POINTER EVOLUTION POINTS TO GRADUAL EVOLUTION OF HIERARCHICAL COMPLEXITY

SVERKER JOHANSSON [*1]

[*]Corresponding Author: sja@du.se
[1]Dalarna University, Falun, Sweden

Chomsky (e.g. 2010) and others regard unlimited Merge as the defining feature of language, that cannot evolve gradually. The neural implementation of Merge is not well understood (Rizzi 2012, Zaccarella et al 2017), but must involve something functionally equivalent to pointers in working memory. Every Merge requires two pointers, and full syntactic trees may require dozens. Other syntactic paradigms also need pointers.

Humans do hierarchies in general better than chimpanzees. Any hierarchical thinking requires nested pointers in working memory, but they are neurologically expensive and degrade with depth (Crawford et al. 2016). Humans have larger working-memory capacity than chimpanzees, which has been proposed as key to human cognitive evolution (Read 2008, Coolidge & Wynn, 2005). Gradual evolutionary growth of pointer capacity will allow gradually increasing syntactic complexity, without saltations in the underlying computational machinery. Both depth degradation and pointer capacity naturally limit Merge even in modern humans, consistent with corpus data (e.g. Karlsson 2010).

1. Can infinity evolve?

Language is commonly said to be infinite, and this is true at least in the limited sense that there are no limits to what can be said. But some popular linguistic paradigms, notably minimalism (Chomsky 1995), postulate that language is inifinite in a stronger, more literal sense, in that the language faculty can, in principle, generate an infinite number of infinitely long sentences. The generating hierarchy-building operation (Merge, in Chomsky's case) can build trees of arbitrary depth, without limits. This kind of inifinity is sometimes invoked as an argument against the gradual evolution of the human language capacity, roughly along the following lines (e.g. Chomsky 2010, Berwick & Chomsky 2015):

1. Language is based on Merge, which is the defining feature of language.
2. Merge is unlimited.
3. Language, generated by unlimited Merge, is infinite.

4. There is no gradual path from the finite to the infinite, no possible intermediate "semi-infinite" proto-language.
    a. Having a limited Merge, and then gradually evolving an increased limit, will not get us to infinity.
5. Merge, and thus language, must have arisen in a single step. There is no such thing as "half-Merge", no intermediate stage between no Merge and full Merge (Berwick & Chomsky 2019).
6. *Conclusion*: Language did not evolve gradually.

But humans are indubitably descended from language-less ape-like ancestors. And the notion that something as complex as language could have arisen in one fell swoop, by a single super-mutation, is not tenable (e.g. Tallerman 2014, de Boer et al. 2019).

If the human language capacity did evolve in several steps, one of the points above must be incorrect. The question is which one.

2 Merge is limited

I will focus first here on assumption #2, that Merge is unlimited, and show that this is not a correct description of the actual human language capacity.

The term "Merge" *sensu stricto* is specific to Chomskyan minimalism, but regardless of linguistic paradigm there will be a need for some kind of hierarchy-handling neural machinery in the language capacity, as human languages indubitably do have some hierarchical structure. The form that the hierarchies take differ widely between paradigms, but my argument here is intended to apply over a broad range of hierarchy-handlers. I will use "Merge" here as a label, both because it is the best-known hierarchy handler, and because most proponents of the infinity argument above are Chomskyans, but I use it *sensu lato*, as a label for a generic hierarchy-handler.

From an empirical perspective, language is clearly finite. The human brain has a finite size, and the human lifespan is finite, so infinite production is obviously impossible in practice. Actual language usage shows that Merge in practice is not just limited, but limited to fairly shallow depths – in written corpus data, it is exceedingly rare to find examples of phrasal embedding more than three layers deep, and the limits appear even tighter in spoken corpora (Karlsson 2010). Similarly, the accuracy of grammaticality judgements approaches chance level as embedding depth increases beyond what is commonly used (Christiansen & MacDonald 1999).

These performance limits keep real human languages from being infinite. Languages are still very large; a few levels of embedding combined with a normal

vocabulary still allows for an astronomical number of different sentences – unlimited for all practical purposes but not infinite.

However, it has been argued since Chomsky (1965) that performance is not interesting, that performance limitations just distract attention from the real underlying linguistic competence. The latter is what linguists should study, and the latter is postulated to be infinite.

But is competence, a theoretical entity that is never directly observed, actually the proper target for language *evolution* studies? Isn't it enough to account for the evolution of actual language usage?

3 How can Merge be implemented in the brain?

The neural implementation of Merge is not well understood (Rizzi 2012, Zaccarella et al. 2017). Discussions in the literature are mainly about the computational machinery, but I will focus instead on memory needs, that are non-trivial. In order to build a syntactic structure in the brain, two types of objects need to be stored in memory: (1) lexicon storage in long-term memory (LTM), and (2) syntactic nodes in working memory (WM). LTM and WM are distinct types of memory, with distinct characteristics (Norris 2017). The leaf nodes in the syntactic tree in WM must somehow refer to lexical items in LTM. Unless entire lexical items are copied into WM nodes, this must involve something functionally equivalent to *pointers* (e.g. Reilly 2003) in working memory, where a WM node contains a reference to a LTM item (Takac & Knott 2016); this would be 0-merge *sensu* Rizzi (2016).

A Merge operation will create a new object in WM, which consists of two pointers, one to each object that is merged (plus features and whatever else is stored at each node). If both pointers refer to lexical items in LTM, we have 1-merge *sensu* Rizzi (2016). The brain must be able to handle and refer to this new composite object as a single entity, a *chunk* of memory (cf. Gobet et al. 2016, Isbilen & Christiansen 2018), for purposes of further merging; chunking is a prerequisite for Merge.

For the next level of merging, it is not enough with pointers from WM to LTM; pointers from one WM location to another WM location are also needed. WM-to-WM pointers are likely neurally distinct from WM-to-LTM pointers, as the address space is different in kind. This means three different types of Merge nodes are needed in WM:

- Merging two lexical items (1-merge of Rizzi 2016).
- Merging a merged item with a lexical item (2-merge of Rizzi 2016).
- Merging two merged items (3-merge of Rizzi 2016).

All this is done in WM, and it can be noted that, while many animals do have working memory (Carruthers 2013), there is fair evidence that humans have more of it than even our closest living relatives (Read 2008, Coolidge & Wynn, 2005).

In recent works, Chomsky does note the need for working memory (or "workspace", as he calls it), for syntactic processing, but does not discuss either implementation or limitations (e.g Chomsky et al. 2017).

4 Pointers

Pointers – having one memory location contain a reference to another memory location – are used extensively in computer programming. Biological brains are doing many operations for which a computer would use pointers. Something functionally equivalent to pointers must be neurally implemented.

4.1 Who needs pointers?

Pointers in the brain are needed as soon as the brain in one location manipulates information that is stored elsewhere in the brain. Notably, any WM computations involving LTM items will require pointers. It is rather pointless to have both types of memory, unless you have pointers as well. But both WM and LTM are widespread among non-human animals. Fish have both WM (Hughes & Blight 1999) and LTM (Lucon-Xiccato & Bisazza 2017), so presumably all vertebrates do, though most likely not identical to human memory in capacity or capabilities (Carruthers 2013). Pointers, at least WM-to-LTM, thus have an ancient origin. This means that 0-merge (Rizzi 2016) is available to all vertebrates. For 1-merge (or higher), chunking is required, for which there is evidence in e.g. rats (Fountain & Benson 2006) as well as some other mammals and birds, but negative results for fish and amphibians (Wickelgren 1979), suggesting a more limited distribution that nevertheless includes many (all?) non-primate mammals.

For 2+merge, WM-to-WM pointers are needed. Any WM operations involving hierarchical structures would be evidence of WM-to-WM pointers. There is some evidence of hierarchical cognition in non-human primates (e.g. Seyfarth et al. 2005), but it is not strongly compelling. The jury is still out on 2+merge in non-humans.

Humans, however, are hierarchical thinkers *par excellence*, to the extent that Fitch (2014) labels us "dendrophiles" for our propensity to use hierarchical thinking and impose hierarchical structure on anything and everything. Martin & Doumas (2017) propose that this general mechanism for thinking hierarchically can be repurposed for linguistic structures.

4.2 Neural implementation of pointers

Pointers are inherently difficult to handle in the brain, as memory addressing is not a matter of just storing the number of the addressed memory cell, like it is in a computer. The content addressable memory in the brain requires a fundamentally different type of pointers, that are neurally quite expensive. This is particularly true when attempts are made to scale up proposed pointer models to a human-sized address space; most models do not scale well and cannot address a human-sized memory with the number of neurons available in a human brain (Blouw et al. 2016). Exact pointers are particularly vulnerable to scaling issues, whereas different types of approximate pointers fare better (Crawford et al. 2016, Legenstein et al. 2016).

The model of Crawford et al. (2016) is attractive in this context, as it has explicitly been shown in simulation to manage the full human lexicon with a reasonable number of neurons that will actually fit within the relevant brain areas. This model is based on lossy compression of information; accuracy remains adequate for single pointers, but degrades rapidly with depth when pointers are nested in recursive structures; the degradation mimics actual human performance (as opposed to theoretical competence) on multi-level embeddings.

But pointers remain expensive and consume WM fast, especially if you have hierarchical structures with multiple pointers-to-pointers.

5 The gradual evolution of limited Merge

As reviewed in the previous section, many non-human animals have LTM and WM, of limited size, as well as WM-to-LTM pointers, which implies 0-merge. Chunking, and thus 1-merge, likewise can be found in a fair range of animals (Wickelgren 1979), though the evidence for 2+merge outside the human lineage is more limited.

Contrasting this with the capacities of modern humans, we can conclude that the evolutionary changes in the human lineage, after we parted ways with the other apes, most likely include:
- Expanded WM.
- Expanded LTM (including possibly dedicated lexical storage).
- Dendrophilia.
- WM-to-WM pointers, and likely generally enhanced pointer handling.
- Node structures for 2-merge and 3-merge.
- Nested pointer handling.
- Enhanced chunking?

None of these changes need to be linguistically motivated; they are plausible components in the general human cognitive enhancement that took place concurrently (cf. Sherwood et al. 2008). There has obviously been considerable selective pressure along the human lineage for enhanced brain size and presumable enhanced cognition (Bailey & Geary 2009). Such a selective pressure explains the evolution of expanded working memory, as there is a strong correlation between WM size and general intelligence (e.g. Colom et al. 2008).

Given the WM cost of hierarchical structures, the WM expansion will make a big difference in hierarchy handling capacities. With a small WM, you cannot build any significant tree structures even if full-blown Merge is computationally available. A larger WM invites the evolution of enhanced pointer handling, including WM-to-WM pointers and recursively nested pointers, which were pointless before. There is also more scope for chunking in WM, when there is room for more than a few chunks. Likewise, with a larger WM, dendrophilia (Fitch 2014) starts making sense.

This leaves us with a proto-human who has a fair-sized WM, a basic set of pointer operations, including WM-to-WM pointers, and a general Merge-like operation that can do all the merge levels of Rizzi (2016). Such a proto-human can combine lexical items into two-word phrases, and can combine two phrases into a composite utterance. Dedicated syntactic machinery such as feature-checking is still missing, leading to a rather anarchic proto-language, but the basic hierarchical structure is there. Pointers remain limited in both address space and nesting depth, imposing limits on both lexicon size and tree size. But both limits can be relaxed simply by gradually adding more neurons to the pointer machinery.

If there is selective pressure towards more expressive language, with more complex syntax and enhanced narrative capacities, this can be dealt with in at least three ways:
- Further WM expansion.
- Pointer expansion as above, to handle both a human-sized address space and the corpus-attested (Karlsson 2010) nesting depth.
- The general Merge operation already available may be augmented with language-specific add-ons, gradually adding all the operations that modern Merge does beyond the actual merging (feature-checking etc.), in order to make linguistic processing more precise and efficient.

The end result is modern humans, with a modern human language capacity. Merge, or whatever hierarchical operation is actually running the human language capacity, has gradually been enhanced to its modern full-fledged form, with no infinity paradoxes blocking the way.

References

Berwick, Robert C. & Chomsky, Noam. (2015) *Why only us?* Cambridge, MA: MIT Press.

Berwick, Robert C. & Chomsky, Noam. (2019) All or nothing: No half-Merge and the evolution of syntax. *PLOS Biology* 17:e3000539.

Bailey, D.H., Geary, D.C. (2009) Hominid Brain Evolution. *Hum Nat* **20,** 67–79.

Blouw et al. (2016) Concepts as semantic pointers: A framework and computational model. *Cognitive Science* 40:1128-1162

Carruthers, Peter (2013) Evolution of working memory. *Proc Nat Acad Sci* 110:10371-10378.

Chomsky, Noam. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press

Chomsky, Noam. (2010). Some simple evo devo theses: how true might they be for language? In Larson et al. (Eds.), *The Evolution of Human Language. Biolinguistic Perspectives*. Cambridge: Cambridge University Press.

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

Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. (2017). *Generative grammar and the faculty of language: Insights, questions, and challenges.* Unpublished manuscript. https://ling.auf.net/lingbuzz/003507

Christiansen, Morten H. & MacDonald, Maryellen C. (1999) *Processing of recursive sentence structure: Testing predictions from a connectionist model.* http://www.academia.edu/download/39800748/Processing_of_Recursive_Se ntence_Structu20151108-5695-17yb6hj.pdf (retrieved 2019-09-11)

Colom et al. (2008) Working memory and intelligence are highly related constructs, but why? *Intelligence*, 36:584-606.

Coolidge, Frederick L & Wynn, Thomas (2005) Working memory, its executive functions, and the emergence of modern thinking. *Cambridge Archaeological Journal* 15:5-26.

Crawford, Eric & Gingerich, Matthew & Eliasmith, Chris (2016) Biologically plausible, human-scale knowledge representation. *Cognitive Science* 40:782-821.

de Boer, Bart, Bill Thompson, Andrea Ravignani & Cedric Boeckx. (2019) *Evolutionary dynamics do not motivate a single-mutant thoery of human language.* bioRxiv preprint 517029 https://www.biorxiv.org/content/10.1101/517029v1

Fitch, W Tecumseh (2014) Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys Life Reviews* 11:329-364

Fountain, Stephen B. & Benson, Don M. (2006) Chunking, rule learning, and multiple item memory in rat interleaved serial pattern learning. *Learning and Motivation,* 37:95-112.

Gobet, F. & Lloyd-Kelly, M. & Lane, P. C. R. (2016) What's in a name? The multiple meanings of 'chunk' and 'chunking', *Frontiers in Psychology* 7:102

Hughes, Roger N. & Blight, Christine M. (1999) Algorithmic behaviour and spatial memory are used by two intertidal fish species to solve the radial maze. *Animal Behaviour*, 58:601-613.

Isbilen, E S & Christiansen, M H (2018) Chunk-based memory constraints on the cultural evolution of language. *Topics in Cognitive Science* 2018:1-14

Karlsson, Fred (2010) Syntactic recursion and iteration. In Harry van der Hulst, ed., *Recursion and Human Language.* Berlin/New York: Mouton de Gruyter,

Legenstein et al. (2016) Assembly pointers for variable binding in networks of spiking neurons. q-bio.NC/1611.03698 (https://arxiv.org/abs/1611.03698)

Lucon-Xiccato, Tyrone & Bisazza, Angelo (2017) Complex maze learning by fish. *Animal Behaviour,* 125: 69-75.

Norris, Dennis (2017) Short-term memory and long-term memory are still different. *Psychological Bulletin* 9:992-1009

Martin, A. E. & Doumas, L. A. A. (2017) A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biology* 15:e2000663.

Read, Dwight W (2008) Working memory: A cognitive limit to non-human primate recursive thinking prior to hominid evolution. *Evolutionary Psychology* 6:676-714.

Reilly, Edwin D. (2003). *Milestones in Computer Science and Information Technology.* Westport: Greenwood Press.

Rizzi, Luigi (2012) Core linguistic computations: How are they expressed in the mind/brain? *Journal of Neurolinguistics* 25:489-499.

Rizzi, Luigi. (2016) Monkey morpho-syntax and merge-based systems. *Theoretical Linguistics* 42(1-2): 139–145

Seyfarth, Robert M. & Cheney, Dorothy L. & Bergman, Thore J. (2005) Primate social cognition and the origins of language. *Trends in Cognitive Sciences,* 9: 264-266.

Sherwood, C.C., Subiaul, F. and Zawidzki, T.W. (2008), A natural history of the human mind: tracing evolutionary changes in brain and cognition. *Journal of Anatomy*, 212: 426-454.

Takac, M & Knott, A (2016) *Mechanisms for storing and accessing event representations in episodic memory, and their expression in language: A neural network model.* In Proceedings of the 38th Annual Conference of the Cognitive Science Society

Tallerman, Maggie. (2014) No syntax saltation in language evolution. *Language Sciences* 46: 207-219

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review, 86*(1), 44–60.

Zaccarella et al (2017) Building by syntax: the neural basis of minimal linguistic structures. *Cerebral Cortex* 27:411-421.