Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews ••• (••••) •••–•••

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Comment

Evolutionary basis for human language
Comment on "Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition" by Tecumseh Fitch

Mark Steedman

Received 2 June 2014; accepted 4 June 2014

Communicated by L. Perlovsky

I will argue, primarily, that you cannot learn a language whose terms express semantic properties not expressed by the terms of some language that you are already able to use.—Jerry Fodor, *The language of thought* [10, p. 61]

Fitch [8] advances the "Dendrophilia Hypothesis" that "humans have a multi-domain capacity and proclivity to infer tree structures from strings, to a degree that is difficult or impossible for most non-human animal species" as an evolutionary basis for human language acquisition and musical ability. The present article argues that, to the extent that this is true, the trees in question are essentially semantic in nature, and are obtained not from strings, but by grace of a capacity shared with some animals for deliberatively planning action in the world. The reason humans have language of a kind that animals do not, is because of a qualitative difference in the nature of human plans.[1]

1. The language faculty

What do we know about human language? Two points that are often remarked are that languages appear to be very diverse in form, but that human children can nevertheless acquire any of them in roughly the same amount of time in interaction with their speakers. Moreover, no other animal, not even those closest to ourselves in evolutionary terms, appears to use or be able to acquire any comparable system. Since the divergence of the human line from that of the chimpanzees occurred only around 6M years ago, and the onset of human language may have been much more recent than that, its evolution seems to have been very rapid indeed, despite its singular nature.[2]

Two kinds of explanation for the rapid evolutionary development of language seem to be on offer.[3]

DOI of original article: http://dx.doi.org/10.1016/j.plrev.2014.04.005.

[1] Thanks to Frank Keller and Ron Petrick. The work was supported by ERC Advanced Fellowship 249520 GRAMPLUS and EC FP7 IP grant 270273 Xperience.

[2] There is circumstantial evidence of this rapid onset in the comparative maladaptation of the human vocal tract for its original dual purposes of breathing and swallowing, relative to that in other primates.

[3] Explanations of a third kind, appealing to universal grammar and asserting the singular nature of language and its divergence from other forms of animal cognition and communication in terms of a language instinct or organ, via notions like emergence and saltation, and the biolinguistic nature

http://dx.doi.org/10.1016/j.plrev.2014.06.010

The first kind of explanation is based on the idea that the tree-like structures that are characteristic of human language, and which have been claimed to distinguish it from the kind of symbol systems that animals actually *can* learn, predate the existence of language itself in evolutionary terms on a much larger timescale. The evolution of language itself is then presumably seen as a comparatively simple attachment of word-like concepts to a preexisting structural component. Jespersen [18], following Darwin [4], seems to have believed something like this when he fancied the first utterances of speech to have been "like something between the nightly love-lyrics of puss upon the tiles and the melodious love-songs of the nightingale" [18, p. 434].

Much recent work on structure and learning of birdsong, including Fitch's own, seems to have this aim of showing that systems of communication including recursive or treelike structure can evolve independently of human language, and therefore offer a possible precursor to it (e.g. [17,14]).

The other kind of explanation is based on the assumption that there is a preexisting non-linguistic, non-communicative, but nevertheless symbolic homologous conceptual representation, onto which articulated communicative language can be hung via its semantics. Since individual languages differ considerable not only in their word order, but in the way they view the world semantically, as partially revealed by their morpholexicons, that universal conceptual representation must be independent of linear order, and of a degree of elaboration which will support all of those different world views. (In other words, different grammatical devices across languages may correspond to different but overlapping substructures of this conceptual representation.) Early exponents of this view were Pinker [26] and (in very different terms), Wexler and Culicover [32].

One version of the idea is implicit in the suggestion of Hauser, Chomsky and Fitch [13, p. 1578] that structural recursion may have evolved for non-linguistic purposes of navigation and "number quantification", and might possibly be shared by other animals.

These two styles of explanation at first seem very different. The first emphasizes the primacy of structure in evolution and acquisition. It seems to depend on the assumption that unsupervised learning of grammars from strings alone is possible (an assumption which seems implicit in Fitch's statement in Section 6.3 of the Supra-Regular Hypothesis).

The latter assumption is questionable. It may be possible to learn finite inventories of candidate phones and morpheme boundaries from mere exposure to native speech (although even that is not trivial). However, learning natural language grammars from exposure to nothing but the strings of the language is an open problem in computational linguistics on which very little progress has been made, though there have been many attempts over many years.

The second style of explanation emphasizes the primacy of cognition. It has the advantage that we know that it works. If we assume that something homologous to the semantics is available to the child, then we are talking about *supervised* learning, analogous to the induction of parsers from treebanks, which is the way that all currently successful wide coverage parsers are induced [2,3].[4]

The child's task of learning a grammar from strings and meaning representations in a homologous "language of the mind" is of course a little harder than inducing a parser from the Penn treebank of newspaper text annotated with syntactic trees. The child has to also discover the alignment of the structures of the meaning representation with the words of the target language. Doing all this in a single incremental pass through the data involves building a parsing model for all possible languages that are consistent with the observed data, rather than just the one language whose model will eventually predominate. Kwiatkowski et al. [20] show how this can be done for real child-directed speech using a Variational Bayes version of the incremental EM algorithm [23].[5]

These two views are not as different as they might seem. As Jespersen realized, the view that structure precedes language necessarily assumes a mechanism for correctly combining the words that are supposedly attached to preexisting structures, projecting them in a way that supports compositional semantic interpretation. Jespersen's proposal was that these birdsong-like structures were initially attached to complex meanings as analyzed holophrastic labels,

of the enterprise [5], seem to be essentially methodological, combining a useful restatement of the problem with advocacy of an antireductionist methodology relying primarily on linguistic data. In terms of actual explanations, its adherents generally seem to fall into one of the two following camps.

[4] When we say that the child's language learning is "supervised" we do not of course imply the presence of a human supervisor or teacher, but refer rather to the mere availability to the child of information concerning the correct analysis.

[5] I refer to the homologous conceptual substrate as a "language of mind" because Fodor insists on defining his "language of thought" as the language-specific logical form of the language, essentially homomorphic to its syntax and lexicon. The language of thought in Fodor's sense must be assembled from a more primitive and ramified universal language of mind during child language acquisition.

Please cite this article in press as: Steedman M. Evolutionary basis for human language. Phys Life Rev (2014), http://dx.doi.org/10.1016/j.plrev.2014.06.010

and were later decomposed into the phrasal and word-level units of the target language. This process is of course only possible if the meaning representations themselves have homologous structure, as assumed under the other view, that meaning precedes language specific syntax, and it is in fact in broad outline the way that the variational Bayes model of Kwiatkowski et al. works in practice.[6]

As another example of the interdependence of the structure-first and meaning-first views, it is interesting to consider George Miller's [22] report on *Project Grammarama*, which Fitch cites in support of his *Supra-Regular Hypothesis* that "*when presented with sets of strings*, humans have both a capacity and proclivity to infer hierarchical structures wherever possible" (emphasis added). Miller makes the following remark in the conclusion to his chapter, concerning the nature of the information necessary for successful language acquisition:

> "When skilled behavior can be analyzed into independent responses, either overt, or covert, that can be reinforced individually, and assembled without significant interaction, the principles of learning derived from conditioning experiments may be applicable. Independent components, however, are not characteristic of rule-guided human behavior, and the systemic aspects cannot be avoided. Under those conditions, therefore, it seems reasonable to assume that *the feedback must convey information at least as complex systemically as the rules to be learned*." (Emphasis added.)

A similar point is made by Fodor [9,10], as for example in the epigraph to this article: you can't learn a language unless you already know a homologous language [10, p. 64].

2. Plans and the structure of mind

If you can't learn a language unless you already know an equally expressive language, where does that preexisting language come from? If we are to escape infinite regression, we know it cannot by definition be learned.

However, it can be evolved. Learning has to be done with the bounded resources of individual finite machines. Evolution has virtually unbounded resources, with numbers of processes limited only by the physical resources of the planet, and processing time limited only by the latter's continued existence. It essentially works by trying every possible variation on every viable variation so far.[7]

The only plausible source for the universal proto-language is as a result of half a billion years of chordate evolution, resulting in a symbolic language of mind, grounded in our physical being in the world, which we must share to varying degrees with our animal cousins, in proportion to their evolutionary proximity.

This observation of course raises the further question of why, in that case, even our closest animal relatives (who must share almost all of this precursor) show no signs of being able to learn anything like natural language as a formal system. There must have been *some* evolutionary advance, but it must be very minor to have arisen so quickly.

I am very skeptical towards the idea of this advance being something as complicated and singular as the sudden introduction *de novo* of trees, recursion, or the related "merge" operation of the Chomskian Minimalist program. First, I am going to argue that those mechanisms are more ancient, and have evolved over a very long period for a more cognitively general purpose which we share with some animals, namely deliberative planning of action in the world [28]. Second, I am going to argue that our failure to show that animals can learn recursive definitions cannot be taken as evidence against this claim, because it is extremely difficult to prove that even *human language* is recursive on the basis of finite stringsets alone.

In fact, far from being something totally singular, like recursion, the advance that supports language must be something that evolution can come up with *easily*. And if it is easy, then it is likely to be something familiar, something that evolution *already has* come up with repeatedly, which if added to a mix of other traits from our immediate relatives, will provide a conceptual base homologous to human language. One plausible candidate is cooperation [29,30], and in

[6] Jespersen also believed that children could do this for themselves, provided there was more than one of them and they could stay alive, a position for which he offered evidence from differential Native American language diversity between California and Arctic North America, and from certain rare cases of language development in isolated twins.

[7] It is a little more complex than this. Even genomes must be structured programs, and even evolution needs the occasional mass-extinction to escape overfitting.

Please cite this article in press as: Steedman M. Evolutionary basis for human language. Phys Life Rev (2014), http://dx.doi.org/10.1016/j.plrev.2014.06.010

particular cooperative breeding [1,15], both of which have arisen a number of times in mammalian evolution, though apparently not among apes other than ourselves and perhaps some of our hominid ancestors.

We will not rehearse Tomasello's and Hrdy's arguments here, except to note that when they invoke the idea of cooperation they are talking specifically about deliberative planning involving calculation about other minds, rather than mere collective action that reinforcement learning or evolutionary selection has found to work better when there are multiple agents, such as hunting in packs or flocking behavior.

Deliberative planning—in particular, planning involving tools—is an ability that we do share with other species, notably chimpanzees and some of the more recently evolved birds. Chimpanzees really can solve the so-called "monkey and bananas problem" using such tools as crates in stacks [19]. This kind of planned action is quite unlike the undirected reactive behavior that results when Skinnerian shaping with reinforcement is used in attempts to get pigeons to solve monkey-and-bananas problems.[8]

There are some specific differences observed by Köhler between the kind of planning that apes and humans are capable of that it will be useful to bear in mind. While apes can form quite complicated plans involving multiple tools (such as towers of many crates), their doing so depends on the perceptible availability of the tools. Plans which involve going to another room and fetching crates, even if they have in the recent past been observed there, are much harder for the animal to attain than for humans. The apes also show deficiencies in comparison with humans in their abilities to make plans involving other minds.

There is an interesting connection between these two kinds of planning difficulties in apes. If I need to talk to my brother then a good plan is to grab my phone and dial +1803318338. But suppose I don't know that number? A bad plan is to grab my phone and dial every number until my brother replies. A better plan involves the idea of a *plan variable* [7,25]. I should grab my phone, look up my brother's number X, and dial that number X. Another good plan is to grab my phone and dial +44207263482, ask my sister to tell me my brother's number X, hang up, and dial that number X. The first involves the use of a non-present tool (the phonebook on my phone). The second involves the use of a non-present other mind. But they seem pretty similar.

Apes seem good at making plans like $\mathit{place}\ \mathit{box}_1; \mathit{climb\text{-}on}\ \mathit{box}_1$ ("place that box and climb on it"), but specifically bad at finding plans that call for plan variables, such as $\mathit{find}\ x \wedge \mathit{affords}(\mathit{climb\text{-}on}\ x)\ x; \mathit{place}\ x; \mathit{climb\text{-}on\text{-}x}$ ("find something that affords climbing, place it, and climb on it"), and $\mathit{find}\ x \wedge \mathit{affords}(\mathit{ask}\ x\ y)\ x; \mathit{call}\ y$ ("find someone to ask the number, then call it").

Thus, tools and people seem rather similar in terms of human plans. It is an open question in evolutionary and developmental terms whether the use of tools precedes and is a model for the use of agents and other minds, or whether it is the other way round, with evolved cooperation modeling tool use. It seems easier to believe the latter, at least in terms of evolution.

The distinctive character of the latter kind of plan is that they are *functions over (possibly multiple) tools and/or agents such as people*, whose result is a plan of the first kind, with the variables instantiated by values, such as specific persons or boxes. Since the number of variables in plans like "persuade someone to find someone to authorize someone to send me the account number" seems to be essentially unbounded.

3. Planning, search, and recursion

Deliberative planning is very well understood in computational terms. It involves search for a desirable state in a directed acyclic graph whose arcs correspond to causal actions represented in a search-efficient logic of change, such as STRIPS or PDDL [21]. Searching such a graph involves the use of a stack or push-down automaton (PDA), essentially to keep track of all the alternative possible action at each branching node. So animals like chimpanzees must have a PDA.

The PDA is of course the automaton that is necessary and sufficient to parse all context-free (CF) languages. It follows that the reason that apes don't seem to use even CF syntax must lie elsewhere than in their lack of such a device.

One might ask why in that case it is so hard to get animals to solve recursive puzzles such as recognizing palindrome languages. The answer is that, as Miller [22] showed in his Grammarama project, *even humans* find these tasks very

[8] See YouTube for examples.

Please cite this article in press as: Steedman M. Evolutionary basis for human language. Phys Life Rev (2014), http://dx.doi.org/10.1016/j.plrev.2014.06.010

difficult *unless they are provided with a semantics.* (Miller made ingenious use of simple generative graphics to provide access to meanings for the strings of his CF language.)

One might further ask whether providing similar graphical semantic support would help animal subjects to learn properly CF grammars in the same way. The answer lies in Fodor's observation in our epigraph that "you cannot learn a language whose terms express semantic properties not expressed by the terms of some language that you are already able to use". Recognizing the property of being a balanced tree or a palindrome requires having mastered concepts such as *symmetry* and *right vs. left*. There is evidence from early studies of non-signing deaf children that such concepts, unlike more perceptually hard-wired ones like shape and color, are extremely hard to attain in advance of learning human language [24,11,12]. Thus there is every reason to believe that this semantics is only accessible to humans who have already mastered their native language, via the conceptual grounding in real world interactions of its semantics.

But apes also live in the real word, equipped with a PDA, so we still haven't explained why the apes haven't made the same progression to language for themselves.

The answer must lie in the prelinguistic conceptual language of the mind that human infants command and the apes do not. More specifically it may reflect the *kinds* of plans that humans and apes can respectively make. We have already noted that apes appear to be unable to make plans that require unboundedly many plan variables, notably those that involve other minds. Such plans are semantically recursive, and show up in recursive structures in natural languages in sentences like the following (indices associate verbs with their subjects etc.):

(1) a. that we$_1$ let$_1$ the children$_2$ help$_2$ Hans$_3$ paint$_3$ the house$_3$.
b. daß wir$_1$ die Kinder$_2$ dem Hans$_3$ das Haus$_3$ streichen$_3$ helfen$_2$ lassen$_1$.
c. das mer$_1$ d'chind$_2$ em Hans$_3$ es huus$_3$ lönd$_1$ hälfe$_2$ aastriiche$_3$.

In English (1a) this construction is tail-recursive, but in German (b) it is center-embedding, while in Zurich German (c) it is center embedding with crossing dependencies. In the latter case, it is known that such languages are non-context-free [27].[9]

In that case, before any verb combines with its subject, the syntax must derive a verb cluster *lönd*$_1$ *hälfe*$_2$ *aastriiche*$_3$ with a logical form isomorphic to the plan with variables $let(help(paint\, z\, y)\, x)\, w$, "someone lets someone help someone paint something". I conjecture therefore that it is the language of planning with plan variables that is the preexisting language of mind, in terms of which language-specific syntax and semantics are defined, rather than the dendrophile's empty tree structures.

Recognizing or parsing such sentences or building the corresponding plan both require a slightly more powerful automaton than the PDA called a Linear Embedded Push-Down Automaton, or LEPDA [31]. The EPDA resembles the PDA in operating with a single push-down store or stack. However, the items on the stack may themselves be stack-valued. The LEPDA is a special case where only one stacked element carries an unbounded stack value.

Since humans (like chimpanzees and all other real physical computational systems) are in fact finite state machines, the PDA and the slightly more powerful LEPDA must be "virtual machines", simulated using finite resources. We may accordingly expect that there will be limitations on the complexity of the plans that chimpanzees (and ourselves) can find. These limitations may to some extend be overcome by learning, in the form of storing successful and robust plans as compiled or assembled functions that can be used as single steps in more ambitious plans that could not have been achieved starting from scratch.

This process of storing functions assembled from more primitive functions is all that is needed to assemble a lexicon corresponding to the syntactic and semantic elements of natural language on the basis of exposure to string paired with meanings in the prelinguistic language of mind.

We may therefore conjecture that the following progression, spanning a couple of hundred million years of mammalian evolution, provided the necessary substrate for the essentially instantaneous subsequent development of human language and the other cognitive faculties we have mentioned.[10]

---

[9] In German and Zurich German other word orders are allowed. This does not affect the argument.

[10] I do not intend to suggest that parts of this progression may not have occurred in non-mammals. The observations of Emory and Clayton [6], Huber and Gajdon [16] suggest that partial parallel evolution may have occurred in species of Crow and Parrot.

Please cite this article in press as: Steedman M. Evolutionary basis for human language. Phys Life Rev (2014), http://dx.doi.org/10.1016/j.plrev.2014.06.010

1. Pure reactive planning with a nonrecursive action-representation is finite-state.
2. Deliberative planning with a nonrecursive action-representation, requires a (simulated) PDA.
3. A PDA supports recursion in the action- and state-representation language of the mind used for planning.
4. The evolution of human cooperative behavior under such pressures as neotenous child rearing requires plans with variables.
5. Planning with plan variables in a recursive state-action language requires simulating a (linear) embedded PDA (LEPDA).
6. An LEPDA supports the limited but trans-context-free complexity of attested natural language grammars.

4. Conclusion

The central claim of Fitch's paper is that humans are subject to a dendrophylic drive to parse strings into trees, compute probability distributions, then use them to generate novel trees and attendant strings. While other animals may have trees as a byproduct of motor control, navigation, and social control, only humans "directly encode" such structures, independent from their sensory motor origins. The process of language acquisition then corresponds to the attachment of language to this preexisting content-free armature.

This commentary argues that the preexisting tree structures are there because they constitute a preexisting symbolic language, related to sensory motor planning, of a kind that we share much but not all of with chimpanzees, and which constitutes the substrate to the semantics of human languages, and one of the inputs to child language acquisition. The reason other animals don't show evidence of language arises from a qualitative difference in their plans. Human plans can deal with situations that are not the sensory motor present, and with actions upon other minds that change knowledge. These show up formally in the planning language as "plan variables", and have their origin in an evolved propensity for cooperation with other minds, of the kind hypothesized by Tomasello, possibly arising from pressures of human child rearing of the kind discussed by Hrdy.

References

[1] Burkart J, Hrdy SB, Van Schaik C. Cooperative breeding and human cognitive evolution. Evol Anthropol 2009;18:175–86.
[2] Charniak E. Statistical parsing with a context-free grammar and word statistics. In: Proceedings of the 14th national conference of the American Association for Artificial Intelligence. Providence (RI). July 1997. p. 598–603.
[3] Collins M. Three generative lexicalized models for statistical parsing. In: Proceedings of the 35th annual meeting of the Association for Computational Linguistics. Madrid: ACL; 1997. p. 16–23.
[4] Darwin C. The descent of man. John Murray; 1871.
[5] Di Sciullo MA, Boeckx C. The biolinguistic enterprise: new perspectives on the evolution and nature of the human language faculty. Oxford University Press; 2011.
[6] Emory N, Clayton N. The mentality of crows: convergent evolution of intelligence in corvids and apes. Science 2004;306:1903–7.
[7] Etzioni O, Hanks S, Weld D, Draper D, Lesh N, Williamson M. An approach to planning with incomplete information. In: Proceedings of the 2nd international conference on knowledge representation and reasoning (KRR-2). 1992. p. 115–25.
[8] Fitch T. Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. Phys Life Rev 2014. http://dx.doi.org/10.1016/j.plrev.2014.04.005 [in this issue].
[9] Fodor J. How to learn to talk: some simple ways. In: Smith F, Miller G, editors. The genesis of language. Cambridge: MIT Press; 1966. p. 105–22.
[10] Fodor J. The language of thought. Cambridge (MA): Harvard; 1975.
[11] Furth H. The influence of language on the development of concept formation in deaf children. J Abnorm Soc Psychol 1961;63:386–9.
[12] Furth H. Thinking without language: psychological implications of deafness. New York: The Free Press; 1966.
[13] Hauser M, Chomsky N, Fitch T. The faculty of language: what is it, who has it, and how did it evolve? Science 2002;298:1569–79.
[14] Hilliard A, White S. Possible precursors of syntactic components in other species. In: Bickerton D, Szathmáry E, editors. Biological foundations and origin of syntax. Cambridge (MA): MIT Press; 2009. p. 161–84.
[15] Hrdy SB. Mothers and others. Cambridge (MA): Belnap/Harvard University Press; 2009.
[16] Huber L, Gajdon G. Technical intelligence in animals: the Kea model. Anim Cogn 2006;9:295–305.
[17] Jarvis E. Learned birdsong and the neurobiology of human language. Ann NY Acad Sci 2004;1016:749–77.
[18] Jespersen O. Language: its nature, development, and origin. Allen and Unwin; 1922.
[19] Köhler W. The mentality of apes. New York: Harcourt Brace and World; 1925.
[20] Kwiatkowski T, Goldwater S, Zettlemoyer L, Steedman M. A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their meanings. In: Proceedings of the 13th conference of the European chapter of the ACL (EACL 2012). Avignon: ACL; 2012. p. 234–44.

[21] McDermott D, Ghallab M, Howe A, Knoblock C, Ram A, Veloso M, et al. PDDL the planning domain definition language. In: DCS TR-1165. New Haven (CT): Yale Center for Computational Vision and Control; 1998.

[22] Miller G. Project Grammarama. In: Miller G, editor. The psychology of communication. New York: Basic Books; 1967. p. 125–87.

[23] Neal R, Hinton G. A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Jordan M, editor. Learning in graphical models. Cambridge (MA): MIT Press; 1999. p. 355–68.

[24] Oléron P. Conceptual thinking of the deaf. Am Ann Deaf 1953;98:304–10.

[25] Petrick RPA, Bacchus F. A knowledge-based approach to planning with incomplete information and sensing. In: Proceedings of the sixth international conference on artificial intelligence planning and scheduling (AIPS-2002). Menlo Park (CA): AAAI Press; 2002. p. 212–21.

[26] Pinker S. Formal models of language learning. Cognition 1979;7:217–83.

[27] Shieber S. Evidence against the context-freeness of natural language. Linguist Philos 1985;8:333–43.

[28] Steedman M. Plans, affordances, and combinatory grammar. Linguist Philos 2002;25:723–53.

[29] Tomasello M. The cultural origins of human cognition. Cambridge (MA): Harvard University Press; 1999.

[30] Tomasello M. Why we cooperate. Cambridge (MA): MIT Press; 2009.

[31] Vijay-Shanker K, Weir D. The equivalence of four extensions of context-free grammar. Math Syst Theory 1994;27:511–46.

[32] Wexler K, Culicover P. Formal principles of language acquisition. Cambridge (MA): MIT Press; 1980.