DESPERATELY EVOLVING SYNTAX?

JUAN URIAGEREKA
Linguistics, University of Maryland at College Park,
College Park, MD 20742, USA

The Chomsky Hierarchy gives us a first approximation as to where human syntax lies: the relevant generating device accepting appropriate languages should be slightly more powerful than a standard Push-Down Automaton (a PDA+), although for familiar reasons not much more so, yielding mere 'mild' context-sensitivity. An evolutionary study of syntax ought to give us some clues as to how a PDA+ could have emerged. The present paper explores this in relation to the Procedural Deficit Hypothesis (PDH) on the role of the *FOXP2* in the regulation of procedural memory, which is taken to be a pre-requisite for the workings of a PDA+ embodied in an animal brain. It is argued that while this gene is probably involved in the evolution of human syntax, in all likelihood it got recruited from the action of a different (set of) genes, which should be responsible for its sudden emergence. The goal is to provide an approach that is informed both about contemporary considerations in the Evo-Devo paradigm and standard results in the study of syntax.

1. Syntactic Boundary Conditions

To examine the evolution of any system one has to make implicit or explicit assumptions about how it works. Syntax is one such system, and in this talk I will assume some major results from its theoretical study. Among these:

(1)
a. Syntactic dependencies arrange themselves in terms of formal objects that can be quite *high within the Chomsky Hierarchy* (specifically, context-sensitive).
b. Context-sensitive dependencies are generally *triggered* (or 'last resort'), *structure-dependent* (or 'uniform'), and limited by *locality* (e.g. 'minimality') considerations.
c. Semantic dependencies are determined by syntactic dependencies and obey the mapping principles of *Full Interpretation*, *Compositionality*, and *Conservativity*.
d. Morphological variation, of the sort patent across languages, in many instances involves *uninterpretable* formatives.
e. Core language acquisition involves the fixation of a few fixed, normally morphological, *syntactic options* ('parameters').

Observe that (1a) and (1b) are purely structural specifications, while (2c) and (2d) are interface conditions, at least the latter one of which has a significant bearing on matters of language acquisition in a naturalistic setting. There may, of course, be further properties that syntax has as a system, but for the broad purposes of evolution this should suffice –and seems relatively uncontroversial.

Let's briefly consider each of these conditions in turn, going from the last to the first. (1e) is perhaps the least established among the results in (1), as no full theory of realistic language acquisition has been developed to date. What seems clear is that an alternative model of acquisition that equates children as hypothesis testers is unworkable, while a broadly 'parametric' approach (and I include in this rubric a variety of Markedness proposals in the market) at least

has a prayer. The important point for us here is that the options the child faces are considerably limited and, for them to be minimally plausible, of a very ostensive sort, possibly even just morpho-phonemic and/or lexical variants. While this model of acquisition is far from established, it is, at least at the level of abstraction that I am presenting it, 'the only game in town'.

The rest of the assumptions in (1) should be even less hard to accept, starting with the factual claim in (1d). One only has to look at the distribution of, say, accusative Case features in all sorts of thematic dependents of a verb to isolate clearly uninterpretable features. These, in a system that is seen to emerge at the interface with interpretive components, are deeply surprising, even somewhat troubling if one takes a minimalist take on the language faculty.

While (1c) is less surprising than (1d) –inasmuch as it asserts the full interpretability of linguistic symbols with semantic import– it is nonetheless remarkable that the particular form of systematicity that the language faculty takes should be of the compositionality sort. This actually goes to an extreme in the case of binary quantifiers –that is, elements that relate an arbitrarily complex restriction and an even more complex (and in many instances syntactically unrelated) scope. There is no logical way to guarantee that natural language quantifiers should be conservative, meaning by this that they relate to their restriction in a tighter, more direct way, than they do to their scope. In other words, the simply observable property, to this date never violated in examples from all languages studied, must be a consequence of how the system is articulated, not its effable conditions or broad cognitive considerations.

It is because of conditions of the sort in (1c) and (1d) that the claim in (1a) is standardly made, qualified as in (1b). To understand this claim in the necessary detail, it is useful to remind ourselves that levels within the Chomsky Hierarchy of automata invoke degrees of systemic memory deployment: from its absence in the less structured (finite-state) layer to unlimited capabilities at the other extreme (the Turing machine). In a very precise sense human syntax falls in between context-free and context-sensitive relations. The former are the least one needs to express compositional dependencies, impossible to express in full generality at simpler the finite-state level. The latter are required to express long-distance morphological dependencies and the conservative property of quantifiers. It should be emphasized that, within minimally realistic assumptions about cognition, this is a formal result, thus not subject to rational discussion.

Imagine one implements the brand of context-sensitive relations found in language by way of Weir's (199*) "linear indexed" grammars. This issue emerges because of conditions of the sort alluded to in (1b): linguistically viable long-distance relations are limited, and Weir's mechanism formalizes such a fact. This approach, in effect, packs some relevant information about syntactic structures in phrasal non-terminals, which are then operated on at a lower level of the Chomsky Hierarchy. Now the overall form of the grammar is still high enough within the hierarchy for its formulas not to be accepted by a Push-Down Automaton (PDA). Similar considerations apply to other forms of coding the relevant relations: type lifters, slash categories, threads, traces, copies. For our purposes all of these involve what we may characterize as a PDA+ automaton, requiring more derivational memory than a PDA –so that internal manipulations within phrases are permitted– but less than the next one in the series –so that they are not generalized beyond observable limits.

The importance of this exercise is to establish a minimal system for

evolutionary theorists to model the evolution of. If put in computational terms, the automaton that must have evolved for accepting relevant linguistic structures is of the PDA+ sort. This could have taken place in various ways, depending on what one takes formal syntax to be modeling. In my view the most straightforward interpretation is the one advocated by Marantz (2005), in his revamping the old Derivational Theory of Complexity (DTC) within the broad confines of the Minimalist Program (MP). But the least that should be modeled is *some* PDA+ properties in *any* theoretical framework one cares to explore.

2. A Brief Critique of Standard Approaches

My reading of the relevant evolutionary literature, admittedly as an outsider, does not fulfill that minimal requirement just posed. For example, the classic adaptationist Pinker and Bloom (1990) virtually ignores syntax, even as understood at the time of its writing. But this piece shows considerably more expertise on the topic than what is shown by notable authors like Michael Arbib, Terence Deacon or even Phillip Lieberman. These are all major players, but I am still to see any detailed discussion in their works of the sorts of facts (1) outlines, presupposing a PDA+ architecture of grammar.

In fairness, a second group of authors does get closer to the concerns just discussed. Thus, for instance, Carstairs-McCarthy (1999) takes phrasal syntax to be an outgrowth of syllabification, while for Calvin and Bickerton (2000) it is an exaptation from pre-existing thematic relations. Unfortunately, it is unclear how either approach bears on whatever goes beyond such local dependencies. As far as I can see such stories would remain unchanged even if the language faculty did not present discontinuous dependencies, or they obeyed conditions that are the inverse of what actually holds (for instance, anti-c-command as opposed to c-command).

Other relevant pieces are designed computationally. Thus Kirby (2000) explores the learners' capacity to segment utterances and generalize over chance coincidences in their meanings of identical segments. Again, such a system does not go beyond phrasal associations and the meaning compositionality thereof, at best reaching PDA capabilities. A similar approach is taken by Hurford (2000), still emphasizing broad generalizations and regularizations, or Nowak et al. (2002), which shows –also through modeling– that beyond a threshold of active word usage it is advantageous for a system to deploy phrases. I believe we knew that important fact already, in formal terms discussed in Berwick (1982).

Again, constituent structure is a step in gearing syntactic complexity; what would be impressive is to show how entangled (discontinuous) constituents emerge. Given the Chomsky Hierarchy, the following step –a transition to a PDA+– *presupposes* the PDA describing mere constituents, and so an evolutionary change taking an organism into PDA+ territory automatically carries it over intermediate realms. In other words, we don't know whether syntax got to be PDA+ directly or through PDA stages, as these studies assume. Of course, it may well be that a PDA+ automaton is an even more effective way to compress the sort of dependencies that arise with linguistic symbols. But if so, why should the grammar go through a "mere" PDA stage?

That is important because, implicit in the discussion about "emergent systems" is that structure "is there", so that learners reflect it. But the question is how it "got there". A PDA could, indeed, effectively compress information that

would be only clumsily describable in terms of a finite-state automaton (FSA). And certainly if such a result and the corresponding system is *not specific to language*, a priori it is better to go into an explanation of this sort, instead of having to be cornered into positing language-specific devices. But facts show us that what the language faculty requires is a PDA+ automaton –and unfortunately familiar communication systems, even elicited ones, do not obviously go beyond FSA conditions, let alone PDA ones. So if all there were to language structuring is some clever packing that compactness favors, why should we be the only lucky species that stumbled onto that wondrous feat of general cognition?

In my view this is where things stand, with one exception I return to. The evolution of syntax as in (1) won't be modeled by ignoring syntax, or by asserting that it should follow from effective packing or similar considerations. The latter claim would move professionals if we were shown how any of the basic properties we routinely work with do indeed emerge, in detail –and moreover why they are unique to this system. Short of this, the sensation many of us get is that researchers are attempting to desperately evolve syntax.

The third problem arises from the fact that it's life we're studying. It may be advantageous for an organism, as modeled in a computer, to use PDA resources, or even fly. It is a subtler task to show how it gets the biological means to do so.

Finally, a third line of research develops Hauser et al. (2002), a programmatic piece sensitive to the matters raised here. In short, it conceptually divides those aspects of language that humans share with other species (the Broad Faculty of Language, FLB) from those that are unique, if any exist (the Narrow Faculty of Language, FLN). The piece furthermore raises the empirical challenge of deciding which features of syntax fall within each. For example, *recursion* is seen as a property of information-exchange systems that has not, so far, been encountered in non-human species, and so is hypothetically declared part of FLN. In contrast, some phonetic and semantic features are shown to have been spotted, in some form, in other creatures, and are thus taken to be part of FLB. The paper suggests that syntax may be part of FLN.

Hauser et al. has raised much discussion, even controversy. Jackendoff and Pinker (2005) attack it on the basis that there is more that is unique to language than recursion, and that the model framing the evolutionary picture –nothing but the dominant paradigm– is not of their liking. To defuse such criticisms we must extend our coverage.

Any evolution-of-syntax theory must at least give an indication of how PDA+ conditions arose. This is especially so because no other animals use a comparable device for information exchange.

Hauser et al (2002) conceptually separates those aspects of language that humans share with other species (the Broad Faculty of Language, FLB) from those that are unique, if any (the Narrow Faculty of Language, FLN). The

article raises the empirical challenge of deciding which specific features of syntax fall within FLB or FLN, offering instances of each. For example, *recursion* is seen as a property of information-exchange systems that has not been encountered in non-human species, and so is hypothetically declared part of FLN. In contrast, many phonetic and semantic features are shown to have been spotted in other creatures, and are thus taken to be part of FLB. The spirit of this paper is that syntax –the customary object that professionals work with– is part of FLN. So this is a natural framework to pursue the question above.

We also assume that evolutionary studies must be grounded on archeology and supporting disciplines. In addition, we may need to reflect on the very structure of evolutionary theory, itself an evolving science.

Language evolution leaves no direct fossils, yet it's interesting to consider human behavior from the perspective of its computational limitations. Uniquely anthropological behaviors exhibiting computational conditions found in language may actually be related to it. Camps and Uriagereka (2006) argue that knot-tying, a specifically human ability, presupposes a context-sensitive mechanism manipulating a string. If so, it must be parasitic on the linguistic system or a reflex of some deeper ability underlying both. Studies show a significant overlap between Developmental Coordination Disorder –one of whose manifestations is knotting difficulties– and Specific Language Impairment (Hill 2001). Considerations like these lead Pierpont and Ullman (2005) to postulate a common underlying syndrome for relevant co-morbid instances, based on their Procedural Deficit Hypothesis (PDH). This theory gives a central role to so-called procedural memory, of a kind that psychologists find related to ruled-governed behaviors.

Coolidge and Wynn (2004) suggest that the two known sub-species of *homo sapiens* differ in the sort of memory that determines rule-governed behavior, an argument that can be focused in computational terms, via knots – which Neanderthal didn't seem to exhibit. Our species is also genetically distinct from Neanderthals, having separately evolved in Africa, from where it migrated elsewhere carrying the same underlying syntax seen throughout the world's languages. This is significant because at least one gene has been identified that seems central to the language faculty (Lai et al (2002)): *FOXP2*, the human allele of a regulating gene that mutated to its present form within the last 200,000 years (Enard et al (2002)). All indications so far are that Neanderthals didn't have the *FOXP2* allele, which could be very relevant to their kind of mind. Though then the question is how the language faculty can suddenly bloom, into the system it must have been already in the Upper Paleolithic, from its virtual non-existence 100,000 years prior.

Enigmas for evolutionary theory arise with any system that emerges rapidly, such as Adaptive Immunity (AI), which appeared abruptly in the cartilaginous fish –though not in the sister lamprey lineage. Key to AI is a kind of memory, which preserves both the recognition of signal sequences and the *RAG* proteins, responsible for initiating the recombination gene segments of the antigen-recognizing receptors. Piattelli-Palmarini and Uriagereka (2004) argue that the accepted theory of how AI evolved, via the horizontal insertion of a (viral in origin?) *transposon*, is a model story for systemic emergence. Aside from pertinent to some of the properties seen in syntax (e.g. the memory it requires, which resembles AI conditions), thinking that way addresses a

difficulty for linguistic emergence, emphasized by Lewontin (1990). New traits need not lead to more offspring, even if advantageous to the individual showing them; a linguistic trait may be detrimental for an individual, against the norm of its group. The AI logic suggests that *an entire sub-group* may have shown the relevant trait, epidemically.

FOXP2 doesn't seem to have been horizontal in origin, although this gene may have been recruited as a result of a brain reorganization precipitated by some earlier, more drastic, mutation. Curiously, expression of the avian variant of the gene, in an area connecting their acquisition and production brain circuits, increases during the critical early age at which the young male acquires its song, which differs regionally and seasonally (Scharff and White 2004). Moreover, the mRNA of the gene *sub*-regulates in this area as adult males sing to themselves, effectively practicing variants of the song, while when the action is female-directed it *up*-regulates (Teramitsu and White (2006)). A parsing strategy could be crucial at both the acquisition and production stages, an issue arising for any complex structure that has to be 'squeezed out' of a brain, into a unidimensional channel like chirping –or speech. Given Ullman and Pierpont's PDH, it is possible that FoxP2 might help regulate a memory window. If it narrows variations are expected, if it widens more structural components will be held in the memory buffer.

The Evo-Devo project shows how species arise from gene arrays, invoked without any simple correlate in the phenotype. FoxP regulating genes are found all the way up to yeast, performing various tasks, in different tissues; but what demanded the *FOXP2* recruitment, apparently for procedural memory regulation? Solving that mystery relates to criticism of Hauser et al. (2002), which used recursion as a (good) example of FLN. While recursive thought may have preceded FLN, its *use* in an information exchange system seems uniquely human. As Hurford (2000) notes, self-embedding –a defining property of recursion– poses special processing problems (online storage cannot easily distinguish several tokens of the same processed category). According to both Bickerton (2000) and Carstairs-McCarthy (2000), the shift to language came with a steep rise in signal processing capacity, which may have addressed the successful identification, in on-line terms, of complex recursion. The last mutation in *FOXP2* may touch on that, as individuals with mutations in it lack the ability to process complex recursive structures (Vargha-Khadem et al. (1995)).

All of that may perhaps seem to some like attempting to 'desperately evolve syntax'. But such an admittedly drastic line of reasoning may well be necessary, for the reasons discussed –indeed more generally for other systemic emergences too. In any case, one should seriously examine what alternatives are available in the market, with comparable data coverage, including the essential tenets of syntax. That is, after all, what we seek to evolve, to start with.

References and other Relevant Works by the Author

Bickerton, D. 2000. "How protolanguage became language." In *The Evolutionary Emergence of Language*, eds. Chris Knight, Michael Studdert-Kennedy, and James R. Hurford, 264-284. Cambridge: CUP.

Camps, Marta, and Juan Uriagereka. 2006. "The Gordian Knot of Linguistic Fossils." In *The Biolinguistic Turn*, eds. J. Martin and J. Rosselló, 34--65. Publications of the University of Barcelona.

Carstairs-McCarthy, Andrew. 2000. "The distinction between sentences and noun phrases: An impediment to language evolution?" In *The evolutionary emergence of language:Social function and the origins of linguistic form*, eds. C. Knight, M. Studdert-Kennedy and J.R. Hurford, 248-63. Cambridge: CUP.

Enard, Wolfgang, Molly Przeworski, Simon E. Fisher, Cecilia S. L. Lai and Victor Wiebe. 2002 "Molecular evolution of FOXP2, a gene involved in speech and language." *Nature* 418.6900, 869-872.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. "The faculty of language." *Science* 198.5598: 1569–1579.

Hill, Elisabeth L. 2001. "Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments." *International Journal of Language and Communication Disorders* 36 (April): 149-171.

Hurford, J. 2000. "The Emergence of Syntax." In *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*, eds. C.

Knight, M. Studdert-Kennedy and J. Hurford, 219-230. Cambridge: CUP.

Lai, Cecilia S. L., Simon E. Fisher, Jane A. Hurst, Faraneh Vargha-Khadem and Anthony

P. Monaco. 2001. "A forkhead-domain gene is mutated in a severe speech and language disorder." *Nature* 413.6855: 519-523.

Lewontin, R. C. 1990. "How much did the brain have to change for speech?" *Behavioral and Brain Sciences* 13.4: 740-741. Piattelli-Palmarini, M., and J. Uriagereka. 2004. "The Immune syntax." In *Universals and Variation in Biolinguistics*, ed., L. Jenkins, 341-377. London: Elsevier.

Piattelli-Palmarini, M., and J. Uriagereka. 2005. "The Evolution of the Narrow Faculty of Language: The skeptical view and a Reasonable Conjecture." In in Denis Delfitto, Giorgio Graffi and Sergio Scalise (eds.), *Lingue e Linguaggio*, IV-1, 27-79, Special issue on Language Evolution.

Piattelli-Palmarini, Massimo and Juan Uriagereka. Forthcoming. "Why Knots, Finches and More: Extending our Data Base." In *Biolinguistics: Language Evolution and Variation*, ed. Anna Maria di Sciullo. Cambridge, MA: MIT Press.

Piattelli-Palmarini, Massimo and Juan Uriagereka. In progress. "Still a Bridge too Far? A Biologinguistics Project for Grounding Language on the Brain." For *Physics of Life Reviews.*

Piattelli-Palmarini, Massimo, Pello Salaburu and Juan Uriagereka, eds. 2008. *Of Minds and Language: A Conversation with Noam Chomsky*, New York: OUP.

Scharff, Constance, and Stephanie A. White. 2004. "Genetic Components of Vocal Learning." In *Behavioral Neurobiology of Birdsong*, eds. H. Philip Zeigler and Peter Marler, 325–347. New York: New York Academy of Sciences.

Teramitsu, Ikuko, and Stephanie A. White. 2006. "FoxP2 Regulation during Undirected Singing in Adult Songbirds." *The Journal of Neuroscience* 26 (July): 7390-7394.

Ullman, Michael T., and Elizabeth I. Pierpont. 2005. **"**Specific Language

Impairment is not specific to language: The Procedural Deficit Hypothesis." *Cortex* 41.3: 399-433.
Uriagereka, Juan. 1998. *Rhyme and Reason*. Cambridge, MA: MIT Press.
Uriagereka, Juan. 2007. "What can finches tell us about linguistic evolution?" *Seed Magazine*, New York, July 2007.
Uriagereka, Juan. Forthcoming. "Evolution of Syntax." In Patrick Hogan (ed.), *The Cambridge Encyclopedia of the Language Sciences.*
Vargha-Khadem, Faraneh, Kate Watkins, Katie Alcock, Paul Fletcher and Richard Passingham. 1995. "Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder." *Proceedings of the National Academy of Sciences* 92, 930--933.
Wynn, Thomas, and Frederick L. Coolidge. 2004**. "**The expert Neanderthal mind." *Journal of Human Evolution* 46 (April): 467-487.