

Structure Dependence, the Rational Learner, and Putnam's 'Sane Person'

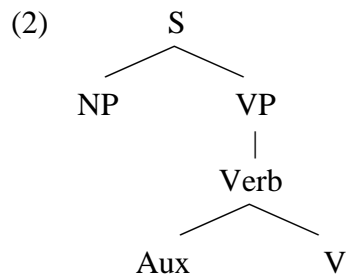
Howard Lasnik and Juan Uriagereka
lasnik@umd.edu Juan@umd.edu

University of Maryland

Part A: Where Does Structure Dependence Come From?

I. The Classic Aux-Inversion Paradigm and Analysis

- | | | |
|------|-------------------------------|-------------------------------|
| (1)a | John would like it | Would John like it |
| b | John has seen it | Has John seen it |
| c | John is reading it | Is John reading it |
| d | John has been writing it | Has John been writing it |
| e | John could be swimming | Could John be swimming |
| f | John could have laughed | Could John have laughed |
| g | John could have been sleeping | Could John have been sleeping |
| h | John likes it | Does John like it |



- (3) Aux → C (Modal) (have en) (be ing)

- (4) C → $\left\{ \begin{array}{l} S \text{ in the context } NP_{\text{sing} -} \\ \emptyset \text{ in other contexts} \\ \textit{past} \text{ in any context} \end{array} \right\}$

- (5) T_q - optional

Structural analysis:
 $\left\{ \begin{array}{l} NP - C - V... \\ NP - C+M - ... \\ NP - C+have - \\ NP - C+be - ... \end{array} \right\} \dots$

Structural change:
 $X_1 - X_2 - X_3 \rightarrow X_2 - X_1 - X_3$

- (6) "Affix Hopping" (informal statement)
Attach an affix (S, Ø, past, en, ing) to a v (V, M, have, be) immediately to its right.
- (7) "do Support" (informal statement)
Attach *do* to an affix that failed to hop.

This sort of analysis directly captures the felt relatedness among the pairs in (1). It also provides a natural description of the examples in the second column of (1) and succeeds in using the same mechanism for the superficially very different (1)h.

II. Fundamental Properties of the Analysis

- (8) "... there are languages (in our general sense) that cannot be described in terms of phrase structure, but I do not know whether or not English is itself literally outside the range of such analysis. However, I think that there are other grounds for rejecting the theory of phrase structure as inadequate for the purposes of linguistic description. The strongest possible proof of the inadequacy of a linguistic theory is to show that it literally cannot apply to some natural language. A weaker, but perfectly sufficient demonstration of inadequacy would be to show that the theory can apply only clumsily; that is, to show that any grammar that can be constructed in terms of this theory will be extremely complex, ad hoc, and 'unrevealing', that certain very simple ways of describing grammatical sentences cannot be accommodated within the associated forms of grammar, and that certain fundamental formal properties of natural language cannot be utilized to simplify grammars." Chomsky (1957, 34)
- (9) "We can greatly simplify the description of English and gain new and important insight into its formal structure if we limit the direct description in terms of phrase structure to a kernel of basic sentences (simple, declarative, active, with no complex verb or noun phrases), deriving all other sentences from these (more properly, from the strings that underlie them) by transformation, possibly repeated." Chomsky (1957, 106-107)
- (10) "The simplest class of sentences not included in our grammatical sketch in §§72.2, 72.3 is the class of interrogative sentences taking a yes-or-no answer. For any sentence of the form $NP - VP_A - VP_I$ (hence any sentence derived in §72.2), we can form a corresponding question by inverting the NP and an initial segment of the VP_A (auxiliary verb phrase). Thus such a question will be of the form $X \wedge NP \wedge Y \wedge VP_I$ where $X \wedge Y$ is the VP_A of the corresponding declarative." Chomsky (1955, 418) [This assumes a slightly different Σ, F grammar than the one in Chomsky (1957), but not in any way relevant to the issue at hand.]
- (11) "It turns out ... that the treatment of "do" as an element automatically introduced to carry an unaffixed affix will have a considerable simplifying effect on the grammar. The ... treatment ... permits a uniform treatment of "do," "does," and "did," as the bearer of a displaced affix." Chomsky (1955, 419)

- (12) "The transformational analysis ... is clearly somewhat simpler than the extension of the kernel grammar ... even in absolute terms as these analyses now stand. Thus we must accept the transformational analysis of yes-or-no questions, and drop them from the kernel. The difference in complexity between these two analyses would become more compelling if we could show that [the crucial parts of the analysis] are needed anyway, for other transformations. But this is in fact the case ...

12 I saw the play and so did he
I will see the play and so will he
I have seen the play and so has he
I have been seeing the play and so has he" Chomsky (1955, 423)

- (13) "... given the transformational analysis of yes-or-no questions, sentences of the form 12 can be introduced transformationally in a very simple way, with no independent characterization necessary... In other words, we have found that two phenomena which on the level of phrase structure are distinct and complex become, in transformational terms, instances of a single generalization. But this naturally leads us to choose the transformational analysis in this case and, in particular, to drop yes-or-no questions from the kernel.

Investigation of sentences like
22 I saw the play and he did too
etc., gives additional support to this analysis." Chomsky (1955, 425)

III. Structure Dependence (and Poverty of Stimulus)

- (14) "A theory that attributes possession of certain linguistic universals to a language-acquisition system, as a property to be realized under appropriate external implies that only certain kinds of symbolic systems can be acquired and used as languages by this device. Others should be beyond its language-acquisition capacity... In principle, one might try to determine whether invented systems that fail these conditions do pose inordinately difficult problems for language learning, and do fall beyond the domain for which the language acquisition system is designed. As a concrete example, consider the fact that, according to the theory of transformational grammar, only certain kinds of formal operations on strings can appear in grammars - operations that, furthermore, have no a priori justification. For example, the permitted operations cannot be shown in any sense to be the most "simple" or "elementary" ones that might be invented. In fact, what might in general be considered "elementary operations" on strings do not qualify as grammatical transformations at all, while many of the operations that do qualify are far from elementary, in any general sense. Specifically, grammatical transformations are necessarily "structure-dependent" in that they manipulate substrings only in terms of their assignment to categories. Thus it is possible to formulate a transformation that can insert all or part of the Auxiliary Verb to the left of a Noun Phrase that precedes it, independently of what the length or internal complexity of the strings belonging to these categories may be. It is impossible, however, to formulate as a transformation such a simple operation as reflection of an arbitrary string (that is, replacement of any string a_1 ...

a_n , where each a_i is a single symbol, by $a_n \dots a_1$), or interchange of the $(2^n - 1)$ th word with the 2^n th word throughout a string of arbitrary length, or insertion of a symbol in the middle of a string of even length. Similarly, if the structural analyses that define transformations are restricted to Boolean conditions on *Analyzability*, as suggested later, it will be impossible to formulate many "structure-dependent" operations as transformations – for example, an operation that will iterate a symbol that is the left-most member of a category (impossible, short of listing all categories of the grammar in the structural analysis), or an operation that will iterate a symbol that belongs to as many rightmost as leftmost categories). Hence, one who proposes this theory would have to predict that although a language might form interrogatives, for example, by interchanging the order of certain categories (as in English), it could not form interrogatives by reflection, or interchange of odd and even words, or insertion of a marker in the middle of the sentence. Many other such predictions, none of them at all obvious in any a priori sense, can be deduced from any sufficiently explicit theory of linguistic universals that is attributed to a language-acquisition device as an intrinsic property."
Chomsky (1965, 55-56)

- (15) "Consider, for example, the simple fact that grammatical transformations are invariably *structure-dependent* in the sense that they apply to a string of words [fn. More properly, to a string of minimal linguistic units that may or may not be words.] by virtue of the organization of these words into phrases. It is easy to imagine structure-independent operations that apply to a string of elements quite independently of its abstract structure as a system of phrases. For example, the rule that forms the interrogatives of **71** from the corresponding declaratives of **72** (see note 10 [I should emphasize that when I speak of a sentence as derived by transformation from another sentence, I am speaking loosely and inaccurately. What I should say is that the structure associated with the first sentence is derived from the structure underlying the second.]) is a structure-dependent rule interchanging a noun phrase with the first element of the auxiliary.
- 71 a. Will the members of the audience who enjoyed the play stand?
 b. Has Mary lived in Princeton?
 c. Will the subjects who will act as controls be paid?
- 72 a. The members of the audience who enjoyed the play will stand.
 b. Mary has lived in Princeton.
 c. The subjects who will act as controls will be paid.
- (16) In contrast, consider the operation that inverts the first and last words of a sentence, or that arranges the words of a sentence in increasing length in terms of phonetic segments ("alphabetizing" in some specified way for items of the same length), or that moves the left-most occurrence of the word "will" to the extreme left – call these O_1 , O_2 , and O_3 , respectively. Applying O_1 to **72a**, we derive **73a**; applying O_2 to **72b**, we derive **73b**; applying O_3 to **72c**, we derive **73c**:
- 73 a. stand the members of the audience who enjoyed the play will
 b. in has lived Mary Princeton
 c. will the subjects who act as controls will be paid
- The operations O_1 , O_2 , and O_3 are structure-independent. Innumerable other operations of this sort can be specified.

- (17) There is no a priori reason why human language should make use exclusively of structure-dependent operations, such as English interrogation, instead of structure-independent operations, such as O_1 , O_2 , and O_3 . One can hardly argue that the latter are more "complex" in some absolute sense; nor can they be shown to be more productive of ambiguity or more harmful to communicative efficiency. Yet no human language contains structure-independent operations among (or replacing) the structure-dependent grammatical transformations. The language-learner knows that the operation that gives **71** is a possible candidate for a grammar, whereas O_1 , O_2 , and O_3 , and any operations like them, need not be considered as tentative hypotheses.

If we establish the proper "psychic distance" from such elementary and commonplace phenomena as these, we will see that they really pose some nontrivial problems for human psychology. We can speculate about the reason for the reliance on structure-dependent operations [fn. See G. A. Miller and N. Chomsky, "Finitary Models of Language Users, Part II," in R. D. Luce, R. Bush, and E. Galanter, eds., *Handbook of Mathematical Psychology*, Vol. 2 (New York: Wiley, 1963), for some proposals regarding this matter.], but we must recognize that any such speculation must involve assumptions regarding human cognitive capacities that are by no means obvious or necessary. And it is difficult to avoid the conclusion that whatever its function may be, the reliance on structure-dependent operations must be predetermined for the language-learner by a restrictive initial schematism of some sort that directs his attempts to acquire linguistic competence." Chomsky (1968, 61-63)

- (18) "Consider, for example, the way in which questions are formed in English. Consider the sentence "The dog in the corner is hungry." From this, we can form the question "Is the dog in the corner hungry?" by a simple formal operation: moving the element "is" to the front of the sentence. Given a variety of examples of question formation, a linguist studying English might propose several possible rules of question formation. Imagine two such proposals. The first states that to form a question, we first identify the subject noun phrase of the sentence, and we then move the occurrence of "is" following this noun phrase to the beginning of the sentence. Thus in the example in question, the subject noun phrase is "the dog in the corner"; we form the question by moving the occurrence of "is" that follows it to the front of the sentence. Let us call this operation a "structure-dependent operation," meaning by this that the operation considers not merely the sequence of elements that constitute the sentence but also their structure; in this case, the fact that the sequence "the dog in the corner" is a phrase, furthermore a noun phrase. For the case in question, we might also have proposed a "structure-independent operation": namely, take the leftmost occurrence of "is" and move it to the front of the sentence. We can easily determine that the correct rule is the structure-dependent operation. Thus if we have the sentence "The dog that is in the corner is hungry," we do not apply the proposed structure-independent operation, forming the question "Is the dog that – in the corner is hungry?" Rather, we apply the structure-dependent operation, first locating the noun-phrase subject "the dog that is in the corner," then inverting the occurrence of "is" that follows it, forming: "Is the dog that is in the corner – hungry?"

- (19) Though the example is trivial, the result is nonetheless surprising, from a certain point of view. Notice that the structure-dependent operation has no advantages from the point of view of communicative efficiency or "simplicity." If we were, let us say, designing a language for formal manipulations by a computer, we would certainly prefer structure-independent operations. These are far simpler to carry out, since it is only necessary to scan the words of the sentence, paying no attention to the structures into which they enter, structures that are not marked physically in the sentence at all. Mathematicians have studied structure-independent operations on strings (inversion, shuffling, etc.), but it has occurred to no one to investigate the curious and complex notion of "structure-dependent operation," in the relevant sense. Notice further that we have very little evidence, in our normal experience, that the structure dependent operation is the correct one. It is quite possible for a person to go through life without having heard any relevant examples that would choose between the two principles. It is, however, safe to predict that a child who has had no such evidence would unerringly apply the structure-dependent operation the first time he attempts to form the question corresponding to the assertion "The dog that is in the corner is hungry." Though children make certain kinds of errors in the course of language learning, I am sure that none make the error of forming the question "Is the dog that in the corner is hungry?" despite the slim evidence of experience and the simplicity of the structure-independent rule. Furthermore, all known formal operations in the grammar of English, or of any other language, are structure-dependent. This is a very simple example of an invariant principle of language, what might be called a formal linguistic universal or a principle of universal grammar."

Chomsky (1972, 26-28)

- (20) "Consider the process of formation of simple yes-or-no questions in English. We have such declarative-question pairs as (1):
(1) The man is here. – Is the man here?
The man will leave. – Will the man leave?
Consider the following two hypotheses put forth to account for this infinite class of pairs:
 H_1 : Process the declarative from beginning to end (left to right), word by word, until reaching the first occurrence of the words *is*, *will*, etc.; transpose this occurrence to the beginning (left), forming the associated interrogative.
 H_2 : same as H_1 , but select the first occurrence of *is*, *will*, etc., following the first noun phrase of the declarative.
Let us refer to H_1 as a "structure-independent rule" and H_2 as a "structure-dependent rule." Thus, H_1 requires analysis of the declarative into just a sequence of words, whereas H_2 requires an analysis into successive words and also abstract phrases such as "noun phrase." The phrases are "abstract" in that their boundaries and labeling are not in general physically marked in any way; rather, they are mental constructions.
- (21) A scientist observing English speakers, given such data as (1) would naturally select hypothesis H_1 over the far more complex hypothesis H_2 , which postulates abstract mental processing of a nontrivial sort beyond H_1 . Similarly, given such data as (1) it is reasonable to assume that an "unstructured" child would assume that H_1 is valid. In fact, as we know, it is not, and H_2 is (more nearly) correct. Thus consider the data of (2):

(2) The man who is here is tall. – Is the man who is here tall?

The man who is tall will leave. – Will the man who is tall leave?

These data are predicted by H_2 and refute H_1 , which would predict rather the interrogatives (3):

(3) Is the man who here is tall?

Is the man who tall will leave?

- (22) Now the question that arises is this: how does a child know that H_2 is correct (nearly), while H_1 is false? It is surely not the case that he first hits on H_1 (as a neutral scientist would) and then is forced to reject it on the basis of data such as (2). No child is taught the relevant facts. Children make many errors in language learning, but none such as (3), prior to appropriate training or evidence. A person might go through much or all his life without ever having been exposed to relevant evidence, but he will nevertheless unerringly employ H_2 , never H_1 on the first relevant occasion (assuming that he can handle the structures at all). We cannot, it seems, explain the preference for H_2 on grounds of communicative efficiency or the like. Nor do there appear to be relevant analogies of other than the most superficial and uninformative sort in other cognitive domains. If humans were differently designed, they would acquire a grammar that incorporates H_1 and would be none the worse for that. In fact, it would be difficult to know, by mere passive observation of a person's total linguistic performance, whether he was using H_1 or H_2 .
- (23) Such observations suggest that it is a property of S_0 – that is of $LT(H,L)$ – that rules (or rules of some specific category, identifiable on quite general grounds by some genetically determined mechanism) are structure-dependent. The child need not consider H_1 ; it is ruled out by properties of his initial mental state, S_0 . Although this example is very simple, almost trivial, it illustrates the general problem that arises when we attend to the properties of attained cognitive states." Chomsky (1980b, 39-40)

IV. Structure Dependence and Rational Learners Perfors et al. (2006)

- (24) "The Poverty of the Stimulus (PoS) argument holds that children do not receive enough evidence to infer the existence of core aspects of language, such as the dependence of linguistic rules on hierarchical phrase structure. We reevaluate one version of this argument with a Bayesian model of grammar induction, and show that a rational learner without any initial language-specific biases could learn this dependency given typical child-directed input. This choice enables the learner to master aspects of syntax, such as the auxiliary fronting rule in interrogative formation, even without having heard directly relevant data (e.g., interrogatives containing an auxiliary in a relative clause in the subject NP)."
- (25) "The phenomenon of auxiliary fronting in interrogative sentences is one example of the PoS argument; here, the argument states that children must be innately biased to favor structure-dependent rules that operate using grammatical constructs like phrases and clauses over structure-independent rules that operate only on the sequence of words. English interrogatives are formed from declaratives by fronting the main clause auxiliary. Given a declarative sentence like "The dog in the corner is hungry", the interrogative is formed by moving the is to make the sentence "Is the dog in the corner hungry?"

Chomsky considered two types of operation that can explain auxiliary fronting (Chomsky, 1965, 1971). The simplest (linear) rule is independent of the hierarchical phrase structure of the sentence: take the leftmost (first) occurrence of the auxiliary in the sentence and move it to the beginning. The structure-dependent (hierarchical) rule – move the auxiliary from the main clause of the sentence – is more complex since it operates over a sentence's phrasal structure and not just its sequence of elements."

- (26) "The "poverty" part of this form of the PoS argument claims that children do not see the data they would need to in order to rule out the structure-independent (linear) hypothesis. An example of such data would be an interrogative sentence such as "Is the man who is hungry ordering dinner?". In this sentence, the main clause auxiliary is fronted in spite of the existence of another auxiliary that would come first in the corresponding declarative sentence. Chomsky argued that this type of data is not accessible in child speech, maintaining that "it is quite possible for a person to go through life without having heard any of the relevant examples that would choose between the two principles" (Chomsky, 1971)."
- (27) "In this work we present a Bayesian account of linguistic structure learning in order to engage with the PoS argument on its own terms – taking the existence of structure seriously and asking whether and to what extent knowledge of that structure can be inferred by a rational statistical learner. This is an ideal learnability analysis: our question is not whether a learner without innate language-specific biases must be able infer that linguistic structure is hierarchical, but rather whether it is possible to make that inference."
- (28) "(1) We demonstrate that a learner equipped with the capacity to explicitly represent both hierarchical and linear grammars – but without any initial biases – could infer that the hierarchical grammar is a better fit to typical child-directed input. (2) We show that inferring this hierarchical grammar results in the mastery of aspects of auxiliary fronting, even if no direct evidence is available. (3) Our approach provides a clear and objectively sensible metric of simplicity, as well as a way to explore what sort of data and how much is required to make these hierarchical generalizations. And (4) our results suggest that PoS arguments are sensible only when phenomena are considered as part of a linguistic system, rather than taken in isolation."
- (29) **The corpus**
"The corpus consists of the sentences spoken by adults in the Adam corpus (Brown, 1973) in the CHILDES database (MacWhinney, 2000). In order to focus on grammar learning rather than lexical acquisition, each word is replaced by its syntactic category. Ungrammatical sentences and the most grammatically complex sentence types are removed. The final corpus contains 21792 individual sentence tokens corresponding to 2338 unique sentence types out of 25876 tokens in the original corpus."
- (30) **The grammars**
"Because this work is motivated by the distinction between rules operating over linear and hierarchical representations, we would like to compare grammars that differ structurally. The hierarchical grammar is context-free, since CFGs generate parse trees with

hierarchical structure and are accepted as a reasonable “first approximation” to the grammars of natural language (Chomsky, 1959)."

- (31) "All grammars are probabilistic, meaning that each production is associated with a probability and the probability of any given parse is the product of the probabilities of the productions involved in the derivation."
- (32) "The probabilistic context-free grammar (PCFG) is the most linguistically accurate grammar we could devise that could parse all of the forms in the corpus: as such, it contains the syntactic structures that modern linguists employ, such as noun and verb phrases."

(Two of the target grammars are presented in Appendix A.)

- (33) Note that both of the hypotheses in (20), H_1 and H_2 , present rules that apply to a sentence, deforming its internal structure in some way (to be precise, the rules apply to the abstract structures underlying sentences, but we may put this refinement aside). Both the structure-independent rule H_1 and the structure-dependent rule H_2 make use of the concepts "sentence," "word," "first," and others; they differ in that H_2 requires in addition an analysis of the sentence into abstract phrases. A rule that does not modify the internal structure of a sentence is neither structure-dependent nor structure-independent. For example, a phrase structure rule, part of a phrase structure grammar in the technical sense of the term, is neither structure-dependent nor structure-independent.

Is (33) Juan and Howard's response to Perfors et al. (2006)? Well, it could be. But it was actually Chomsky's response to Putnam (1980). (Putnam was presenting another sort of counter to Chomsky's poverty of stimulus argument.) Here's more of the discussion:

- (34) " H_1 has never been "put forth" by anyone, nor would any sane person put it forth ..."
Putnam (1980, 287)
- (35) "Putnam considers my two hypotheses H_1 and H_2 , advanced to explain the formation of yes-or-no questions in English. He observes that the structure-independent rule H_1 would not be put forth by any "sane person," which is quite true, but merely constitutes part of the problem to be solved. The question is: Why? The answer that I suggest is that the general principles of transformational grammar belong to S_0^L as part of a schematism that characterizes "possible human languages.""
Chomsky (1980a, 311)

Part B: Where Does Structure Come From?

- (36) Implicit in the discussion about “emergent systems” is that structure “is there”, so that learners reflect it.
- (37) But how did it “get there?”
- (38) Suppose that a CFG is, indeed, a very effective way to compress information that would be very clumsily describable in FSA terms (and cf. Berwick (1982)).

- (39) So one might want to say that the data “out there” evolved in such a way that if it somehow coalesced into a CFG format, it was in the end easier to learn; so if humans evolved to learn language easily, language had to co-evolve to be learned easily too (cf. Briscoe (1998) and Deacon (1998)).
- (40) First problem with that: E-language. Chomsky has long ago shown how it is totally unclear what it means for something to be a language in those terms, if it has the properties of human language. It leads to virtual paradox.
- (41) But suppose we ignore that (huge) problem to go on with the discussion at face value. *Even in E-language terms*, Chomsky showed us that it is not enough with a CFG.
- (42) You may think: well, it is enough with a CFG *plus some clever coding* (cf. the “linear-indexed grammars” of Weir (1988)).
- (43) For the point we're trying to make, that doesn't matter. Call that a CFG+. It is *not* a CFG with standard first-order symbolic manipulation. You need to index, type-lift, or, well, go into context-sensitive transformations, “mild” as they may be (that's a separate matter, cf. SPH).
- (44) A crucial point in the alternative to Chomsky's argument: whatever “emergent systems” there are, they are *not specific to language*. So a priori it is better to go into an alternative explanation, instead of having to be cornered into positing language-specific devices.
- (45) There are phenomena in nature that naturally fit CFG terms. Take the Fibonacci structures that Lindemayer systems allow (with the trivial rewrite rules $0 \rightarrow 1$, $1 \rightarrow 0, 1$).
- (46) These are found not just in natural morphology; they appear also in behaviors. Much of Gallistel's work on animal cognition (e.g., Gallistel (In press)) can be understood as claiming that at least CFG description is needed for some animal conducts.
- (47) Note: animal conducts, *not animal communication systems*.
- (48) All the communication systems we are familiar with, even elicited ones, do not obviously go beyond FSA conditions (bees, birds, bats, cetaceans, primates).
- (49) Moreover, Gallistel hasn't been able to produce an animal behavior that demands what we called CFG+ conditions.
- (50) For example, that could be shown with animals capable of generalized quantification (not enough to have weaker forms of quantification, not even exact numerosity); if that's too tough, how about “cross-serial” dependencies?
- (51) Possibility: no animal behavior exhibits CFG+ conditions – except the human language faculty.
- (52) Note: there *might* be room for CFG+ conditions in nature more generally (cf., with caution, the work of Searls (2002) on RNA pseudoknots or protein folding, where he shows “cross-serial dependencies”, and more generally Niels Jerne's admonition to biologists in his 1984 Nobel Prize address...)

- (53) But CFG+ conditions only made it to human cognition...
- (54) Or to be fair and specific: to human *language*.
- (55) Other behaviors where one might see the need for CFG+ conditions (math, music, complex planning, etc.) could "easily" all be based on extensions of human syntax.
- (56) So now the question is where all of that leaves (44) above.
- (57) We have CFG+ conditions that are allegedly best seen as general cognition, all purpose informational, or some such thing. But they only appear in the language faculty!
- (58) What, then, is the gain of saying that language has co-evolved these properties with human learners?
- (59) Why did it evolve THOSE properties in the case of human language, but not in other animal behaviors?
- (60) To defuse this argument: just as Chomsky did half a century ago for language, show us unmistakable CFG+ conditions in animal behaviors, or even in aspects of human cognition that do not correlate with language.

Appendix A: 2 of the "structure-dependent" grammars of Perfors et al. (2006) (the least and most comprehensive ones). (Taken from <http://www.mit.edu/~perfors/cogsci06/archive.html>)

Level 1 CFG

S --> wh C	I2 --> NP
S --> wh I	I2 --> AP
S --> NP I	AP --> adj
S --> I2	NP --> det N
S --> wh	NP --> N
S --> co	NP --> NP C
C --> aux I2	NP --> pro
I --> aux I2	N --> adj N
	N --> n

Level 6 CFG

S --> wh AP	I --> aux I2	T --> to VI
S --> wh PP	I --> aux	T --> to
S --> wh NP	I2 --> I2 PP	NP --> NP T
S --> wh IP	I2 --> T	NP --> NP PP
S --> wh C	I2 --> NP	NP --> det N
S --> wh I	I2 --> VI	NP --> N
S --> aux IP	I2 --> VP	NP --> NP CP
S --> aux VI	I2 --> AP	NP --> NP VP
S --> NP V	I2 --> PP	NP --> NP C
S --> NP I	AP --> adj T	NP --> pro
S --> I	AP --> adj	NP --> prop
S --> I2	V --> V NP	N --> adj N
S --> V	V --> V PP	N --> n
S --> CP	V --> V T	PP --> prep PP
S --> wh	V --> V AP	PP --> prep VP
S --> T	V --> v	PP --> prep
S --> co	VI --> vi	PP --> prep NP
CP --> comp C	VI --> VI NP	
IP --> NP I2	VI --> VI PP	
IP --> NP C	VI --> VI T	
IP --> AP C	VI --> VI AP	
C --> aux IP	VP --> part	
C --> aux I2	VP --> VP NP	
C --> V	VP --> VP PP	
C --> NP V	VP --> VP AP	
C --> NP I	VP --> VP T	

Appendix B: Structure Dependency Lives

Subject/Aux Inversion may not be the best way to demonstrate structure-dependency in transformations, given how limited that process is to root-contexts and how unclear its nature is. But the point can be illustrated with scores of other examples in the literature. This appendix shows a sample (see Lasnik and Uriagereka (2005)) for a fuller presentation).

(B1) *Classical Formulation of Transformations* (Emonds (1976))

- (a) Core Transformations:
- i. 'Triggered'.
 - ii. Can be 'unbounded'.
 - iii. 'Structure preserving'.
 - iv. May have scopal effects.

- (b) Stylistic Transformations:
- i. Not 'triggered'.
 - ii. Extremely local.
 - iii. Sensitive to phono
 - iv. Have 'surface' effects.

(B2) *Minimalist Formulation*

- (a) Agree-based Transformations
- i. Obey the LRC.
 - ii. Obey the MLC.
 - iii. Obey the CCU.
 - iv. May have scopal effects.

- (b) Operations without Agree:
- i. Not subject to the LRC.
 - ii. Take place within phases.
 - iii. Sensitive to borders.
 - iv. Have 'surface' effects.

(B3) Structure Preserving Hypothesis (SPH)

No core transformational rule can involve positions X and Y if X and Y do not share property P.

(B4) *Improper Movement*:

*[Who [t seems [t (that) [t to love Mary]]]]
 ^ _ _ / ^ _ _ * _ _ / ^ _ _ _ _ _ /

- | | | | | | | | | |
|------|----|-------------------|----|---------------|----|-------------------|----|----------------|
| (B5) | a. | X' | b. | XP | c. | X' | d. | XP |
| | | / \ | | / \ | | / \ | | / \ |
| | | X ...Y... | | YP X' | | X ...YP... | | Y X' |
| | | /\ / | | ^ / \ | | /\ / | | ^ / \ |
| | | Y X / | | X ...YP... | | YP X / | | X ...Y... |
| | | ^ _ _ / | | _ _ _ _ / | | ^ _ * _ / | | _ _ * _ _ / |
| | | picture-collector | | | | *all-sorts-of- | | |
| | | | | the man's car | | pictures-of-cars- | | *the's car man |
| | | | | collector | | collector | | |

References

- Berwick, Robert. 1982. *Locality principles and the acquisition of syntactic knowledge*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, Mass.
- Briscoe, Ted. 1998. Language as a complex adaptive system: co-evolution of language and of the language acquisition device. In *Proceedings of the 8th Meeting of Computational Linguistics in the Netherlands*, 3-40. Rodopi, Nijmegen.
- Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms. Harvard University, Cambridge, Mass. and MIT, Cambridge, Mass. [Revised 1956 version published in part by Plenum, New York, 1975; University of Chicago Press, Chicago, 1985].
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT press.
- Chomsky, Noam. 1968. *Language and mind*. New York: Harcourt Brace Jovanovich. Extended edition, 1972
- Chomsky, Noam. 1972. *Problems of knowledge and freedom: The Russell lectures*. New York: Random House.
- Chomsky, Noam. 1980a. Discussion of Putnam's comments. In *Language and learning: The debate between Jean Piaget and Noam Chomsky*, ed. Massimo Piatelli-Palmarini. Cambridge, Mass.: Harvard University Press.
- Chomsky, Noam. 1980b. On cognitive structures and their development: A reply to Piaget. In *Language and learning: The debate between Jean Piaget and Noam Chomsky*, ed. Massimo Piatelli-Palmarini, 35-52. Cambridge, MA: Harvard University Press.
- Deacon, Terrence. 1998. *The symbolic species: The co-evolution of language and the brain*. New York: Norton.
- Emonds, Joseph. 1976. *A transformational approach to English syntax*. New York: Academic Press.
- Gallistel, Charles R. In press. The foundational abstractions. In *Of minds and language: The Basque country encounter with Noam Chomsky*, ed. Massimo Piatelli-Palmarini, Juan Uriagereka, and Pello Salaburu. Oxford: Oxford University Press.
- Lasnik, Howard and Juan Uriagereka (with Cedric Boeckx). 2005. *A course in minimalist syntax: Foundations and prospects*. Oxford: Blackwell.
- Perfors, Amy, Joshua B Tenenbaum and Terry Regier. 2006. Poverty of the Stimulus? A rational approach. In *Twenty-Eighth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum, Vancouver, Canada.
- Putnam, Hilary. 1980. What is innate and why: Comments on the debate. In *Language and learning: The debate between Jean Piaget and Noam Chomsky*, ed. Massimo Piatelli-Palmarini, 287-309. Cambridge, MA: Harvard University Press.
- Searls, David. 2002. The language of genes. *Nature* 420: 211-217.
- Weir, David J. 1988. *Characterizing mildly context-sensitive grammar formalisms*. Doctoral dissertation, University of Pennsylvania, Philadelphia.