonology (cf. Ross, 1972, 1973a, b). In traditional derivational terms there has to be some kind of constraint on derivations such that they almost all end up with the same general surface form. This is despite the computational fact that each underlying or logical form could be reflected in a unique surface 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 use for further stages of acquisition.

This facilitates the formation of a surface schema based on statistical dominance of the pattern. The second property above (14b) relates that schema to a standard semantic interpretation. The third property is particularly important if the child is to discover that there are actual derivations in which a given surface form can express different patterns of thematic relations. The fourth property contributes further to the child's ability to interpret sentence types for which it does not yet have a syntactic analysis. For example, raising (16a) or passive constructions such as (16c) can be interpreted via a schema based on a simple preterite construction (16b and 16d).

(16) a. Harry seemed/looked/happened-to-be happy.
b. Harry was happy.
c. Harry was hit (by Bill).
d. Harry was happy (about Bill).

The combination of these characteristics guarantee that the child can transcend the 'poverty of the stimulus' problem—the child can create and then analyze his/her own set of form/meaning pairs based on these generalizations. In addition, this solves an important problem for any learning scheme, namely the problem of how children understand sentences for which they do not yet have a correct syntactic analysis (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 Analysis by Synthesis 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, Valian,1999).

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. Slobin (1978) reported extended interviews with his child demonstrating a similar sensitivity: 'she rarely uses some of the [strong] verbs correctly in her own speech; yet she is clearly aware of the correct forms.' He reports the following dialogue at 4;7 (pps. 52–53).

Dan: Did Barbara read you that whole story?
. . .
Heida: Yeah
. .
and
.
mama this morning after breakfast, read ('red')
Dan: You don't know when she what?
Heida:
.
she readed the book
.
Dan: M-hm
Heida: That's the book she read. She read the whole, the whole book.
the whole book
. .
I don't know when she readed ('reeded')
.
Dan: That's the book she readed, huh?
Heida: Yeah
...
read! (annoyed)
Dan: Barbara readed you Babar?

Heida: Babar, yeah. You know cause you readed some of it too
...
she readed all the rest.
Dan: She read the whole thing to you, huh?
Heida: Yeah,
. . .
nu-uh, you read some.
Dan: Oh, that's right; yeah, I readed the beginning of it.
Heida: Readed? (annoyed surprise) Read!
Dan: Oh, yeah -read.
Heida: Will you stop that Papa?
Dan: Sure

Heida's complaint shows that she 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.

These facts and considerations offer an explanation for the set of peculiar facts above (14)—peculiar in the sense that the computational architecture of syntax does not in itself require any of the properties (14a)–(14d). Rather, those properties are functional if language learning is to proceed based on the formation of generalizations, and the subsequent analysis of syntactic derivations for sentences described by those generalizations. That is, those properties exist in attested languages because they make language learnable, using a general hypothesis formation procedure.

There are also implications for linguistic universals that have been taken to be part of universal grammar. A particularly clear example is the Extended Projection Principle (EPP, Svenonious, 2002, McGinnis & Richards (in press). This was first noted as the structural/configurational requirement that sentences in English and a few other languages must always 'look' as though they have subjects, even when the subject NP is clearly without semantic content (17) (cf. Lasnik, 2001, Epstein and Seely, 2002, Richards, 2003, for a general description).

(17) 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 is an embarrassment to syntactic theory, especially in the minimalist framework, because it simply has to be stipulated, rather than following from a minimal set of structure building principles. But, we can see that it actually can be interpreted as the instantiation of the canonical form constraint, not a structural principle, but a fact of attested languages that plays a role in their learnability.

Of course, a proposal like this requires fuller investigation to check out all the scientific nooks and crannies. Most important, I have not elaborated the general learning theory here in enough detail to be convincing on its own. But this discussion serves as an outline of how a simplified model of language acquisition, based on a general model of human learning, can explain universal properties of attested languages, including some properties that have been taken to be structural. When Slobin decided to explore the question of the child's acquisition of canonical forms in different languages, it began a scientific journey that showed the universality of such a stage. That, in turn, has licensed consideration of how such statistical generalizations in attested languages might be explained by a general acquisition theory, which develops not only those generalizations but also arrives at a structural knowledge involving derivations. That acquisition theory in turn can explain why many computationally possible languages do not exist. It also explains some specific apparent structural features of language. It explains the overwhelming feature of languages as used, displaying canonical forms with canonical pattern of thematic relations, which nonetheless also have specific exceptions to those patterns.

References

Bates E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 197–193). 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 and Sanz, 1997 Empty categories access their antecedents during comprehension. *Linguistic Inquiry, 28,* 68–91

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

Carpenter, P.A., and Just, M.A. (1975). Sentence Comprehension. *Psychological Review, 82,* 45–73.

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. (1957). *Syntactic Structures.* The Hague; Mouton

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

Chouinard, M. M. & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language, 30*, 637–669.

Curtin, S, Mintz, T. and Christiansen, M. (2005) Stress changes the representational landscape: evidence from word segmentation. *Cognition, 96,* 3, 233–262.

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.

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. R. (1990). The structural sources of verb meanings. *Language Acquisition: A Journal of Developmental Linguistics, 1*(1), 3–55.

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.

Karmiloff-Smith, A., & Inhelder, B. (1973). If you want to get ahead, get a theory *Cognition, 3,* 195–212.

Lasnik, H. (2001). A note on the EPP. *Linguistic Inquiry, 32,* 356–362.

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.

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.

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 (2000). *A first language taught and learned.* Baltimore, MD: Brookes Publishing.

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.

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.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13*(2–3), 129–191.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425–469.

Richards, N. (2003). Why there is an EPP. *Gengo Kenkyu, 123,* 221–256.

Ross, J. (1972). The category squish: Endstation Hauptwort. In Paul Perntau et al (Eds.), *Chicago Linguistic Society, 8,* 316–328.

Ross, J. (1973a). A fake NP squish. In Bailey & Shuy (Eds.), *New ways of analyzing variation in English* (pp. 96–140). Washington: Georgetown University Press.

Ross, J. (1973b). 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.

Slobin, D. I. (1966) Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior, 20,* 120–136.

Slobin, D. I. (1978). A case study of early language awareness. In A. Sinclair, R. Jarvella, & W. J. M. Levelt (Eds.), The child's conception of language (pp. 45–54). Heidelberg: Springer-Verlag.

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.

Townsend, D., & Bever, T. G. (2001). *Sentence comprehension.* Cambridge: 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.

Yang, C. (2006). *The infinite gift.* New York: Scribner.