Towards a ‘Converging Theories’ Model of Language Acquisition:

Continuing Discontinuity

Joseph Galasso

English Department
California State University, Northridge

joseph.galasso@csun.edu

Incomplete Working Draft: Spring 2003
Abstract

The Dual Mechanism Model credits the Brain/Mind with having two fundamentally different cognitive modes of language processing—this dual mechanism has recently been reported as reflecting inherent qualitative distinctions found between (i) regular verb inflectional morphology (where rule-based stem+affixes form a large contingency), and (ii) irregular verb constructions (where full lexical forms seem to be stored as associative chunks). In this paper, we examine the Dual Mechanism Model and broaden its scope to covering the overall grammatical development of Child First Language Acquisition.

Proposal

This paper proposes new accounts of old issues surrounding child first language acquisition. The general framework of our proposal is based upon hybrid theories—proposals stemming from recent investigations in the areas of PDP-style connectionism, as well as from more naturalistic studies, and sample-based corpora of Child Language Acquisition. Much of what is sketched out here attempts to converge the leading tenets of two major schools-of-thought—namely, Associative Frequency learning and vs. Symbolic Rule learning. Cast from this new tenor, proponents calling for a Dual Mechanism Account have emerged advocating a dual cognitive mechanism in dealing with processing differences found amongst regular and irregular verb inflection morphology (inter alia). The main task of this paper is (i) to broaden and extend the dual mechanism account—taking it from the current slate of morphology to the larger syntactic level, and (ii) to spawn some theoretical discussion of how such a dual treatment might have further reaching implications behind more general developmental aspects of language acquisition (as a whole), namely (though not exclusively), the twin benchmarks of syntactic development regarding Lexical vs. Functional (staged) grammar, etc. Our central claim will be that whatever factors lead to a deficient morphology, say, at a given stage-1 of development—factors that may potentially lead to the postulation of a non-rule based account—these same factors are likely to be carried over, becoming a factor of deficiency in the overarching syntax. Thus, the tone of the discussion is dualistic throughout. Our main goal is two-prong: first, to assert as the leading null hypothesis that language acquisition is Discontinuous in nature from that of the adult target grammar, and that this discontinuity is tethered to maturational factors which lay deep seeded in the brain—factors which yield fundamental differences in the actual processing of linguistic material, (a so called ‘Fundamental Difference Hypothesis’), and second, to show that this early multi-word non-target stage can be attributed to the first leg of this dual-mechanism—i.e., that side of cognitive/language processing that governs (i) (quasi-) formulaic structures along with (ii) non-parameterizations. We attribute the generation of this two-stage development to
maturational scheduling—viz., a Non-Inflectional stage-1 and vs. an Optional Inflectional stage-2 (where formal grammatical relations are first learned in a lexical bottom-up fashion and then later regroup to generalize across the board in a word class top-down fashion). It is our understanding that the two-staged development involves and shares both a relevant associative style theory of learning (cf. Skinner/Associative style learning, for our former stage-1), while preserving the best of what syntactic rule-driven theories have to offer (cf. Chomsky/Rule style learning, for our latter stage-2)—hence, the entitled term Converging. By analyzing much of what is in the literature today regarding child language acquisition, as well as drawing from the rich body of work presently being undertaken in connectionism, it is our hope that a new hybrid converging theory of child language acquisition can be presented in a way that captures what is inherently good from both schools—an alternative theory that bears more flavor of truth than camp rhetoric.

<>  Why—I don’t need any ‘rule’ to see this tree here in front of me. My eyes work just fine. That is, insofar as there exists a single tree. But, how is it that my ‘tree’ gets destroyed once I move my head ever so slightly to the east and fall into view of a second tree? The mystery of it all lies somewhere in the dismantling, between a single torn branch of lifted foliage, that forces the rule—for how was I ever to know that this second tree was indeed a tree after all? (Poem based on Plato’s form).

<>  “Humans use stories that they tell themselves in order to get themselves to work on this or that. These stories often deal with confrontation between areas and ideas. From some point of view, it is almost always the case that these high-level stories are relevant only as motivation and not really relevant to what eventually happens in terms of technical understanding”. (Allen Newell)  

<>  Sometimes, stories within a certain school split—e.g., formalist debates on the amount of functionalism Chomsky can and should afford to surrender (cf. Pinker & Bloom). Sometimes differing stories converge—Neo-Behaviorists seeking out an innately based architecture (cf. Elman). In any event, differing schools-of-thought are prosaic at best, ripe with unfortunate misunderstandings that lead to the fanfare of debate. There is no clarion call behind dueling rationales; one is left merely to one’s own devises, scrambling to gather fuel for the worthy debate. All reduce to subtle argument of fine detail—very rarely is there really substantial differing. The world as we see it ultimately provides very little in the way of such dividend: perhaps ontogeny recapitulates phylogeny in every respect.
0. Overview

Periodically, say every two or three generations, our vows on science are renewed by a sweeping change of reasoning—cerebral airs that deliver their own inextricable kind of ‘off-the-beaten-path’ hedonism. These solemn changes are few and far between and constitute what the philosopher of science Thomas Kuhn called ‘Paradigm Shifts’ (a new-way of thinking about and old-something). Unfortunately, these generational spurts often provide very little in the way of true original thinking, and much of what is behind the fanfare quickly reduces to little more than the recasting of old ‘brews’ into new ‘spells’. Perhaps a glimmer of true original thought (a ‘new-something’) comes our way every two hundred years or so. We are in luck! One of the greatest breakthroughs in science has been borne in the latter half of the last century and has made its way onto the scene shrouded by questions surrounding how one should go about rethinking the Human Brain/Mind—questions that have led to eventualities in Computer Programming, Artificial Intelligence (AI), Language/Grammar, Symbolic-Rule Programs and Connectionism. Much of what sits here in front of me, at my desk, can be attributed in one way or another to this ‘new-something’, and whenever there is a new-something, whether it be steam-locomotives to transistors to tampering with DNA, there’s bound to be an earful of debate and controversy. And so remnants of this debate have edged their way ever so slowly onto the platform—from the likes of the psychologist Donald Hebb (1940s-50s) (and his revolutionary notion of ‘nerve learning’ based on oscillatory frequency), to the great debates between two great personalities in the AI field, Marvin Minsky and Frank Rosenblatt (1950s-60s), to those in the realm of language, Noam Chomsky (1960s-80s). More recently, the debates have taken on a vibrant life of their own by the great advances in computer technology. The most clearly articulated of these recent debates has come to us by two leading figures in the research group called Parallel Distributed Processing (PDP)—namely, Jay McClelland and Dave Rumelhart (1980s). Most recently, the debates have come to carry a portmanteau of claims—chief among them is the claim that human brain function, and thus human computation, is not analogues to (top-down) symbolic-based computers, but rather, the brain and its functional computations should be considered on a par with what we now know about (bottom-up) nerve functions and brain cell activations. In other words, the paradigm shift here occurs the moment one rejects the computer as an antiquated model of the brain (and language), and instead, prompts up a newer model of language and thinking based on connections and connectionism (as understood in neurological studies). In this vain, it is fair to say that we should no longer view language as a mere gathering and shaping of atomic particles or logical symbols—much like how one might view the atomic nature of computer language as it is composed of a serial string of 0’s and 1’s—rationing out sub-parts of the structure in more-or-less equal portions in hope at arriving at a larger and more cohesive general frame of language. It could be argued by connectionists that language is not only much more fluid than what any strict rule-driven/symbolic function could provide, but also that language requires a greater measure of freedom and flexibility at the bottom end. Whereas rules originate top-down, it may likely turn out that bottom-up processes better reflect what is actually going-on, at least in the initial learning processes of language. (One nontrivial note here to remember is that there is a fundamental and crucial difference between (AI) artificial computer (chips) and living brain cell (neurons): the latter must secure survival. There is no sense in the notion that silicon
chips need to secure survival, since there is no death of a chip. Cells are living organisms that must somehow ensure its survival, and this survival apparatus certainly for the individual cell, must be organized in a bottom-up fashion. Along these lines, much of what is coming out of West Coast schools-of-thought (connectionism) affords the old school of Gestalt psychology a new lease on life. Some connectionists find themselves talking-up the fact that language can’t simply be a cohesion of atoms put together in very elegant ways, but that some ‘higher-order’ of fluidness must exist. Human cognition is more fluid, more context driven. In a token manner of speaking, Kohler might carry-on here about mysterious magnetic fields which suddenly arise in the brain which pull sub-particle visual stimuli together—any notion of a gestalt brain, of course, has long been disputed (I think). However, it should be noted that Gestalt psychology continues to pave a way for a serious return in the contexts of connectionism. (In addition, as a historical footnote, let’s not forget that while Rosenblatt’s work originated with visual perception, it is now viewed that his work, if carried-out in today’s climate, would have had potentially serious linguistic implications.). And so let us turn to language. With specific regards to grammar, the Word-Perception Model of Rumelhart and McClelland (1981, 1986) has made a dramatic impact in the field. Not only has it provided us with a new way of looking at potential brain processing (a quantitative way of looking with regards to weights of connections, thresholds, memory storage, etc.), it also made rather precise claims about what kinds of material (qualitative) would be difficult to process in such a model: (the need for hidden units regarding 2-degree complex structures and paradigms, recursive complexity and back-propagation, etc.). Clearly, when one can predict with a fair amount of certainty where problems will be had, and then attempt to account for the nature of the problem in terms of the model, then surely the criterion of explanatory value is close to being met. For example, the now conceded fact that ‘hidden units’ must be installed (p.c. Jeff Elman, as part of the innate apparatus) in order for the full complexity of language to be process via any PDP, I believe, speaks volumes to where we stand today in explanatory value—in fact, hidden units have now become the main rallying cry for those who postulate for rule-based accounts of language (not to mention the nativists among us. See Marcus vs. Elman on this matter). Finally, the typical intransigence that often will shape and define opposing views has given way to a certain amount of movement leading to an ultimate compromise between the two leading schools of thought—as noted by the likes of Steven Pinker and Alan Prince. For instance, Pinker & Prince’s somewhat tentative and partial acceptance of a connectionist model regarding specific lexical processes, if nothing else, has buttressed their own alliances in the pursuit of upholding counter-claims against proponents for a pure ‘Single Mechanism Model’ (strictly based on associative learning). And so out of this twist of faiths, a renewed and rejuvenated interest in rule-driven processes has been gathering momentum in an attempt to seek more narrowly confined rule-based analogies for dealing with specific aspects of language/grammar as a whole. Finally, as suggested by Newell above, long-standing dichotomies often provide a variety of clever means to think about a wide range of topics. It goes without saying that as a pedagogical device at least, students not only crave a good debate, but more importantly, they often report that new material introduced in the form of a debate procures a much higher level of understanding. Well, this singular debate has been ongoing for centuries, merely masked under several different labels: nature vs. nurture, innate vs. learned, hard-wire vs. soft-wire abilities, instinct vs. learning, genetic vs. environment, top-down vs. bottom-up strategies, and as presented herein, the Single vs. Dual Mechanism Model.
Introduction

1. It is a fact that children do not produce ‘adult-like’ utterances from the very beginning of their multi-word speech. And so much of the debate ongoing in child first language acquisition has been devoted to the nature and extent of ‘What gets missed out where’. Theory internal measures have been spawned every which way in effort to account for the lack of apparent adult-like language in young children—Theories abound. Despite some evidence that would seem to point to the contrary, more robust syntactic theories from the outset continue to view the very young child as maintaining an operative level of language closely bound to abstract knowledge of grammatical categories (Pinker 1984, Hyams 1986, Radford 1990, Atkinson 1992, Wexler 1996, Radford & Galasso 1998). For instance, Pinker (1996) has described early language production in terms of a first order (general natives) cognitive account—suggesting a processing ‘bottleneck’ effect which is attributed to limited high-scope memory to account for the child’s truncated syntax of Tense/Agr/Transitive errors (e.g., Her want), and over application Tense errors (e.g., Does it rolls?). Radford (1990), Radford and Galasso (1998), on the other hand, has maintained a second order (special nativist) maturational account affecting syntactic complexity in order to explain the same lack of adult-like speech. It should be noted that these two nativist positions share a common bond in that they are reactions to much of what was bad coming on the heels of work done in the 1970s—theories which sought to account for such errors on a purely semantic level e.g., Bloom (1975), Braine (1976) and to some extent Bowerman (1973). Steering away from potentially non-nativist associative/semantic-based accounts to proper syntactic-based accounts was viewed by most to be a timely paradigm shift—acting as a safeguard against what might be construed as bad-science Behaviorism (of the purely semantic kind). This shift brought us toward a more accurate ‘Nativist’ stance swinging the Plato vs. Aristotle debate back to Plato’s side, at least for the time being (as witnessed in Chomsky’s entitled book ‘Cartesian Linguistics’) —a move keeping in line with what was then coming down the pike in Chomskyan linguistics. One thing that seems to have caught the imagination of developmental linguists in recent years has been to question again the actual infrastructure of the child-brain that produces this sort of immature grammar—namely, a rejuvenated devotion has reappeared in the literature circumscribing new understandings of age-old questionings dealing with Theory of the Brain/Mind.

2. For instance, proponents of Behavioral/Associationist Connectionism today (cf. Jeff Elman, Kim Plunkett, Elizabeth Bates, among others) are more than ready to relinquish the old Chomskyan perspective over special nativism (‘special’ in that language is viewed as coming from an autonomous region in the brain, unconnected to general cognition or other motor skill development, pace Piaget and vs. general nativism), and have rather shifted their locus on an innateness hypothesis based not on natural language (per se) but rather on a type of innateness based on the actual architecture itself that generates language (architecture meaning brain/mind: viz., an innate Architecture, and not an innate Universal Grammar).
3. For Chomsky, it was this autonomous language faculty (that he refers to as a language organ) that allowed this innate language knowledge to thrive and generate grammar. For the connectionist movement, it is the very architecture itself that is of interest—the input/output language result being a mere product of this perfected apparatus. So in brief, the debate over innateness has taken on a whole new meaning—today, perhaps best illustrated by this more narrow debate over General vs. Special Nativism. We shall forgo the meticulous details of specific theories at hand and restrict ourselves to the rather prosaic observation that the child’s first (G)rammar (G1) is not at all contemporary with the adult (T)arget grammar (Gt). Notwithstanding myriad accounts and explanations for this, for the main of this paper, let’s just simply examine the idea that the two grammars (child and adult)—and we do consider them as two autonomous and separate grammars—must partake in some amount of Discontinuity: (Gt is less than equal to G1, or Gt < G1) and that such a discontinuity must be stated as the null hypothesis tethered to maturational/biological differences in the brain. Hence, G1 represents the (B)rain at B1…(B2…B3…Bt ), while Gt represents the brain at Bt).

4. Discontinuity theories have at their disposal a very powerful weapon in fighting off Continuity theories—whether it be language based, or biological based (noting that for Chomsky, the study of Language, for all intents and purposes, reduces to the study of biology). This great weapon is the natural occurrence of maturational factors in learning. In fact, on a biological level, maturation is taken to be the null hypothesis—whether it be e.g., the emergence and consequent loss of baby teeth, to learning how to walk-talk, to the onset of puberty. In much the way the adult achieves, the achievement can be attributed to the onset of some kind of scheduled-learning timetable—for language, it’s an achievement mirroring a process in which the nature and level of syntactic sophistication and its allocation is governed in accordance to how the brain, at the given stage, is able to handle the input.

5. It is common knowledge that (abstract) grammatical relations are frequently a problem for language acquisition systems. Early reflection on this was made by Brown when he discovered that one could not explain why some grammatical morphemes were acquired later than others simply in terms of input. The question was posed as follows: If all morphemes are equally presented in the data-stream at roughly the same time—contrary to what might be believed, parents’ speech toward their children is seldom censored so as to bring about a reduced mode of grammatical communication/comprehension—then, what might account for the observed asymmetrical learning? Similarly, Pienemann (1985, 1988) has made claims for a grammatical sequencing of learning second language based on complexity of morphology. This question led to early notions of a linguistic maturational timetable much like what Piaget would have talked about regarding the child’s staged-cognitive development—maturation being the only way to address such a staged development. Likewise, a Chomskyan position would have it that there must be something intervening in the child’s (inner) brain/mind (albeit not tied to cognition) that brings about the asymmetrical learning since there’s no change in the (outer) input. Well, one of the first observations uncovered by Brown was that a child’s linguistic stage-1
(with multi-word utterances (MLUw) lower than 2) went without formal functional grammar. In other words, Brown noted that a telegraphic stage of learning was absent of abstract grammar such as Inflection, Case and/or Agreement.

6. One consequence of this style of learning was that children were considered to learn by rote-methods, associative means similar to what Skinner had earlier advanced in (his ‘bad science’) Behaviorism. It was somewhat tentatively suggested here regarding a very early stage-1 that children didn’t start learning language as a set of rules of logic (as Chomsky would have us believe in his notion of generative grammar), but that children would first grapple with the linguistic input by gathering and constructing formulaic chunks. Children would only later on, say at a stage-2 of language acquisition, start to employ Chomskyan rules to generate a target grammar (as a consequence, see ‘U-shape learning’ discussed below). For example, Bellugi (1967), Klima and Bellugi (1966), Bellugi (1971), initially allowed for a certain amount of formulaic miscalculation to enter into the accounting of non-adult-like stage-1 structures. More specifically, recently Rowland and Pine (2000) have similarly suggested that e.g., early Subject-Auxiliary inversion errors such as *What he can ride in? (inter alia) (along with the optional target structures showing inversion What can he ride in?) cannot be accounted for by a rule-driven theory—viz., if the child has access to the rule, the theory would then have to explain why the child sometimes applies the rule, and sometimes fails to apply it. Rowland & Pine rather suggest an alternative account by saying that as a very early strategy for dealing with complex grammar (e.g., Aux. Inversion, Wh-fronting) children learn these semi-grammatical slots as lexical chunks—a sort of lexicalized grammar—whereby they establish formulaic word combinations: e.g., Wh-word + Auxiliary as opposed to Auxiliary + Wh-word combinations. It was shown that aspects of error rate and optionality (as opposed to rule-driven mechanisms) highly correlated to high vs. low frequency rates of certain combinations in the child’s input. This early non-rule-based strategy was then able to account for the vast array of the child data—viz., where the number of non-inverted Auxiliaries vs. inverted Auxiliaries were at a significantly higher rate at the initial stage-1 of development. As an example of a non-rule-based account here, they show that when inversions did occur, they typically involved only a certain select few Wh-words, and not the entire class. Hyams (1986, p.85) somewhat agrees with such a reduced structure when she asserts the following: By hypothesis, the modals (or Aux. Verbs) are unanalyzable during this period. Such overall claims strongly support Stromswold’s (1990) statistical data analysis which clearly demonstrated that children at a very early stage-1 might not productively realize an utterance string containing [don’t, can’t] in e.g., I/me [don’t] want, You [can’t] play as the syntactic elements [{Aux} + clitic{n’t}], but that such strings were more limitedly realized as quasi-formulaic representations of a negative element. In other words, the claim could be extended to mean that for the child at this stage-1, the lexical item don’t/can’t reduced to the one-to-one sound-meaning of not: e.g., Robin [don’t] [=no(t)] play with pens (Adam28) where the verbal inflection {-s} goes missing since it isn’t analyzed as an Aux Verb. Likewise, Brown came to similar tentative conclusions by recognizing that (i) verbal inflection seemed not to be generalized across all verbs in the initial stages, and therefore, that (ii) children didn’t
really start with rules, but rather employed a strategy of ‘lexical-learning’. Early stage-1 inflected verbs might then be learned as separate verbs (chunks) thus explaining observable optionality: since, as the story was then told, either you knew a rule (and so you always applied it) or you didn’t. Optionality of verbal inflection was then seen as a dual process of word acquisition in the brain: both uninflected and inflected words were stored as two different items in the lexicon. (See Bloom 1980 for comments). This notion of a stage-1 learning via non-rule-based means implied that the stage was a formulaic stage, and set-up in such a way as to learn by associative processes buttressed by frequency learning.

The Dual Mechanism Model

7. It has recently been hypothesized that the language faculty consists of a dualistic modular structure made up of two basic components: (i) a Lexical component—which has to do with formulating lexical entries (words), and a (ii) Computational component—which is structured along the lines of algorithmic logic (in a Chomskyan sense of being able to generate a rule-based grammar). It is argued that these two very different modes of language processing reflect the ‘low-scope’ (1st order) vs. ‘high-scope’ (2nd order) dichotomy that all natural languages share. Low/High scope would be described here in terms of a how and where certain aspects of language get processed in the brain (see also section # below on brain studies). In addition to newly enhanced CT brain imaging devices, multidisciplinary data (e.g. linguistic, psychological, biological) are starting to trickle in providing evidence that a dual mechanism is at work in processing language. Results of experiments indicate that only a dual mechanism can account for distinct processing differences found amongst the formulations of irregular inflected words (e.g., go>went, foot>feet) and regular inflected words (e.g., stop>stopped, hand>hands). The former (lexical) process seems to generate its structure in terms of stored memory and is taken from out of the mental lexicon itself in mere associative means: these measures are roughly akin to earlier Behavioristic ideas on frequency learning, etc. fashionable in the 1940s-1960s and made notable by the experimental work of D. Hobb and B.F. Skinner. The latter regular mode of generating structure is tethered to a Chomskyan paradigm of (regular) rule-driven grammar—the more creative, productive aspect of language/grammar generation. Such regular rules can be expressed as [Stem]+[affix] representations, whereas a stem constitutes any variable word <X> (old or novel) that must fit within the proper categorization (parts-of-speech) stem. For instance, using a simplified version of Aronoff’s realization pair format (1994, as cited in Clahsen 2001, p. 11), the cited differences in parsing found between e.g., (i) a regular [Stem + affix] (decomposed) construction vs. (ii) an irregular copular ‘Be’ [Stem] (full-form) lexical item can be notated as follows:

a.<[V, 3sg, pres, ind], X+s>
b.<[V, 3sg, pres, ind, BE], is>
The regular 3Person/Singular/Present rule in (a) spells out the bracketed functional INFLectional features of Tense/Agreement by adding the exponent ‘s’ to the base variable stem ‘X’. The features in (b) likewise get spelled; but rather than in the form of an exponent, the features are built into the lexeme ‘BE’ by the constant form is. Once the more specific, irregular rule is activated, the default regular rule-base spell-out is blocked-preventing the overgeneralization of *bes.

8. **INFlection.** Recent research conducted by Pinker (MIT), Clahsen (et al.) (Essex), among others has shown that a dual learning mechanism might be at work in acquisition of a first language. The research first focuses on terminology. It is said that there are two kinds of rules for Inflection: an Inflection based on lexical rules, and an Inflection based on combinatory rules. In short, the types of rules are described as follows:

   (i) **Lexical Rules:** Lexical rules (or lexical redundancy rules) are embedded in the lexical items themselves (‘Bottom-up’). Lexical rules may be reduced to being simple sound rules somewhat akin to statistical learning; for instance, associative regularities are built-up from out of the sequencing of lexical items—e.g., the <$sing$sang$sung -> $ring$rang$rung> sequencing of an infix (vowel change) inflection (presented below)

   (ii) **True Rules:** Word inflection of the former type (i.e., lexical rules) is cited as an inflection not based on rules, but rather encoded in the very lexical item itself. True Rule (or affixation), on the other hand, would be a combinatory symbolic process based on variables—a creative endeavor not bound by associative input (‘Top-down’). Whereas lexical-based inflections are exclusively triggered by frequency and associative learning methods—i.e., they are not prone to deliver the creative learning of novel words with inflection—Novel word inflection is generated (by default) once the true rule-based grammar is in place. One simple example that Pinker and Clahsen give in illustrating lexical/associative Inflection is the irregular verbs construction below:

9. **Irregular Verb Constructions: The #ing>#ang>#ung paradigm**

   **Table 1**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a). sing &gt; sang &gt; sung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b). ring &gt; rang &gt; rung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c). *bring &gt; *brang &gt; *brung</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   The cause of this commonly made error in (9c) is due to the fact that the phonological patterning of rhyme #ing>#ang>#ung—as a quasi-past-tense infix (lexical-rule) form—is so strong that it often over-rides and outstrips the default regular (true-rule) form of V+{ed} inflection for past tense. (Spanish offers many similar examples
where frequency of irregular verbs affect the paradigm such as the irregular (incorrect) *Roto (=Broke) over-generalization generated from the regular inflection Romp-ido.*) (*marks ungrammatical structures)

10. The erroneously over-generated patterns of *bring>brang>brung (for English) and *Roto (for Spanish) are heavily based on statistical frequency learning in the sense that the sound sequences of other irregular patterns (e.g., ring>rang>rung,) contribute to the associative patterning. Recall that structured lexical/associative learning merely generalizes, by analogy, to those novel words that are similar to existing ones. Regular grammatical rules (true rules), on the other hand, based on affixation, may apply across the board to any given (variable) syntactic category (such as Verb, Noun). In one sense, the ultimate character of ‘true rules’ is that which breaks the iconic representation of more primitive, associative-based processes, whether it be a neuropsychological process or some other process.

11. The point that the actual over-generalized strings (bring>brang>brung) are not found in the input demonstrates that there is some aspect of a rule evoked here—albeit, a rule based on rhyme association, and thus not a ‘pure rule’ where true (non-associative) variables would be at work. In other words, these lexical rules are to be generalized as a form of associative pattern learning, and not as a true rule, since they are associated with sound sequencing only. One crucial implication of an Inflection generated by a true-rule is that such inflection could be easily applied to novel or unusual words: viz., words never before heard in the input (contrary to frequency learning of lexical rules discussed above—cf. Brown (1957), Berko (1958).

12. Expanding on previous studies which examined differences in priming effects between Derivational and Inflectional morphology, Clahsen concludes that difference in priming effects can only be accounted for by a dual mechanism of learning—interpreting the data to show that high priming effects were connected with productive inflecational forms not listed in the mental lexicon, where as low priming effects were connected to productive derivational forms associated with stem entries.

13. With regards to German forms of pluralization, Clahsen (et al. p. 21) note that the same argument can be made for a dual mechanism process—viz., the high priming regular (default) plural ‘-s’ (auto-s) contrasts with the low priming of the irregular plural ‘-er’ (kind-er). The raw findings here suggest that certain irregular inflections in German (e.g., participle {-n}, plural {-er}) might be stored in the lexicon as undeveloped form chunks and that these two processes of storage are activated in very different places and manners in the brain—viz., the findings that irregular inflections spawn reduced priming as compared to regular inflection suggest that regular inflections are built forms based on rules that contain variables which make the basic unmarked stem/root available for priming. It is clear from the table below that regular inflected word forms such as {-t} participles and {-s} plurals produce full priming and no word-form frequency effects. For irregular inflected forms such as {-n} participles, {-er} plurals and (irregular) {-n} plurals the opposite pattern appears. The data suggest that irregular forms are stored as underdeveloped stems—hence the
emergences of full form frequency effects. Regular forms are captured by the full rule process and are stored in a computational manner that works off of variable+stem algorithms—hence, the lack of full-form frequency effects. These differences in German morphology seem to parallel what we find between English (i) Inflectional morphology and (ii) Derivational morphology where the former seeks out specific rule formulations—e.g., V + {ed} = Past, or N + {s} = Plural, etc. and where the latter seeks out associative style sound-to-meaning learning approaches (as in irregular verbs/nouns e.g., go>went, tooth>teeth, etc.) Applying fMRI brain imaging techniques, a consensus has begun to emerge suggesting that the lexical storing of derived stems + suffixes (e.g., teach+{er}) may actually be processed as one single word chunk in the otherwise lexical (word/recognition) temporal-lobe areas of the brain, and not, as intuition would have us believe, as a dual segmented [stem + suffix] lexical structure which has undergone a process much like a morpho-syntactic string). This may be an apparent economical move keeping in line with the classic one-sound-one-meaning association. In noting this, there seems to be a natural tendency in the diachronic study of language to move from (i) rule-driven Inflectional morphology—with more complex rule-driven infrastructures [+Comp] (Comp=complex) to less complex [-Comp] structures—to (ii) association-driven Derivational morphology. This tendency can be easily captured by looking into the way words have evolved over a duration of time—e.g., Breakfast /breɪkˈfast/ has evolved from a twin morpheme structure [[Verb Break] + [Noun Fast]] > to Breakfast /brɛːkˈʃt/ [Noun Breakfast] composed of a single morpheme chunk.

Table 2  Summary of experimental effects  
(Taken from Clahsen et al. 2001: p.26)

<table>
<thead>
<tr>
<th>Representation</th>
<th>Full priming effect?</th>
<th>Full-form frequency effect?</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>-t participles: ge[kauf]-t</td>
<td>yes</td>
<td>no</td>
<td>Sonnenstuhl et al. (1999), Clahsen et al. (1997)</td>
</tr>
<tr>
<td>-s plurals: [auto]-s</td>
<td>yes</td>
<td>no</td>
<td>Sonnenstuhl &amp; Huth (2001), Clahsen et al. (1997)</td>
</tr>
<tr>
<td>-er plurals: [kinder]</td>
<td>no</td>
<td>yes</td>
<td>Sonnenstuhl &amp; Huth (2001)</td>
</tr>
<tr>
<td>-n participles: [gelogen]</td>
<td>no</td>
<td>yes</td>
<td>Sonnenstuhl et al. (1999), Clahsen et al. (1997)</td>
</tr>
<tr>
<td>-n plurals I: [bauern]</td>
<td>no</td>
<td>yes</td>
<td>Sonnenstuhl &amp; Huth (2001)</td>
</tr>
<tr>
<td>diminutives: [[kind]chen]</td>
<td>yes</td>
<td>yes</td>
<td>Clahsen et al. (2001)</td>
</tr>
</tbody>
</table>
14. In sum, Pinker and Clahsen assume that the language faculty has a dual architecture comprising of (i) combinatory rule-based lexicon (leading to the lack of full-form effects) and (ii) a structured non-rule-based lexicon (leading to full-form effects). Questions on specifics will surface in the following sections-namely: How are these two methods represented in the brain?

15. A Stage-1 Language Acquisition. There is a huge and ever-growing body of data today being collected by developmental linguists in the field which suggests that the brain of a child matures in incremental ways which, among other things, reflects the types of ‘staged’ language development produced by the child for a given maturational stage. The collected data suggest that children’s early multi-word speech demonstrates ‘Low- Scope’ lexical-specific knowledge, and not abstract true-rule formulations attributed to grammar. Somewhat akin to Piagetian notions of language development (see general nativism below): One difference being that it need not be tied here, exclusively, to a cognitive apparatus. This staged, maturational theory of language development accounts for the lack of specific linguistic properties by suggesting that the brain is not yet ready to conceptualize higher and more abstract (High- Scope) forms of linguistic conceptualizations.

16. The idea behind ‘What gets missed out where’ in child speech production has given those linguists interested in morphology and syntax a particularly good peek at how the inside of a child’s brain might go about processing linguistic information—and other information for that matter. As stated above, research initially carried out by Brown and his team (1973), working under a Chomskyan paradigm of linguistic theory, and consequent work by others (cf. Radford) suggests that there is a stage-1 in language acquisition that tightly constrains the child’s speech to simple one-to-two word utterances with no productive forms of verb or noun inflection. One child that appears in the early studies, Allison, provides transcripts between 16-19 months showing no signs of the onset of formal inflectional grammar—only later-on close to two years of age (22-24 months) does inflectional grammar/syntax emerge, and then only in what could be said as a sporadic, optional manner.

17. This stage-1 is considered to be a grammatical stage with an MLUw (Mean Length of Utterance word) of 2 words or less. More specifically, in the sense of the apparent lack of formal grammar, this shouldn’t be confused with the idea of an earlier a-grammatical stage well before the onset of multi-word speech. (Surely, there can be no grammar or syntax of which to speak if there aren’t multi-word constructions). This grammatical stage-1 therefore differs with the notion of a one-word stage (MLU=1) where supposedly absolutely no grammar/syntax is at work. The grammatical stage-1 is said to begin roughly with the onset of multi-words at about the age of 18 months (+/-20%). It is reasonable to suppose that such a stage would have target semantic meaning—even though, say the arbitrary ‘one-to-one sound-to-meaning’ relationship is not of the target type (e.g., onomatopoeia forms /wuwu/=dog, /miau-miau/=cat, etc.).
18. The above notions beg the question: At what point do we have evidence of grammatical categorization? For example, the traditional distributional criterion that defines the Noun class as that category which may follow Determiners (a/the/many/my/one) made not be available to us if, say, Determiners have yet to emerge. Hence, distributional evidence may be lacking in such cases. One way around the dilemma has been to suggest that early stage-1 grammar is categorical in nature simply owing to a default assumption that categorization is part of the innate ability to acquire language (in Chomskyan terms, part of the richly endowed LAD or Language Faculty) and that words are both inherently categorical and semantic in nature. Pinker (1984) claims that the categorization of early stage-1 words should be roughly pegged to their inferred semantic properties. Radford (1990), in a slightly different approach, prefers to consider such early multi-words at stage-1 as lexical in the sense that (i) they have built-in default lexical categorization abilities (forming classes of Nouns, Verbs, Adjectives, Adverb, and Prepositions), but, at the same time, (ii) rely heavily on their semantic-thematic properties. In any event, either description starkly contrasts with a connectionist view—which claims that e.g., the class ‘subject’ emerges through rote-learning of particular framed constructions. Subject-hood is learned as a category via rote associative learning of thematic relations. Now, it remains unclear to me precisely how close such thematic links to category-hood get to Radford’s 1990 interpretation. I would only venture to say that both views share the belief that semantics hold the central cognitive underpinnings upon which syntax can later be built.

19. This account of stage-1 has been labeled as the Lexical thematic stage-1 in language acquisition (Radford 1990). It is unclear how far Radford would like to go in accepting his stage-1 as cognitively based: the labeling here of lexico-thematic (the term thematic referring to argument structures pegged to semantics) certainly permits some amount of semantics to leak into the discussion. Nevertheless, Radford emphatically rejects the notion that a stage-1 syntax could be exclusively based on semantics. It is here that Radford gets full mileage out of his two-prong converging Lexical-Thematic stage-1 grammar: a stage-1 that is both—

(i) ‘thematic’ in the sense that it leans towards general nativism since simple utterance types at the earliest MLU get directly mapped onto their thematic argument structure; while,

(ii) ‘lexical’ in the sense that the child seems to be fully aware that they are dealing with words based on lexical grammatical categories, and not semantic. This is made apparent by how children know the morphological range of category (e.g., Noun, Verb) selectiveness along with inflection distribution.

20. One argument against a semantically based stage-1 was that from the very beginning, children’s productive multi-word speech (MLU= 2+) yielded Inflectional plurals {+s} and gerund {+ing} endings—the first two morphemes to be acquired according to Brown’s morpho-sequencing list. These endings were only attached to syntactic categorial word-classes: e.g., {s} to nouns, {ing} to verbs, etc. There seemed to be no
attempt by the young child to generalize such inflections onto pure semantic categories. In other words, if children’s word classes at this stage-1 were thematic, rather than syntactic in nature, we would expect that specific inflections would be distributed along semantico-thematic lines: e.g., plural {s} to agent, gerund {ing} to action words, etc. (Radford 1990, 41). Such findings are not reported in the data. It was this absence of semantically based grammars which led discussions about possible a priori innate grammatical categories, a grammar based on a syntax (without meaning) rather than a syntax based on semantics (meaning) (cf. general vs. special nativism). Although it is indeed correct to suggest that there seem to be no purely semantically based Inflections at stage-1, one argument against the conclusion of the claim, and seemingly in support of a semantically-based stage-1, would be to suggest that, in fact, most utterances at this stage are instances of formulaic constructions. Only at a later stage-2 would we find instances of real productive inflection—viz., even though on the surface, inflection appears to be utilized at stage-1, the surface structure only mimic input driven phonological patterns.

21. This ‘mixed bag’ of a grammatical stage is indeed an argument against ‘too-strong-of-claim’ syntactic-based model of early grammar (assuming that a syntactic version holds as a buttress for Continuity—we shall take some comfort in it however due to the fact that this strong claim we take will be short lived and relegated to the very earliest of grammatical stages: (=MLU below 2). There is a caveat here. One argument, however, against interpreting from no evidence-namely, the observation that no inflection shows up on argument-themes might be the following: If our stage-1 were in fact formulaic, and not rule-based, then there indeed would be no utterance of an improper formulaic inflection attached to a semantic category simply because this would not have been available in the phonological input. Formula constructions come out of the input in a highly regular manner—based on high frequency, saliency and churn out as formulaic un-analyzable chunks. (See 40 for an account of apparently correct parameterized word order found at an otherwise non-parameterized stage of acquisition).

22. The argument could run as follows. The fact that children at stage-1 never produce e.g., the action-inflection ‘-ing’ to semantically classed action-words like *up-ing/down-ing/over-ing/on-ing, etc. merely indicates that such strings are not part of the available input (particularly note worthy given that our stage-1 is semi-formulaic in nature). It will be argued that the very earliest of stages (stage-1), addressed herein, is indeed the very earliest of staged developmental grammar—what may have been even termed a-grammatical in previous theories (viz., the one word stage (cf. Atkinson, 1992; Radford, 1990; among others). Let it be known that I am all too ready to acknowledge and agree that language is indeed built upon pure syntax at our stage-2 of development, (and not on semantics): the classic evidence for a syntactic-based language at the earliest stages has been taken from the child’s inflectional system at work on the basis of grammatical categories. Notwithstanding early attempts to cast syntactic analyses to early stages of language, there has been attempts in the child language acquisition literature to construct a dual model for stage-1 based on (i) semantico-thematic relations on one hand, and (ii) categorial syntax on the
other. This hybrid model has been considered as a lexical-thematic stage-1 of child language acquisition where mere semantic properties tied together those lexical syntactic categories void of any functional material (as related to the functional categories IP & CP). The most fully articulated version of this hybrid theory could be found in Radford (1990):

```
(Stage-1)    VP   (Radford 1990: 45)
     / \  
    NP  V'
   / \  
Agent V NP
   \  
  Patient
```

23. The question is then put to us in the following form: Is there any evidence at the earliest phases of stage-1 (say MLU < 2) that the child actually analyzes strings as a syntactic structure—as opposed to a formulaic speech-utterance (i) which may be tethered to a variety of gradient meanings, and (ii) which may reduce to mere surface-level syntactic phenomena)? In other words, what may appear on the surface as syntax proper, may in all actuality simply be a result of the surface formulae learned and that real tacit syntactic knowledge is not represented. There seems to be little that hinges on the possible alternatives:

   If, on the one hand, we consider such semi-formula as syntax proper-making our stage-1 (MLU < 2) a syntactic stage—then so be it. We are then forced to reconciling our syntactic stage-1 to the one word stage as previously thought and nothing is lost.

   If, on the other hand, a lexical-thematic stage-1 involved itself with bridging this narrowing gap between formula and syntax—then so be it. The benefits we have gained by adapting this measure is that it allows us a nice continuity bridge onto the later phrases of stage-1 (MLU +2).

24. One interesting by-product of such a lexical-thematic stage-1 is that it doesn’t specify Word Order: word order being traditionally tied to functional parameterization (see Atkinson, 1992; Tsimpli, and Galasso, 1999). Coming on the heels of such semantic-based models of language acquisition, claims have been made suggesting that the cause of a semantic stage-1 is due to memory deficits. As part of a Maturational timetable, the child starts off with a very limited memory attention span—this memory deficit (maturational based) triggers the more ‘robust & primitive’ semantic-lexical level of language (since the lexical component is more salient) to kick start productive communication (see Newport’s ‘Less-is-More Hypothesis’, S. Felix’s non-UG/cognitive approach to L2 learning, as well as J. Elman’s work in relation to connectionism. For evolutionary accounts, see Bickerton’s Proto-language, 1990).
Less-is-More Hypothesis. According to Newport's 'Less-is-More' Hypothesis, a Radfordian style maturational time-table—dividing our stage-1 from stage-2—would be linked to 'working memory' deficits: Stage-1 starts with early limited memory and thus can solely rely on the more primitive and robust rote-learned and formulaic structures. (One needn't say that all possible structures at stage-1 are rote or formula—let it suffice to say that the flavor of the stage suggests little if any evidence for 'true-rule' formations or parameterizations, citing stage-1 variant Word Orders and null INFLections). This handicap of low memory actually works as an advantage for the child in that it serves to constrain the perceived input to basic degree-0 SV(X) structures—the structures are ready-made by the lower-level cognitive processes and made available to the stage-1 child. Lower-level memory seeks out idiomatic lexical-based categories or lexical based morphemes as opposed to functional, syntactic based morphemes/categories (termed 'l'-morphemes vs. 'f'-morphemes respectively by Pesetsky (1995) as understood in Distributional Morphology (see §50). (N.B. Felix (1981) as well as Krashen claim that it is precisely this over-production of the cognitive apparatus/high memory that makes second language learning so fraught with difficulty—having to 'learn' language overtly instead of naturally 'acquiring' it in a natural setting.)

We can better frame arguments that claim for a cognitive/memory dependence for language acquisition by addressing the very nature of syntax. First, syntax requires much more in the way of computational memory. (Or perhaps the question is better framed conversely—viz., more memory forces the computation to reorganize itself by way of syntax.) The emergence of syntax coincides with the onset of higher (quantity) amounts of language material—i.e., a higher number of memorized words/strings leading to longer and a richer complexity of sentences, etc. For instance, Degree-zero structures (say, basic SV sentences, order irrelevant) come at the expense of lower memorizations, while, et vice versa, Degree-1 structures, (embeddings, binding, recursiveness) come at a much higher cost with regards to memorization. Why is that? Well, in one manner of speaking the reason is self serving: simply due to the fact that in order to have a degree-1 sentence, the empirical (maturational) data dictates that a child must have, at some prior time, gone through a degree-0 stage, a process that mirrors memorization capacity. But more to the point, the reason for this mental/computational juggling has to do with how our brains go about making the most out of our limited memory capacity. The very nature of these high amounts of material forces a shift in how the brain can process (parse) the material. It is believed in the neuro-linguistic community that the shift here—both in the quantity and quality of language—triggers the already over burdened process of rote-learning and memorization to be lifted, triggering the share of burden to be replaced by rule-based processes (variables, categories, etc.). Such rule-based learning frees up space in the lexical component of the brain (say, the list of words stored) and allows new routes to be mapped. In other words, such a huge volume of material forces new ways of organizing the input (hence, categorization). In sum, the two-prong development as sketched out above might proceed as follows:
At the Micro-Development level (stage-1) the data stream is reduced for the child in terms of its cognitive saliency: (the output data is not changed, rather it’s the intervening deficiency of the child’s mental processing that overall affects that data). The child, working with a primary memory ‘tool-kit’, allows a small subset of language input, this in turn allows the child to ultimately deal with less data enabling rote-learning to take place. (N.B. It is generally acknowledged that any memory deficit or trauma resulting in language attrition would first affect the more abstract levels of language/syntax).

At the Macro-Development level (Stage-2) the data stream is affected by the upsurge in memorization that in turn expands what becomes salient for the child. Perhaps having to do with the triggering of hidden units at the end of stage-1, the child now is in a position of capably taking the data and applying paradigmatic structures—all which lead to formal (stage-2) grammar. Thus, Macro development makes available more memory which in turn spawns new ways of handling the material—the initial process of stage-1 rote association and memory is no longer adequate and syntax proper emerges as a way of handling both the quantity and quality of this newfound material.

What syntax allows the brain to do is categorize and form analogies based on the vast amount of input, rather than to memorize and store all input as meaningful chunks (with an associative sound-to-meaning relationship imposed). This results ultimately in a finite array of neuro-linguistic networks in the brain. Hence, in a basic input-output model—similar to what we understand to be happening in behaviorist stimulus and response associative models—quantity of input equates to quality of brain processing. As is evident, the classic enigma (chicken and the egg scenario) remains: Is it this newly wired brain which now seeks out the formations of paradigms and variable rules that is responsible for the quantum leap of quality of language, or is it this quality leap in language that somehow drives the changes in the brain? This is tantamount to the classic Nature vs. Nurture debate. My hunch here is that (i) the nature of the raw Data as it is (ii) tied to cognitive processing may be the driving force behind any structural changes that occur in the brain—in other words, language changed the brain and not the other way around. (It may ultimately be impossible to separate the one from the other). But this is only a hunch, and again, it reduces to the same catch-22 scenario (if it is the data that is the driving force behind the change, how do we account for a maturational protracted development, and secondly, surely, how the brain handles and processes the data must be part of the equation for any theory that attempts to account for developmental stages of language). In a certain sense, Newport’s ‘less-is-more’ hypothesis simply restates this same paradox. Regarding architecture and the nature vs. nurture debate, clearly all linguists suppose now that some connection must be made between genes and environment Thus, a two-staged development follows:
(i) Stage-1 comes with low-level memory with strong correlates to semantics and rote-learning. As a consequence, one-to-one sound-to-meaning correspondence ensues explained by more prosaic economic constraints placed on cognition.

(ii) Stage-2 comes with increased memory that (for reasons having to do with processes of parsing, etc.) triggers high level categorization and syntax. One-to-many/many-to-one relations are evoked triggering a highly rich paradigmatic grammar.

27. Radford (2001) more recently has gone on the record as saying that the Language Faculty specifies a universal set of features—namely, that a child acquiring language has to learn which subset of these features are assembled into the lexical items as universal (all other features awaiting parameterization via a maturational timetable). The problem for the child is assembling the features into lexical items. To a certain degree, the child needs to build-up lexical items feature at a time (see Clahsen’s Lexical Learning Hypothesis). Thus, the issue for Radford is that there are innate architectural principles—loosely referred to as an Innate Grammar Construction Algorithm—which determine how lexical items project into syntactic structures. This begs the following question: How much of this initial learning deficit cited for our lexical stage-1 is owed to the child’s protracted language development being exclusively tied to a maturational based low-scope cognitive template—a potentially semantic based template upon which later formal abstract categories (such as functional categories) can be mapped? It is clear at least that more abstract functional categories come on-line later in the course of development.

28. General vs. Special Nativism. This is a nice place to pause and examine the role that our lower-scope cognitive processes might play in deciphering between Stage-1 vs. stage-2 grammar. In brief, there are two schools of thinking on this, both of which could maintain general ties to a Chomskyan paradigm. One school takes an evolutionary stance (Pinker & Bloom) and basically claims that lexical learning leading to grammaticalization is heavily based on what are preexisting cognitive constraints (much in the manner of former Piagetian models of language development). Such linguists would disagree with the notion that a special module in the brain must exist in order for language to manifest. Recall, Chomsky in his strongest claims suggests that the Language Faculty (LF) is an independent autonomous organ found somewhere in the mind/brain (similar to say the liver or the stomach) and that this LF organ shares very little in the way of general cognitive processes—a language module all to its own and without common lineages to other regions or modules of the brain. This notion is referred to in the language acquisition literature as a Double Disassociation Hypothesis (disassociation between formal language and cognition) (see Smith and Tsimpli for some discussion). The second anti-Neo-Darwinian position suggests that a special module in the brain is required for language, and that language learning can be accounted for by reduced/non-cognitive means.
Regarding the debate over General vs. Special Nativism, it is still unclear how the debate should be viewed. Much of the argument quickly degenerates into the classic aforementioned ‘chicken-and-the-egg’ dilemma of being circular in nature: e.g., (i) The Special Nativist claims that the child first needs syntax to uncover the underlying semantics (syntactic-bootstrapping), while (ii) The General nativist insures that in order to properly construct a syntax category in the first place, general properties of (inherent) cognitive-semantics must be observed (semantic-bootstrapping). (Interesting, Chomsky’s most recent work on Minimalism suggest that there may be economical constraints on language processing (from out of Logical Form). While it is still unclear how to interpret the wide range of claims on the minimalist table, and Chomsky himself often remains agnostic at these levels of inquiry, such economic constraints could be interpreted as indeed not pertaining to consideration of pure syntax, and rather adhering to more cognitive levels of processing: e.g., Minimalist notions of shortest move, minimal amount of rules, and to a certain degree, the objective essence behind the (PF) phonological form of language as versus the (LF) logic form, etc.). On one hand however, it seems to me that a dualist approach to acquisition (as presented herein) would initially favor a first order semantic-bootstrapping view, given that semantics seem to play an essential role in language acquisition early on before the onset of syntax. (There is no conclusion drawn here, as nothing argued in this paper hinges on that debate).

Why—I don’t need any ‘rules’ to see this tree. My eyes work just fine. That is, insofar as there exists a single tree. How is it that my ‘tree’ gets destroyed once I move my head ever so slightly to the east and fall into view of a second tree? The mystery of it all lies somewhere in the dismantling, between a single torn branch of lifted foliage, that forces the rule—for how was I ever to know that this second tree was indeed a tree after all?

Well, the above passage makes for a nice analogy, but it merits a closer look. When I look at this cup of coffee in front of me, reach out for it, and drink its contents, it certainly appears to me that I do little more than what my own cognitive abilities lets me achieve—I don’t perform any ‘abstract rule’ formulations, procedures as such, although, I do agree that one could possibly uncover all of the aforementioned procedural content coming together such as e.g., Gestalt psychology, visual cortex processing, contextual/meta-linguistic background of say [+liquid] => drink => mouth, along with muscle motor coordination that allows me to see into space reaching and holding the cup without breaking the glass (etc.). In face of all this possible theory, it remains somewhat natural for me to maintain the idea that when I see a tree, I just see a tree (period). But much has come out of Gestalt theory in the past (being somewhat reframed here in the present context of connectionism) that suggests there may be something to this very natural notion of just seeing after all. Gestalt theory on perception states that there are first-order perceptions in which, say, a child might see a line or a slope in a strict iconic representation of the visual field.
No rules apply—and there is a strict Stimulus and Response (S&R) equation involved. Regarding language acquisition, this first-order representation could be illustrated by the early onset of vowel recognition (i.e., environmental sound)—and not sound as filtered through assimilation processes, etc. (as seen in the u-shaped model below). At a later stage of perception, second-order perceptions allow the child to break iconic mappings and allow lines, slopes, etc. to begin to be seen (with less vividness) as e.g., a chair—now, a larger, somewhat more generic unit, which embodies the lower level visual stimuli. It seems to be the case that the role of second-order perceptions is to pull and frame larger aspects of Objects and Events—in linguistic terms, forming Nouns (out of the former) and Verbs (out of the latter). So regarding language, we should be clear that by the time a child reaches the very first stages of language development—where a child is said to begin producing single word utterances—s/he has already moved from the first-order perceptual field into a second-order field. So, the idea that children may have some means to rules, perhaps bootstrapped from Gestalt psychology (the General Nativist Position) may not be totally implausible. However, and more to our point, Newport’s ‘Less-is-More’ hypothesis just as well could be interpreted to fit Gestalt findings: when memory/cognitive capacity is low, children see in a fixed iconic manner, and when memory/cognitive capacity increases, the child reorganized the visual field and must begin to classify according to class—e.g., the child sees a chair (second-order) as opposed to a chaotic string of lines and slopes, etc. So, roughly, the theme throughout holds—memory/cognitive capacity drives computational order. One way though to save our nice analogy is by pinning it down (to a narrow application) to issues surrounding Lexical S&R behaviorism vs. Functional rule-based grammar. Surely, the spirit of the analogy is well taken. Yes, iff (if and only if) I ever saw one tree, I could adhere and maintain an exclusive iconic S&R process; it is when I look and see another tree that I must compare notes and begin to re-organize both visual trees into a class of ‘Tree’ (using Plato’s terminology). Again, Newport’s theme above holds in that too much information, in this case the second tree, forces an adjustment in the computation—corresponding to our data drive axiom. In other words, on one basic and primitive level (order-1), visual transmission is nothing more than sensory input directly stimulating the sensory cortex). However, at a more abstract and functional level (order-2), perception is not fully determined by sensory input, but is dependent on intervening processes of Gestalt psychology. Hence, a dual mechanism account likewise credits a purely cognitive behavior such as vision as having two distinct modes of processing—(i) Bottom-up sensory-driven Transmission and (ii) Top down context-driven Perception. These two approaches could nicely map onto our analogous dichotomy between Skinner and S&R style learning vs. Chomsky’s rule-based symbolic style learning. So, our emerging linguistic schism separating Derivational morphological processes from Inflectional processes may not be a schism relegated to language per se, but may actually be operative in separating other lower-level cognitive procedures as well.
Discontinuity: A Lexical-Thematic stage-1

31. It is now widely reported in the literature that children generally go through a stage during which they optionally project Functional Categories: e.g., Determiner Phrases (DP), Finite Main Verb Phrases which mark Tense (TP), Infinitive and Agreement markings such as infinitive-‘to’ (IP), and 3person/present/singular {s} (AgP) (respectively). Wexler (1994) refers to this stage as the Optional Infinitive stage. In more general terms regarding Inflection, Wexler’s ‘Optional Infinitive’ stage has recently been more accurately characterized as an ‘Optional Inflection’ stage (see Radford & Galasso 1998). More importantly, a picture seems to emerge in the investigation of early child speech that shows an even earlier stage of development—a stage in which the overall deficit well exceeds any notion of Infinitive/Inflectional Optionality. Mainly speaking, there seems to exist a stage-1 in the course of child language acquisition—briefly peaking at around two years of age with MLUw well below 2.5 and then quickly falling off—which indicates ‘No Inflection’ whatsoever. What might have been too hastily claimed a stage-1 in Wexler’s terms, must now be relegated to a stage-2 in Radford & Galasso’s terms. It so happens that Wexler has been a leading proponent of maturational-based theories of language acquisition, so supporting arguments for Continuity based on the optionality data as presented by Wexler and his colleges don’t get a fair play. Likewise interpretations regarding our own work here would most certainly solicit continuity in getting at fair play—that is, however, only if it were the case that Wexler’s stage-1 indeed simply equated to our stage-1. As it turns out, it doesn’t. Our stage-1 is much more systematically void of functional material. In other words, Wexler’s OI-model doesn’t offer us a solid, foul-proof discontinuity model. By definition, ‘optionality’ suggest that the child has some working tacit competence of the adult target grammar—it may only be that the performance level or mastery of such competence is lacking. Certainly, this is a far cry from any possible notion of a ‘strong discontinuity’ theory.

32. One Continuity argument could run as follows. Since, the child at the earliest of conceivable syntactic stages is already marking Inflection (albeit optionally), then there is no justifiable reason to assume (even as the null hypothesis) that child’s grammar is Discontinuous with the adult target grammar. (In this sense, the dual mechanism apparatus has established itself from the get-go, and thus, no child-to-adult discontinuity has to be assumed). The differences found between the child’s grammar and adult target grammar would not be significantly real, in developmental language terms, and could be readily accounted for by a variety of superficial means—such as e.g., saliency conditions, morphological feature spell-out conditions, parameter miss-settings, phonological complexity and general immature cognitive factors bringing about the memory deficits of such non-salient phonological features. Having said this, a very different scenario emerges if indeed our stage-1 is a stage that precedes optionality by showing ‘no inflection’ whatsoever. In such a scenario, a discontinuity hypothesis now seems to emerge as the null hypothesis, as previously cited above, yielding to highly universal biological considerations. My own data (a syntactically-coded naturalistic corpus of well over 10,000 analyzable utterances) presented in the following sections, taken from Radford & Galasso (1998) & Galasso
demonstrate this two-prong stage of acquisition, the consequence of which will buttress our calls for Discontinuity. (N.B. It goes without saying that by postulating a ‘dual mechanism model’ for adult language systems, any working theory claiming a stage in which a child starts-off with a truncated ‘single model’ (for stage-1) would be tantamount to discontinuity).

The Inflection {‘s’}

In examining the ‘portmanteau’-morpheme {s}, the data provide (prima facie) evidence of some relation between the acquisition of Possessive {‘s} and the Third person singular {s}. At our stage-1, there is no evidence of the inflectional marker across the board, it is only with the onset of our so labeled ‘Stage-2’ (from age 3,2 onward) that we begin to see Wexler’s notion of optionality kick-in. The table 3 below shows the relative frequency of use of Poss(essive) {‘s} and 3Psing {s} in obligatory contexts before and after age 3;2:

Table 3

<table>
<thead>
<tr>
<th>Age</th>
<th>3SgPres {s}</th>
<th>Poss {‘s}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;3-3;1</td>
<td>0/69 (0%)</td>
<td>0/118 (0%)</td>
</tr>
<tr>
<td>3;2-3;6</td>
<td>72/168 (43%)</td>
<td>14/60 (23%)</td>
</tr>
</tbody>
</table>

Token sentence examples of the two-staged data are presented below (respectively):

a). That Mommy car (2;6). No Daddy plane (2;8), Where Daddy bike? Batman (2;11 in reply to ‘Whose it is’). It Daddy bike. No Baby bike (3;0).

b). Daddy’s turn (3;2). It’s the man’s paper (3;4). It’s big boy Nicolas’s. It’s Tony’s. What’s the girl’s name? Where is Zoe’s bottle? (3;6).

c). Baby have bottle (2;8). No Daddy have Babar (2;9). The car go (2;11) The other one work ( 3;0). Here come baby (3;1)

d). Yes, this works. This car works. My leg hurts. It rains He comes (3;1-3;2).

Interestingly, the data above suggest a potential parallel between the acquisition of third person singular {s} and possessive {‘s} (see Radford & Galasso for discussion).

But more importantly, they also suggest that whatever discontinuity is at work in the child’s grammar, it seems to manifest across the board in a systematic way. In other words, the lack of inflection here is not categorical specific, but rather is realized
It seems that Poss and 3PS {s} at our stage-1 both reflect general catastrophic agreement failure. Certainly, any notion of a real child-to-adult discontinuity would want to be expressed in such absolute terms—as opposed to any Optional-based theory which might be comined into spinning arguments from what on the surface would appear as mere non-mastery and underspecification of Continuity into arguments for real Discontinuity. As was expressed above, both deficits could be captured by a lack of Agreement—a functional property of adult grammar. Consider then the phrase structure discontinuity of the two stages below:

**Agreement Structure**

Stage-1: \[ \text{IP Mummy [I-agr 0] car} \]
Stage-2: \[ \text{IP Daddy [+agr 's] turn} \]

It is argued herein that both possessive {‘s} and third person {s} are reflexes of an agreement relation between an inflectional head and its specifier—and any omission reflects an agreement failure. The specific issue at hand here is that only an absolute omission stage—as seen with our stage-1—would provide support for true discontinuity. Any optionality here, e.g., [+/- agr] would play directly into the hands of Continuity theories with the mere additional disclaimer that the adult target grammar has indeed been acquired, but simply not mastered. (See Wexler & Schütze (1996) for treatments of under-specification of Agr as would be encountered in our stage-2 data).

**Possessors**

In a similar vain, we find additional support for a non-target grammar in the wake of data showing Case errors e.g., with possessors *enter alia*. The assumption that children’s possessive structures may be initially (i) non-specified, and then later (ii) (optionally) underspecified with respect to agreement also accounts for the wide array of case errors where children (at stage-1) use the default case of objective possessor (*me*) and only later come to acquire the target Case of possessor (*my*), etc. The use of objective possessors e.g., *me* has been reported for Dutch by Hoekstra & Jordens (1994), but not for English. If we look at the earliest first person singular possessor structures produced in the data, we find that objective *me* possessors predominate at ages 2;6-2;8, and that Genitive possessors (prenominal *my* and pronominal *mine*) are initially infrequent (with no cases reported for the use of nominative *I* for possessor):
Table 4  Occurrence of First Person Singular Possessors

<table>
<thead>
<tr>
<th>Age</th>
<th>Objective “Me”</th>
<th>Genitive “My/Mine”</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:6-2:8</td>
<td>53/55 (96%)</td>
<td>2/55 (4%)</td>
</tr>
<tr>
<td>2:9</td>
<td>11/25 (44%)</td>
<td>14/25 (56%)</td>
</tr>
<tr>
<td>2:10</td>
<td>4/14 (29%)</td>
<td>10/14 (71%)</td>
</tr>
<tr>
<td>2:11</td>
<td>5/24 (21%)</td>
<td>19/24 (79%)</td>
</tr>
<tr>
<td>3:0</td>
<td>4/54 (7%)</td>
<td>50/54 (93%)</td>
</tr>
<tr>
<td>3:1-3:6</td>
<td>6/231 (3%)</td>
<td>225/231 (97%)</td>
</tr>
</tbody>
</table>

a). That me car. Have me shoe. Me and Daddy (=Mine and Daddy’s). Where me car? I want me bottle. (2;6-2;8)


c). It is my t.v. Where is my book? Where is my ball? Don’t touch my bike. I want my key. It’s my money. (3;0)

37. In terms of the analysis outlined above, the data seem to suggest that the possessive structures produced early on (=stage-1) are predominately not specified for possessor-agreement, with agreement gradually being specific more and more frequently (until it exceeds 90% mastery at age 3;0). While it is true that we can’t argue here for absolute Non-agreement of Case at stage-1 (whereas for the earliest file, age 2;6, we get at least two examples of correct my), this contrast in acquisition—as compared to what we observed earlier regarding the agreement of the morphological inflection {s}—may be a residual effect of the two types of agreement involved: its seems to be the case that true morphological inflection should be the benchmark of agreement and not lexical equivalents e.g., prenominal/pronominal my/mine (respectively) due to the fact that it is always more difficult to tease apart lexical form functional underlying structure and determine if a lexical item is being properly projected as a functional category, or if merely the lexical ‘shell’ is simply phonologically produced ‘rote-learned’. (Also see §50ff regarding such distinctions placed between the two features as understood in Distributional Morphology). The above examples could be expressed by the same type of Phrase structure presented below:

Agreement Structure

1. Stage-1 [IP me [-agr 0] car]

2. Stage-2: [showing +/-optional Agr(eement)]:
   (i) [IP me [-agr0] pasta] [showing +/-optional Agr]
   (ii) [IP me/my/mine [+/-agr] pasta]
   (iii) [IP my [+agr] pasta]
It could be argued that for our stage-1, ‘adult target’ agreement (acting as a functional and formal feature of language) is set to the default via a [-agr] setting and so renders the possessor case objective. The close to 100% omission of adult-like agreement provides additional support for a discontinuity theory between child and adult grammars.

**Word Order**

One initial assertion that can be made regarding the possible early insensitivity of verbs towards their appropriate position within a sentence has come from early MLU data. For instance, many naturalistic studies of early language development suggest that rather than generating structure via abstract grammatical generalizations, children may actually be tethering their grammars to individual lexical items with respect to functional elements auxiliaries (Kuczaj & Maratsos, 1993, Pine & Lieven, 1997); determiners (Pine & Lieven, 1997) and pronouns (Pine & Baldwin, 1997). Data on early verb/argument structure (see Radford 1990: pp. 213-17 for stage-1 examples) suggest that early MLU verb classes may not adhere to appropriate SVO argument structure in the sense that target transitive verbs take obligatory object arguments. Radford cites very early two-word structures of the *I/(me) want, Her hit*, type where the direct objects required presence in adult speech go missing. Such deficits might suggest the children’s initial knowledge of verb-argument structure to be developed around individual verbs (and not verb type). In addition, semantic over-extensions of Intransitive verbs of the *Me sleeped teddy* type (=sleeped, slept>put to bed) may likewise show over-extensions on an individual verb basis (Tomasello, 1992), or individual frame-basis (Braine, 1976), but show little evidence that the extension carries over to the entire verb class. In view of these data, it remains questionable whether or not children’s very earliest MLU staged grammar operates with abstract, rule-based representations at all—e.g., [+/-] Verb Transitivity. More specific to English SV(X) word order, some questions regarding rule-based word order parameterization for early MLU speech have been formulated. Atkinson (1992) (following the work of Susan Goldin-Meadow with deaf children and ‘Home Sign’) suggests that there may be no theoretical reason to stipulate for a correct target word order at, say, a pre-parameterized stage of development. If children have an inherent abstract understanding of predicate-argument structure (cf. Valian, 1991), they should then be able to understand the differences between the subject and object of a transitive verb and how to apply this to word order.

Although traditional naturalistic studies have typically shown that correct SVO word order usually appears early on in the data (Brown, Cazden, and Bellugi-Klima (1968), Bloom, 1970; Brown, 1973; Radford, 1990), there is mounting literature to suggest otherwise (e.g., Braine, 1971; Bowerman, 1973; Tsimpili, 1992; Galasso, 1999). Mixed word order data to this effect suggest that there may be a very small window in the chronological development of language that doesn’t reflect target word order—i.e. a pre-SV(X) stage for English. In addition, the fact that early child English seems to provide us with correct word order recognition may be accounted for by means other
than linguistic motivation—e.g., non-linguistic, (and perhaps cognitive-based) sequencing strategies based on formulaic aspects of the input, etc. (cf. Atkinson, 1992). Recall that the ‘U-shape’ learning discussed herein shows how possible surface similarities may actually have very different underpinning structural realizations—e.g., (i) *went* (formulaic) ⇒ (ii) *go-ed* (rule based), ⇒ (iii) *went* (rule insertion). While *went* in (i) and (iii) look identical on the surface, they are actually products of two very different processes. Other various studies on novel/non-sense verbs similarly reveal a small window in the duration of staged speech development that gleams word order errors (Olguin & Tomasello, 1993; Akhtar & Tomasello, 1997). The child’s inability to generalize correct word order to novel verbs suggests that word order, at this early MLU stage, may be learned on a ‘low-scope’ memorization level one verb at a time rather than via a rule-based ‘high-scope’ parameterization process. Thus, it remains unclear whether or not children’s very early MLU speech should be credited with having rule-based processes/parameterizations for determining word order. If not, a special nativist position could still be maintained in the sense that functional parameterization has not yet taken place (cf. Atkinson op. cit). In light of a potential stage-1 non-parameterization account for free word order, strong arguments could be devised suggesting that instances of free word order, in fact, demonstrate the early onset of abstract rules—if we take ‘rules’ here to mean the setting (or non-setting) of parameters. Such arguments would counter the general claims being made (cf. Tomasello, Rowland and Pine (ibid)) that stage-1 is more or less entirely rote-learned. The fact that we do find word order errors call for some level of formal rule abstraction—much in the manner of the U-shape learning model discussed above.

41. Keeping to the spirit of Chomsky’s Minimalist Program regarding Word Order, we would like to maintain Richard Kayne’s proposal that word order is indeed a universal hierarchical property of a Spec>Head>Comp relation. One could perhaps go as far as to make the very strong claim that SVO mirrors cognition, and thus a universal order of Subject-Verb-Object is innately given. In any event, Kayne’s universal constraint is seen as keeping to the spirit of Chomsky’s innateness hypothesis, and so we’ll take it as the null hypothesis here and see where we go with it. However, we can only possibly adhere to it insofar as the empirical data bear it out—and it is here that we instantly run into some difficulties. Mainly speaking, if we want to maintain a universal SVO order, we therefore must do so at that stage of development where the child in fact has access to Double Argument String structures (DAS). For instance, a prior Single Argument String stage (=SAS) would have no way of showing the appropriate Spec and Comp distributions. Well, when looking at a good cross section of child acquisition data, it appears that there is no strong evidence pointing to an exclusive SAS stage—(without some small amount of DASs interceding). While this may be the case, a stage does evidence in the data where at least the majority of utterances are indeed not only SASs, but that such SASs show variable word orders amongst the Subject/Object and Verb—rendering SV, VS, OV, VO orders. It is at this juncture that we have to weaken Kayne’s strong universal claim for an SVO order as correlated to his Linear Correspondence Axiom (LCA), and say that such an axiom only holds for a child at (DAS) stage-2 of development—
again, a stage roughly corresponding with the (albeit optional) emergence of abstract rule formulations and functional categories, both which lead to Parameterization. So in one full sweep, what we have done is somewhat preserve Chomsky’s original version of a word order based on Functional Parameterization (pace Kayne’s strong stance for a non-parameterized word order based on his universal LCA) and have added a further Kaynian stipulation by saying that LCA may only work, rendering all structures as base generated SVO orders only after a pre-cursor parameterization has taken place positioning the Object either Leftward or Rightward of the Verb—now providing two basic universal orders: SVO and (the mirror image) OVS: (of course, the latter order is very rare as a base-generated order, though some have claimed Japanese as an OVS base order, and then, via subject movement, derive an SOV order (fn). In any event, Kayne is explicit in stating that his Head Medial Principle, (stipulating that a Head/Verb must remain in middle position and one of the tenets of his axiom) would conceivably permit the four word orders above to be accessed by a child in a SAS stage-1.

42. Looking at the data (Galasso: 1999), we indeed find a strong correlation between SAS strings and mixed word order alongside DAS strings and fixed order.

Table 5

<table>
<thead>
<tr>
<th></th>
<th>Word Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Files: 8-16 SAS SV VS DAS= SVX / Other</td>
<td></td>
</tr>
<tr>
<td>Age: 2:4-2:8 n.=</td>
<td>87 78 290 5</td>
</tr>
</tbody>
</table>

I. Some token examples include:

(a) SV: Daddy cooking. Him go
(b) OV: Dog kick (= I kick dog). A egg cook. (= I cook egg).
(c) VS: Open me (= I open). Work bike (=Bike works)

In terms of structure, before on the onset of DASs, a Proto XP could be assigned to our SAS stage providing the variable word orderings:

II. K (=XP) vs. III. K (=XP)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>L</td>
</tr>
<tr>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>j</td>
<td>M</td>
</tr>
</tbody>
</table>

a’> spec/ | b’> comp m | a’> spec/ | b’> comp m

head | head

(a = SV, b = OV) (c = VS, d = VO)
In addition to general word order variability, Wh-word order patterns emerge in our early files (age 2;4-3;0) showing semi-formulaic consistencies when examined in light of the general acquisition of complex structure—as mentioned above regarding SAS vs. DAS complexity. Our data evidence a pattern showing Non CSV (Non Comp Subject Verb) ordering which could be interpreted as formulaic in nature. This stage roughly overlaps with our SAS stage mentioned above. Like Kayne on Word order, Cinque (1990) has formulated a strong universal position claiming that all Wh-elements universally position within the Spec-of-CP. Recall, that CP is a functional category that should have a delayed onset time under any maturational theory (cf. Radford: 1990). Here too we need to weaken the strong position by adding the stipulation that in order for this Spec-CP analysis to hold, the subject must simultaneously surface forcing the Wh-element to raise and preposition in Spec-CP. Otherwise, very early (stage-1) Wh-arguments (e.g., What, Who) seemingly get initially misanalyzed as base-generated 3Person Pronoun/Quantifiers placed in superficial subject Spec-VP position. This miscategorization often results in Agreement errors where the Wh-word, seen as incorrectly taking the thematic-role of the subject, agrees (by default) with the verb. Consider the structures of the two following CP-structures below:

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Wh-word order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non CSV</td>
</tr>
<tr>
<td>Files 1-21 n.=</td>
<td>78</td>
</tr>
<tr>
<td>Files 22-25 n.=</td>
<td>120 80</td>
</tr>
</tbody>
</table>

I. VP (Non CSV)  
/     \
| Spec V’ |
/     \
| V D(P) |
| | |
| a). What is these?  
d). Who you see?  
b). Who see him?  
e). What him doing?  
c). What is him?  
f). Where you going? |

II. CP (CSV)  
/     \
| Spec C’ |
/     \
| C VP |

In sum, arguments could be devised suggesting that early Wh-structures are prime examples of semi-formulaic strings base generated (VP *insitu*). A later second stage (or even overlapping stage) may thus be seen as converting formulaic processes into rule driven processes whereby syntactic manifestations of Wh-movement occur with or without Auxiliary inversions. (See Stromswold (§6) above for Non-Aux
inversions). Regarding formulaicity, Pine & Lieven (1997), Pine et al. (1998) claim that a non-rule based account is what is behind the formation of early correct wh-questions (a U-shape learning take on the data). While adopting a constructivist account in explaining the high rate of correctly inverted Wh + Aux combinations, they go on to predicted that correctly inverted questions in the child’s stage-1 data would be produced by those wh + aux combinations that had occurred with high frequency in the child’s input. They go on to specify that there is evidence that the earliest wh-questions produced with an Aux. can be explained with reference to three formulaic patterns that begin with a limited range of wh-word + aux. Combinations (e.g., “whydon’t” you/she) (Rowland & Pine, 2000). Such findings on early formulaic structures parallel what Tomasello (1992) and Newport (op. cit.) suggest regarding an initial stage-1 that reflects a processing deficit tied to functional grammar. In other words, child stage-1 processing which shows a bias toward the modeled high frequency lexical input (vs. rule driven analogy) may arise due to constraints imposed by the low memory bottle-neck of distributional learning (Braine 1987, 1988).

**Lexical Stage-1: A Recap**

45. In light of the above data, and the collections of data elsewhere, it could be argued for our stage-1 that the child’s utterances involve pure projections of thematic argument relations. In Minimalist terms, the operation ‘Merge’ would directly reflect thematic properties and that this operation is innately given by the Language Faculty: Verbs directly theta-mark their arguments as in predicate logic expressions:

<table>
<thead>
<tr>
<th>Token Utterance</th>
<th>Predicate Logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d)addy work</td>
<td>work(d)</td>
</tr>
<tr>
<td>(m)ommy see daddy</td>
<td>see(m,d)</td>
</tr>
</tbody>
</table>

The above Word Order/Syntax includes (SV) and (SVO) patterns and is structured below:

\[
[vP[N Dad][v0][VP[V work]]] \quad [vP[N Mom][v0][VP[V see][N dad]]]
\]

(vP=light-verb Phrase).

46. In both example above, the Nouns (Daddy & Mommy) contain no formal features (such as person or case) and so don’t agree with the verb. The verb likewise carries no Tense or Agreement features. In this sense, theta-marking directly maps onto the semantics of lexical word classes—viz., ‘pure merger’ involves only theta-marked lexical items. It is therefore claimed that there is no Indirect theta-marking capacity at stage-1 such that oblique or prepositional markers would enter into the syntax: for
example, the PP ‘to work’ in Daddy goes to work, would be thematically reduced in the operation Merge as Daddy go work (work = Nouns and not infinitive verb). Such utterances are wide spread for our stage-1 as was revealed in the section above. In addition to seemingly direct thematic based syntax/grammar, numerous other studies have shown that, indeed, children inappropriately overextend semantic (causative) alternations of verbs such as giggle vs. tickle by indiscriminately giving them identical thematic argument structures Thematic role ‘Patient’ in their intransitive forms: e.g., don’t giggle me! vs. don’t tickle me! (Bowerman, 1973). If we wish to make claims that such overgeneralizations are a result of some innate linking rule, then clearly some sort of default semantic-based linking rule must be up for discussion. In any event, the lack of non-semantic [Interpretable] formal features certainly dispels the notion of syntax and leads us to look at such early stage-1 lexical items as being stripped of their formal features, and projecting quasi-semantic information on a class of their own-perhaps to the point that each lexical item is learned and projected in isolation

47. In conjunction to an isolative lexicon, and much in the same spirit with Pine et al. above, Morris et al. (ms 1999) has sketched out a theoretical proposal (based on PDP-style connectionism) that relegates verb-argument structures in children’s stage-1 grammar to individual ‘min-grammars’-that is, each word is learned (‘bottom-up’) in isolation in that there are no overarching abstractions (‘top-down’) that link one verb’s argument structure to another. In other words, there are no argument rules, only isolated word meanings each argument structure is a separate grammar unto itself (p. 6). It is only at a second stage-2 that the child is seen as corresponding the semantic as well as the syntax over from one word to another. For example, the verbs eat and drink, hit and kick, etc. will merge at stage-2 in ways that will project this overarching abstract structure regarding transitivity, thematic structure, etc. Hence, stage-2 is defined as the benchmark in emergence of true syntax and rule formation.

48. In sum, what the above sketch has to offer us is the proposal that children start off (stage-1) with rote-learned items and then strive to find commonalities—the child then builds-up this lexicon from brute memory and only later (stage-2) does she slowly start to form levels of abstraction. The claim is that children learn grammatical relations over time—the bottom-up processes mimic the maturational processes behind language acquisition (viz., first a stage-1 ‘bottom-up’ lexical learning followed by a stage-2 ‘top-down’ rule formation).

49. –Insert Interpretable features Radford here—

50. Distributional Morphology. A second but similar line of reasoning, likewise motivated by outcomes in Chomsky’s Minimalist Program (see Marantz 1995, 1997) calls for morphology to be the all encompassing aspect of grammar—doing away all together with the lexicon as maintained under so call ‘lexicalist hypotheses’, as well as dispensing, to a certain degree, with traditional notions of syntax that sought to
derive a syntactic model outside of the lexicon in a seemingly top-down manner. The theory’s basic core calls on a number of assumptions: viz., (i) that syntactic hierarchical structures ‘resonant all the way down to the word’ (or perhaps more accurately described ‘as being essentially derived from the word’); (ii) that the notion of ‘word’ is broken up into two properties— the word shell of phonology, (or as it is termed in DM, the Idiom), and the word’s selectional morphological features. The distinctions are articulated in terms of morphology by the following labeling: the ‘l’-morpheme—which pertains to the idiomatic aspect of the sound-meaning relation—and the ‘f’-morpheme—which correlates to the abstract morphological features. These two labels may be seen as correlating to Radford’s usage of +/-Interpretable features (above) where the [+Interp] feature distinction pertains to lexical item’s semantic properties (part of which would be the Idiomatic aspect of the word as used in DM, along with its phonological make-up (i.e., ‘l’-morpheme), and where [-Interp] would correlate to the more formal and abstract syntactic properties (i.e., the ‘f’-morpheme). The two-prong theory today is seen as part and parcel of a formal language system. Traditional parts of speech such as ‘Noun’ are redefined as a bundle of features that make-up a single l-morpheme type (called Root). The Noun root or ‘l’-morpheme is defined by how the root entertains certain local relations or governing conditions which it imposes on its complement hosts—e.g., how the Noun root might c-command or license its Determiner (in a local Specifier position) or a ‘Verb (in a local Complement position). A classic example here would be how the same lexical item *Destroy* appears as a ‘noun’ *Destru(ction)* when its nearest adjacent licenser is a Determiner (The destruction), or how the item takes on the role of a verb when its nearest adjacent licensors are Tense/Agreement and Aspect (Destroy-(s), (is) destroy-ing, (have) destroy-ed) (marking Tense, and Participle respectively). This model now places the burden of syntax not with exterior stipulations, but rather with interior conditions that seem to flow up-ward from the lexical items itself and into the relevant projecting phrase. In this new definition (taken right out of MP, ‘Bare Phrase Structure’), the ‘phrase’ is reorganized as simply the sum of the total interacting ‘f’-morpheme parts; the ‘word’ is thus redefined as nothing more than a ‘buddle-of-features’ that project out of the phonological shell. This new analysis will hold a number of consequences for how we come to understand language acquisition. For starters, much of what is being spelled out here concerns a two-stage acquisition of language development and that this dual stage can be accounted for the dual mechanism model as advanced in this paper. What I am on about here can be summarized as follows regarding language acquisition:

(i) Syntax, as understood in Chomsky’s Pre-Minimalist’s terms, may for all intents and purposes reduce to specific bundle-of-features that are encoded in ‘parts-of-speech’ words, (rendering a seemingly bottom-up learning mechanism where ‘meaning’ governs not only how words are learned, but how their syntactic properties project).

(ii) Syntax may no longer be considered as a top-down generator of sentence types, and so words have the capacity to emerge in a early stage of
language merely encoded with ‘l’-morphology or [+Interp] features. In this way, one may be able to define an early stage-1 word as exhibiting more or less only the phonological shell of the word void of its otherwise embedded syntax. If this is indeed the case, a viable maturational story can likewise hold for the onsets of ‘f’-morphology [-Interp] features for the given word. Much in the manner of Roger Brown’s observation leading to a sequence of morphological development (staring with -ing and ending with the Aux. Clitic etc.), a similar story could likewise hold regarding how certain features mature and then merge in a word—a maturation of features however which would not delay the onset of the word in phonological terms (or ‘l’-morpheme values), but would only delay the relevant selectional properties (or ‘f’-morpheme values, etc.) associated with its functional grammar.

The twin notions above would ultimately buttress any theory which would see language development as a maturational interplay of features—as captured here in our discussion of a Converging Theories Model.

51. --Insert DP-analysis stage-1 here—

52. A typical Chomskyan syntactic tree asserts that functional features (features having to do with M(ood), T(ense) and Agr(eement)) are assumed to be projected in a top-down way: these functional features are understood to be what is behind the notion of movement—lexical items move up the tree in order to acquire and check-off these features. The following question certainly could be formulated in Chomskyan terms: ‘why can’t lexical items have such features embedded in their sub-categorical entries, and if they can, what then would motivate movement other than some ad hoc stipulation requiring features to be checked-off in a overall top-down environment’? Chomsky’s interpretations are clear here—some top-down (deductive) measure must be constructed in order to establish a proper rule-driven syntax. Well, this may in all likelihood be correct, but where/what is it in the system that says that the syntax must start out in this way. Consider the tree below (reduced showing only M & T/Agr features):

```
CP
  /\  
C   MoodP
    /\    
  Subj  
    /\    
  M     [T/Agr]P
    /\     
  {do, can,..}  {-s, -ed,..} VP
```
The tree above positions the T/Agr features, along with their specific phrases, as having a top-down representation. If such a tree is completely available early-on in language acquisition—as the Continuity view would maintain—than there should be no reason why a child would exhibit 100% omission of say a top-down Agr feature in the way that would affect only certain words and not others. (When only certain words show individual residual affects, e.g., regarding subcategorization, syntax etc., then a strong claim can be made that the overarching phrase structure is not what is behind the phenomenon, but rather specific lexical-parameterizations may be involved.) (See J. Fodor 1997, Baker 2002 for a seemingly bottom-up treatment of lexical parameterization). In other words, if the structure is in place (from top-down) to deliver the feature of Agr (as with Case), than it would be hard to explain away the fact, if observed in the data, that some words could maintain Case while others (which should maintain Case in the target language) do not. Guasti and Rizzi (2001) say: ‘When a feature is not checked in the overt syntax UG makes it possible to leave its morphological realization fluctuating’. Fine. But, this is seemingly a bottom-up problem. It seems that such optionality would have nothing to do with the phrase (per se). What do we say when the feature itself (as projected from the tree top-down) seems to select some words over others regarding inflection? Surely, if this is a top-down venture, then the features should project onto all verbs (for the appropriate phrase), and not just a select few. But this is in fact what we find at our stage-2 of language development—some words may (optionally) inflect/project the specific feature while others completely by-pass it (entirely).

53. --Insert data from Radford & Galasso here--

54. This gives us the flavor of specific words (and not word classes) taking on functional features (bottom-up). The problem here is how does one maintain the higher-ordinance structure of functional grammar originating from the latter two upper layers of the tree while selecting the functional projection on only a select handful of words. One way around the dilemma may be to suggest that the lexical word itself has part of the (upper-branching) tree embedded in the very lexical item itself (as in subcategorization). In this way, a specific word may reflect a specific functional feature or parameter while another word may not (on a specific lexeme by lexeme basis)—in all actuality, what we are talking about here is that (i) the initial process of the acquisition of functional grammar involves one word at a time (in a bottom-up way), and that (ii) only at a later more developed stage does such feature projection extend to the overall class of words (which then extent to phrases). Following in the spirit of Lexical Parameterization (Borer), Janet Dean Fodor in a similar vain has tentatively suggested in some recent work that parameterization may affect certain words (as in lexical feature specificity) and not others (outside of the scope of its word class) (talk presented at the University of Essex, 1997). One outcome of this would assume that children establish parameter values (perhaps piece-meal) and not grammars as wholes. An example of such bottom-up parameterization or say feature specificity (only selecting [+/-Nom] Case marking here) might then be diagrammed in the following manner:
Treelet Pronoun Parameterization of Case

<table>
<thead>
<tr>
<th>Parameterized</th>
<th>Non-parameterized</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+D(P)/ +Nom]</td>
<td>[+D(P)/ -Nom/default]</td>
</tr>
<tr>
<td>He</td>
<td>her</td>
</tr>
</tbody>
</table>

Such an exclusively bottom-up parameterization method would however obscure correlations often found in the data regarding Case and/or Agreement—such as a seemingly top-down holistic correlation which seeks to link (i) [+Nom] Case if in an agreement relation with a Verbal INFL, (ii) [+Gen] Case if in an agreement relation with a nominal INFL, (iii) Default Case otherwise. It may be that such correlations do come on-line after an initial ‘non-phrase’ parameterization stage—hence, an initial and not fully fledged parameterized stage would meagerly work with individual words, delaying class-parameterization to a slightly later stage.

55. A growing body of research recently undertaken by developmental linguistics suggests that children’s (stage-1) multi-word speech may in fact reflect low-scope lexical specific knowledge rather than abstract categorical-based knowledge. As discussed above, this distinction clearly points to a possible language acquisition processes as proceeding from out of a dual mechanism in the brain. For example, regarding verb inflection, studies (Tomasello & Olguin, Olguin & Tomasello, Pine & Rowland) have shown that the control children have over morphological inflection very early in the multi-word stage is largely individually rote learned—that is, there is no systematic relationship between stem and inflection, nor is there any transfer from ‘supposed’ knowledge of an inflection to other stems. In other words, at the very earliest stages of multi-word speech, there is little or no productively of transferring the knowledge of one verb to another. This may suggest a stage-1 based not on complete paradigm formation, but rather on (semi)-formulaicy.

56. Rowland suggests that a distributional learning mechanism capable of learning and reproducing the lexical-specific patterns that are modeled in the input may be able to account for much of what we find in the early stage-1 data. Input of a high frequency nature will then trigger rote learning associations and patterns that will manifest in the speech production of young children. This notion of rote-learned vs. rule-based or non-systematic vs. systematic behavior (respectively) can be further investigated by looking into what has become known as the *U-shape learning curve*. For instance, indications of systematic (rule-based) behaviors can be seen in overgeneralization. In other words, if overgeneralizations appear with, say, the morphological inflection {s} as in the portmanteau forms for either Verb or Noun—e.g., *I walk-s, feet-s* (respectively), than a sound argument could be made that rules have been employed-albeit, rules which have erroneously over-generated. (In fact, if children in the
process of their early language acquisition are never seen to over-generalize rule-like formations, this is very often a sign of potential Specific Language Impairment (SLI), a result of some neo-cortical brain malfunction which has disturbed the normal syntactic structuring of rules and paradigms.) And so, we rightly extend the argument that if rules are being applied at a given stage, than a rule-based grammar has been activated: Right you say. Well, as it turns out, there are some very interesting findings which suggest that apparent ‘look-a-like’ rules at stage-1 are in fact imposters and don’t really behave as ‘true’ rules.

57. *U-Shaped Learning*. One of the most striking features of language acquisition is the apparent so called *U-shaped Learning Curve* found straddling the two stages of language acquisition. In brief, the U-shaped curve is understood in the following way:

(i) *Inflection*. Children’s earliest Inflected/Derivational word types are, in fact, initially correct—that is, it appears to be the case amongst very early MLU that children have correct formulation of rules. (It goes without saying that typical early MLU utterances indeed have no tense markings to speak of (cf. Wexler & Radford’s Maturational Theory). The point here is that whenever a small sampling of Tense does appear in early MLU speech, it always appears correctly). An example of this is the early emergence in the data of the past tense and participle affixes [ed] and [en] e.g., *talked/gone* (respectively). The initial Past Tense and Plural forms are correct, regardless of whether or not these forms are regular (talked/books) or irregular (went/sheep). However, and what is at the heart of this striking development, it also appears that this initially correct performance stage is then followed by a period of attrition during which the children actually regress—that is, at this slightly later stage in development, they do not only lose some forms of affixation, but in addition, produce incorrect over-generalizations in tense forms (*go>* *goed>* *wented*), and plural forms (*sheeps*), as well as non-inflected tensed forms e.g., *talk-Ø/go-Ø (=past tense)*. To recap, the first occurrence of inflectional overgeneralization roughly at age 2 years that supports a rule-based grammar is preceded by a phase without any errors at all.

(ii) *Phonology*. Similar to what one observes regarding a u-shape grammatical/inflectional development, children also appear to follow a u-shape learning curve with regards to phonology. An example of this is the often cited early productions of e.g., (i) *slept /sΛpt/, cooked /kΛkt/, played /plaε:d/ > to (ii) *sleeped /slεπd/, cooked /kΛkt/, played /plaε:d/ > and back to (iii) *slept /sΛpt/, cooked /kΛkt/, played /plaε:d/ (respectively) completing a U-shaped morpho-phonetic curve yielding **/t//d/ > /d//t/**.

What appears to be good examples of ‘rule-based’ inflection and assimilation in (i) and (ii) (above respectively) is in all actuality nothing more than the product of a ‘parrot-like’ imitation sequence—more akin to iconic pattern
processing derived from stimulus and response learning. The child can be said to engage in segmental, phonetic-based rules only when s/he appears to process the rules yielding an incorrect overgeneralization of past marker {ed} typically pronounced as the default /ld/ which forms the middle-dip portion of the u-shape curve. Recall, in terms of phonology, the child has three allophonic variations to chose from:

a. {ed} => /t/ “walked” /wa:k/ 
b. {ed} => /d/ “played” /ple:d/ 
c. {ed} => /ld/ “wanted” /wantld/

It seems that a default setting with regards to phonology (place & manner of articulation) is minus Comp(lex) where [-Comp] denotes one feature distinction over a two or more features (for instance, bilabials /b/ /m/ would have a [-Comp] feature whereas labio-dentals and inter-dental /l/ /d/ (respectively) would have a [+Comp] since both lip and tooth are involved. In addition, it seems that plus voicing [+V] typically wins out over minus voicing [-V]. By using these default settings, we naturally get voiced plosive /b/ /d/ /g/, nasals /m/ /n/, as our very first sequence of consonants along with [+V] vowels. By taking this default status, the /ld/ should be the allophone of choice, and it often is. In this manner of speaking, adherence to the default setting suggests at least some formation of the rule: defaults work within rule-based paradigms and so should be considered as a quasi-rule-based generation as opposed to a pure imitation sequence.

58. The first two stages of development that form this apparent u-shape curve has been interpreted as manifesting the application of qualitatively different processes in the brain—representing different modes or stages in the course of language acquisition. This u-shaped curve arguably provides some support for our stage-1 to be defined in terms of a formulaic stage rather than as a syntactic and true-rule learning stage. The second up-side of the u-shaped curve is found to coincide with an independent syntactic development—the emergence of a Finiteness marker, and that this finiteness marker only emerges at our functional stage-2 (see Clahsen). In sum, the three stages could be described in the following way:

(i) The first period of the first up-side curve (correct production) correlates with a style of rote-learning. This more primitive mode of learning suggests that the mental lexicon is bootstrapped by mere behaviorist-associative means of learning. In such a rote-learning stage, lexical items (either regular or irregular inflections) are stored in an independent mental lexical heavily based on memorization of formulaic chunks and associations and are processed in a different part of the brain. It is of no surprise that irregular verb past inflection (go>went) out number regular verb past inflection (talk>talk-ed): The former being stored in the lexicon as a formulaic chunk, while the latter indicating the morphological rule formation [V+ {ed}]. Hence, our dual converging theories model postulates for a sharp contrast and disassociation between regular vs. irregular inflection. This seemingly early correct production is therefore due to a low-scope, phonological ‘one-to-one & sound-to-meaning’ relationship with
no relevance to rules. Hence, our formulaic past tense inflection is not realized as \([{\text{stem}} + {\text{affix}}] \{\text{talk-}\{\text{ed}\}\}\), but rather as one unanalyzable chunk \([\text{talked}]\).

(ii) The second stage then marks the onset of a rule process (albeit, not necessarily the mastery of it). Here, the child is seen as letting go with the formulaic lexical representation in favor of rule formations: i.e., patterns of concatenate stems appear along side inflectional affixes. Thus, irregular forms often get over-generalized with the application of the rule resulting in e.g., \([\text{goed}/\text{wented}/\text{sheeps}]\). This overgeneralization stage maps onto a chronological functional categorical stage of language acquisition where rule-based mechanisms are becoming operative. Thus, the over-generalized up-swing of the u-shaped curve is linked to children’s syntactic development: over-generalization of inflection appears when the child ceases using bare-stems (as in stage-1) to refer to past events.

(iii) The third and final stage marks the second up-side swing of the u-shaped curve and represents the correct target grammar.

59. It is thus proposed that this tri-staged learning process—from correct to incorrect to correct again—can more properly be accounted for by a dual learning mechanism in the brain: (i) an initial mechanism that has no bearing on rules and is pinned to a type of process best suited for more behavioristic associative learning, such as base lexical learning, irregular verb learning, lexical redundancy formations, etc.

Brain Related Studies

60. Much of the theory behind a dual model of language has become buttressed by recent developments in Brain Related studies. There is now an ongoing stream of data coming in that tells us the brain does indeed process different linguistic input in very different ways. Some of the first analyses using fMRI (functional Magnetic Resonance Imaging), and other brain-related measures show that irregular inflection processes \((\text{go} > \text{went})\) seem to be located and produced in the temporal lobe/motor strip area of the brain, a processing area strictly associated with basic word learning referred to as the lexical component, or Lexicon). On the other hand, regular inflection processes e.g., \((\text{stop} > \text{stopped})\), where the rule \([\text{stem}]+[\text{affix}]\) is applied, point to areas of the brain which generate rule formations, i.e., the computational component. In other words, there seems to be a clear indication that the two types of linguistic processes are dissociated. This same disassociation seems to project between how one processes derivational morphology—here, being equated to irregular and/or whole lexical word retrieval—and inflectional morphology.

61. Wakefield and Wilcox (=W&W) (1994: 643-654) have recently concluded that a discontinuity theory—along the lines proposed by Radford—may have an actual physiological reality as based on a biological ‘maturation’ of brain development. Their work consists of two segments: the first being a theory of the relationship
between certain aspects of brain maturation and certain transitions in grammatical representation during the course of language acquisition, the second being a preliminary investigation to access the validity of the theory by testing some of the specific hypothesis that it generates. In their model, it is the left posterior aspect of the brain, at the junction of the parietal, occipital, and temporal lobes (POT) that generates semantically relevant, modality-free mental representations by allowing signals from all neocortically-represented sensory modalities to converge in a single processing region. In turn, the linguistically relevant contributions of Broca’s area, located in the inferior portion of the left frontal lobe imparts abstract structure to those representations with which it interacts—including (functional) grammatical components as well as the semantic components. The idea here is that we can now tentatively spot functional abstract grammar within the frontal lobe areas of the brain, and show how such grammatical aspects relate to the more primitive, prosaic elements of lexical-semantics (as spotted in the temporal lobe regions). The trick here is to see if the two regions are initially talking to one another (as in neuro-connectivity), say at our grammatical stage-1. Using PET/ERP-language studies, a sketchy two-prong picture emerges suggesting that the neural mechanism(s) involved split along lexical and functional grammatical stages of language development. It is clear that Broca’s area is involved not only with the generation of abstract hierarchical structure, but, with the representation of lexical items belonging to functional categories. However, the studies reveal that in order for Broca’s area to work at this highly abstract level of representation, the frontal lobe which houses Broca’s area must also connect to the POT region of the brain—in this sense, a real conversation must be carried out between the (first order) semantic properties of language (POT) and their functional counterparts. This relationship parallels the lexical-functional dichotomy found in all language.

62. The W&W study suggests that the maturational development of language follows from brain development—and can be summarized below:

a. The lexical stage-1 of language acquisition naturally arises from a disconnect between the more primitive POT (temporal-lobe/lexical-grammar region) and the hierarchical Broca’s area (frontal-lobe/functional grammar).

b. This disconnect has to do with the biological development of myelination in the bundle of axons that connect the two areas together. Myelination of axons is then said to mature at roughly that chronological stage where we find a lexical (staged) grammar merging with a functional (staged) grammar.

c. With respect to the brain/language relationships in the child, it is important to recognize that during the period of time typically associated with the initial stages of language acquisition, the brain is still in a relatively immature state. Neural plasticity begins with the sensory motor-strip temporal area (POT), and then proceeds to move to secondary areas (Broca’s area) related to the frontal lobe region.
Conclusion: A Converging Theories Model

63. In the history of all pursuit of science, it has traditionally been the case that science precedes and develops via different methods and theories. Converging approaches always strive to expose inherent weakness in their opposing theories. It goes without saying that convergence methods go far in peeling away biased assumptions which often lead to half-correct assertions. Taking what is good from one theory and throwing away what is not is just common-sense science. For example, on one ‘converging; hand, Chomsky has asserted that syntax is the result of the creative human brain set-up in such a way as to manipulate true-rules. It creates, from nothing external to itself, the structure of language. (See special nativism above). In restricting ourselves to the point at hand, Chomsky has assimilated much of his arguments from the long line of rational philosophy and has converged such reasoning into how he believes an autonomous language structure (internal) might be construed. His belief that syntax is autonomous directly paves a way for him to distinguish between species-specific (human/hominid) language and other modes of cognitive-based primitive communications (animal/pongid). His now famous debates—first between Skinner (Behaviorism) and later with Piaget (Constructivism)—can be readily reduced back to Converging Methodologies between (philosophy and cognition) which sought to return language to seventeenth century nativist assumptions. Later, he would go on to extend such arguments to fight off pure pragmatic/socio-linguistic pursuits of linguistics—saving the study of language from becoming strictly a ‘humanities’ field of study which emphasized social phenomena with little if any analytical worth: (cf. Quine, Rorty pace Chomsky). Taking his notion of an autonomous syntax further, the natural next step to take would be to say that all other aspects of language (whatever they may be) that can’t fall under this autonomous rule-based syntactic realm might be conversely tethered to both behaviorism and associationism as part of an underlying cognitive mechanism. Chomsky has himself expressed the possibility that general mundane concepts—many of which contain inherent sub-categorial features that are extremely convoluted and abstract, yet from which we go on to readily attach labels (=words)—may be preconceived and innate: however, he goes on to suggest that such conceptual innateness may be tethered to cognition as a universal ability to get at meaning (Chomsky 2000: p.61-62):

These conceptual structures appear to yield semantic connections of a kind that will, in particular, induce an analytic-synthetic distinction, as a matter of empirical fact.

These elements (he cites concepts such as locational nature, goal, source of action, object moved, etc.) enter widely into lexical structure… and are one aspect of cognitive development.
On one hand, what Chomsky seems to be saying is that (i) Functional Grammar, or Syntax (par excellence) is autonomous and disassociated from all other aspects of the mind/brain-including meaning and/or cognition. Thus, syntax is created from out of the mind’s creative and independent eye (with all aforementioned nativist trappings). However, and to the point of this section, Chomsky doesn’t hesitate to attribute those non-syntactic aspects of language, say word learning (based on frequency learning and associationism), to cognition. This, I believe, goes to the heart of the matter—namely, that a form of converging theories has been evoked here and could be summarized as follows:

Chomsky and Converging Theories

1. Syntax proper (labeled herein as Functional Grammar) is creatively formed by a true-rule process via an innately given Language Acquisition Device (LAD) (more recently called the Language Faculty)—comprising of initial grammatical default settings of which are called Universal Grammar. For example, this is where the more abstract Inflectional rules are housed: the functional features of number/person/case/agreement/tense e.g. Plural \([N+s]\), Past Tense \([V+\{ed\}]\), etc. Of course, the ‘Wugs-Test’ of Berko goes directly under this category. Meaning is detached from syntax.

2. Word learning (labeled herein as Lexical Grammar) is formed via a one-to-one iconic association between sound and meaning. This process of both word learning on (i) a phonological level, and word learning on (ii) a semantic/conceptual level, is more akin to past behavioristic notions of learning. Very young children (at our stage-1) may exploit and over-extend such processes—this is apparently what we find regarding formulaic type utterances, Irregular Verb/Noun lexical learning and retrieval, as well as Derivational morphology.

Connectionism

In view of Chomsky’s assertion that Syntax is autonomous, there can be by definition no primitive lower-level capacities at work in syntax—namely, nothing that hinges on perception, sound, object movement, spatio-temporal, etc. Although we share with our primates such low-scope abilities, more than anything else, it is our ability to work with abstract rules which creates the unsurpassable, and ever widening gap between human language and animal communication—the former based on true-rules & syntax, the latter based on more primitive behavioristic modes of learning. Regarding the higher-level processes having to do with syntax/grammar, the bootstrapping problem as discussed above does provide a way for lower-level processes associated with connectionism to serve as a springboard for later rule-based grammar. For instance, it is now widely assumed (cf. Plunkett, Elman, among others) that something like a connectionist system must provide the neurological foundations for the apparent symbolic mind. In other words, a symbol processing system might sit on top of a connectionist implementation of the
neurological system. Such an heteroarchical layered approach to language would be similar to stating that in order to talk about Darwinian Biology, one must first acknowledge the underlying universals of Physics. However, having said this, and more to the point of Chomsky’s reference to autonomous syntax, a symbol processing system would operate according to its own set of principles. Recently, the notion of hidden units/rules providing crucial feedback loops in connectionist processors have been interpreted (much to the chagrin and potential demise of the pure connectionist group) as a form of a quasi innate symbolic devise--cleverly hidden in the actual architecture itself. (See the on-going debates between Marcus vs. Elman, Elman vs. Marcus on this). Nonetheless, it is now becoming commonly accepted in connectionist circles that a number of local architectural constraints are indeed necessary in order to bring about a sufficiently qualitative approximation of computation worthy of language: constraints such as the right number of units (hidden and overt), layers, types of connections etc. Notwithstanding camp rhetoric and inevitable spin involve—again, arguments tantamount to the old nature vs. nurture debate—there however may be something to the notion that such hidden units serve as a bridge between the two systems (and for that matter, the two schools of thought). Moreover, there is a certain degree of truth to the analogy stating hidden unit tabulations spawn symbolic rule paradigms. From this dualistic approach, it is possible to tentatively sketch out some shape of what a “Converging Theories Model might look like in the face of such aforementioned assertions:

66. A Converging Theories Model: Lexical vs. Functional Grammar

1. A Lexical Based Grammar:
--The Lexical categories and their Phrase Projections:
   Nouns, Adjective, Verbs, Adverbs
--The Lexical Projections are based upon a thematic template and do not have language specific parameterization or [-Interp] feature spell-outs-hence, e.g., non-target or absent Verbal Inflection (Tense/Agreement), Nominal Case, and Word Order.
--High saliency value of semantic material:
   Low-scope processes based on cognition/semantics

--Examples: Lexical learning and labeling, lexical redundancy rules, grammatico-semantics, irregular verbs based on sound patterning, ‘one-to-one’ sound-meaning relation.
--Theory: General Nativism:
   Skinner’s associationism, Piaget’s constructivism
2. A Functional Based Grammar:
--The Functional categories:
  Determiner, Quantifier, Auxiliary/Modal Complementiser
--INFLection:
  Tense, Agreement, Case
--Minimum saliency value of semantics:
  Upper-scope processes based on rule-driven syntax

--Examples: (True) Rule-based learning based on variable algorithm, paradigm construction, syntax, ‘one-to-many’ abstract relation.

--Theory: Special Nativism:
  Chomsky’s autonomous syntax  Wexler/Radford’s maturation theory

67. The above lexical categories are substantive in meaning and are akin to more behavioristic processes such as rote-learning, formulaic mimics, frequency-based memorization as discussed in a Piaget-style cognitive learning model. The above functional categories are non-substantive in meaning and are akin to true autonomous syntactic theory (Chomsky), and the creative ability to carry out preconceived rules with novel items (Brown, Berko). From a neurological standpoint however, Chomsky’s idea of an autonomous syntax is treated as a quintessential impossibility—the obvious being that—there must be a syntax positioned in the brain, and thus buttressed by any or all cognitive apparatus. Our Converging Theory Model as sketched above partially gives this argument to the neuro-science position by claiming that indeed one aspect of language is tethered to a low-scope cognitive apparatus. Where we beg to differ is that “higher-scope” rule-driven processes are not cognitive bound and rather rely on their own set of structure dependency conditions to survive—a structure dependency that has no bearing on the exterior cognitive realm. This analogy follows in the wake of Steven Pinker’s hypothesis stating that a dual mechanism of language is at work in language acquisition—(i) one based on cognitive universal in relation to the lexical component of the brain (say, our stage-1 child grammar), and (ii) one based on true-rule formation related to the computational component of the brain (say, our stage-2 child grammar). (Recall our brief discussion on Radford (§19ff) in defining a functional stage-2 grammar based on pure syntactic (high-scope) [+/-Interpretable features).

68-69 A Final note
70. References (In Preparation)