## 1. Main Thesis

There is a deep chasm between what we can do and what other animals cannot: language. What makes human language special is the *Merge Operation*; a compositional operation that takes two objects already constructed X and Y, and forms from them a new object that consists of the two unchanged, {X,Y}. The set notation is used because the objects don't have to be ordered. This operation can be applied recursively without bound, yielding an infinity of digital, hierarchically structured expressions. By combining the Merge Operation with atoms of language (basic lexical concepts) or previously constructed syntactic representations, a sensorimotor interface for externalization (including vocal learning and production), and a conceptual-intentional interface for thought, we start to build an account of what goes on in the brain.

### 1.1 An Evolutionary Account of What Happened

The paleo-archeological record for the lineage of *Homo* shows a sudden appearance of symbolic behavior around 80,000 years ago (Blombos Cave, South Africa). When we talk about symbolic behavior, we mean artifacts with no consequence for the immediate day-to-day survival; a sort of abstraction. Examples include geometric engravings, figurative art, beads for decorations, or burial "goods".The hypothesis is that symbolic behavior happened as a result of *Merge*. Because the records also show evidence of long periods of stasis (hundreds of thousands years) between the appearance of new *Homo* variants and the appearance of new technologies and behaviors, we can set the appearance of Merge back at 200,000 years ago, at the appearance of anatomically modern humans. This is also consistent with the notion of exaptation, that is, evolution by natural selection co-opts existing traits for new uses; there is no foreknowledge that a particular trait will be useful in the future. Innovations arise independently of the function that they will be eventually selected for. If we take unambiguous evidence of symbolic behavior as a proxy for language, modern humans had language before their exodus from Africa at 60,000 years ago. The argument the authors provide for this, is that as far as we are able to make out from historical records, the fundamental parametric properties of human language have remained fixed, varying only within prescribed limits. Then language (and Merge) must have appeared sometimes between 200,000 and 60,000 years ago. When it comes to other branches of the *Homo* lineage, like Neanderthals, the book notes that there is lack of unambiguous evidence for symbolic behavior (proxy for language). There is also evidence that since the human-Neanderthal split (400,000-600,000 years ago), the human brain development has been reshaped through several genetic events. In short, the evidence so far seems to show that Neanderthals didn't have language, but we cannot be certain.

What about nonhuman animals and songbirds? While birds can do linear chunking (called motif), and motifs can be iterated, there are no motifs found that in turn contain other motifs; for example, a tweet-trill combination that is itself contained within a warble motif. There is no Merge in songbirds. Nonhuman primates also seem to suffer from the same limitations as songbirds. The book describes one of the failed-attempts to teach chimps human language: Project Nim, which

attempted to teach to a chimp (named Nim) American Sign Language. What Nim was able to learn about ASL was a kind of rote memorization; short linear sign sequences. He never progressed to the point of producing embedded, hierarchically structured sentences, which a normal child can do by age three or four. If Nim wanted an *apple*, he would run through his catalog of all individual signs that had been associated with *apple*, retrieving *Nim apple, apple Nim, apple knife*, and so on. He didn't actually learn the word, or have the human concept for *apple*. For Nim, an *apple* was the object associated with the knife in the drawer that cut the apple, the place apples were found, and so on. The chimp did have direct connections between particular external stimuli and their signs, but it couldn't abstract concepts to later combine them with other concepts. It had no Merge ability. Compare this to human child, who can quickly get *the doggie*, from examples like *the apple* and *a doggie*.

How did the innovation of Merge arise? By a minor chance mutation (evolution by leap), that may have caused a slight rewiring of the brain. It must be minor, because it happened quickly in terms of evolutionary time. It must be by chance, because there is no evidence for anything else like it. Although in terms of genetic material there was nothing new, the affected areas in the brain may have presented novel input/output properties, which together with adequate input and output connections, performed novel information processing functions. Unfortunately, the actual neural representation for Merge is unknown.

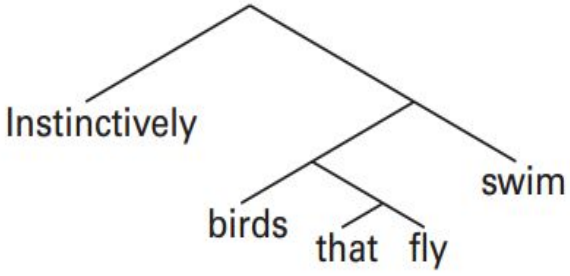
When it comes to the evolution of language, the book strongly suggests that the purpose of language was likely as an inner mental tool, and not as something driven by the need for external communication. One of the arguments given is that all biological functions can be met by a species without language; Wallace's problem. In this context, it's hard to see how something like language is needed for external communication, when simple animal signaling suffices. There is also experimental evidence that humans may use language to integrate representations from geometric and nongeometric modules. In one such experiment, a child sees that an object is placed in a geometrically assymetrical room with one single nongeometric clue, a blue wall; then the object is hidden. After closing his eyes, getting spinned to be disoriented, then reopening his eyes, the child will use the blue wall to locate the unique corner of the object only after he has a nearly full command of language (age 4-5).

**1.2 A Mechanism for Merge in the Brain**

A plausible mechanism will have to account for the *Basic Property of Language* (accounts of language as a biological object) and the *Minimalist System Assumptions* (simplifying UG to the most minimal set of computational principles). Language as a biological object means that we can think of language as an "organ of the body"; it has sufficient integrity that it can be studied in abstraction from its complex interactions with other systems. UG (Universal Grammar) is the theory of the genetic components of the language faculty. The main idea (in mid-1960s) was that

we could solve part of the mystery of language by discovering constraints on the biological systems for language acquisition. These constraints were accumulated in the "Principles and Parameters framework (P&P)". In this framework, the syntax of a language is a set of general principles (i.e. abstract rules or grammars) and specific parameters (i.e. markers, switches) that for particular languages are either turned on or off (e.g. head-initial "read books" in English vs head-final "books read" in Japanese). P&P is claimed to account for the diversity of languages; the different parameters come as a result of externalizing the language by the sensorimotor system, but the internal structure is similar for all languages. Because original attempts at UG turned out too complex to account for evolvability, later work aimed to simplify. The Basic Property requires that 1) language is a finite discrete (words are discrete) computational system yielding an infinity of expressions, with interpretations at interfaces with two other internal systems, 2) the sensorimotor system for externalization, and 3) the conceptual system for what is informally called "thought" (inference, planning, etc). Minimalist System Assumptions require that 1) human language syntax is hierarchical, with linear ordering constraints reserved for externalization; 2) the particular hierarchical structures associated with sentences affects their interpretation; and 3) there is no upper bound on the depth of the relevant hierarchical structure (in practice, it may be limited by processing difficulties; i.e. a sentence cannot be infinitely long).

The claim is that the proposed *Merge* accounts for both. But, the book doesn't provide a mechanism for Merge! It does however provide illustrative examples of what the mechanism should be able to handle. Because the book follows the Strong Minimalist Thesis, Merge must be optimal as determined by efficient computation, and as simple as logically possible. Merge should yield hierarchical expressions, because there is strong reason to believe that the way we handle language internally is hierarchical, and not linear (e.g. linear distance between words, or using counting to form a passive sentence). The sentence "*birds that fly instinctively swim"* is ambiguous. The adverb *instinctively* can modify either *fly* or *swim:* birds either fly instinctively, or else they swim instinctively. Take a second sentence "*instinctively birds that fly swim";* now *instinctively* can modify only *swim.* In terms of linear order of words *instinctively* is closer to *fly*, but on a hierarchical structure it is closer to *swim* in terms of structural distance (picture below). Another point made here is that the mechanism should be able to generate ambiguity, like in the case of the first sentence, or in the case of *deep blue sky*. Something like a finite-state network won't work, because under associativity for string concatenation the expressions *(deep blue) sky* and *deep (blue sky)* are equivalent.



**2. Critique**

The book is not very technical and can probably be categorized as a popular science book. As such, it does a good job of arguing that language makes us special, and that the innovation that made language possible happened relatively quickly in evolutionary time due to a minor mutation. It also does a good job of listing important conditions that a mechanism for such innovation should account for, and introducing relevant properties of language syntax for people with no background in linguistics. *My main problem with the book is that it does not provide a mechanism for that important innovation*, even though it claims that it does so, through Merge. The problem we are dealing with is that of compositionality, a key problem when it comes to understanding how our minds work, and how we can build systems of human-like intelligence. What the book describes as Merge, is just a different label for compositionality, envisioned as a binary operation (taking two objects). Another way to put this, is that Merge provides a linguistic description for the problem of compositionality. Given the history of the field of linguistics (generative grammars), it makes sense why the authors chose the most plausible description available up to date. Yet, the point is, we don't know how compositionality happens in the brain.

Another point is that, by reading the book's account of what the innovation of compositionality is, you get the impression that the innovation per se is natural language, since Merge is described as the basic operation for language syntax. The book also sees Merge as separate from the conceptual interface for thought. My guess is that such a distinction doesn't exist; Merge happens throughout the conceptual interface, and other parts of the brain too. I think that what the innovation of compositionality did is to make compound thoughts possible. Probably all of the "organs" of the brain are learning machines that adapt to the signals that come from their neighboring organs, as well as the sensory and motor inputs that they are connected to. The minor mutation might have made it possible for the organs to talk to each other in "complicated" "language". Here language refers to internal language: local protocols dynamically constructed by neighboring organs of the brain to make more complex communication among those organs possible. Figuring how this happens might amount to figuring out Merge. Given such communication mechanisms supplementing the existing primitive mechanisms (face recognition, object perception, motion planning, naive physics, etc), we can imagine that parts of the brain begin to tell other parts stories, representing information that may be useful to the neighbor. A plethora of internal languages and cultures can develop, implementing something like Minsky's Society of Mind. The next, and probably almost instantaneous step, is that some of those internal languages get externalized to become the human natural language organ. In this view, natural language was an emergent "invention" from complex internal communication of parts of the brain. Here, I'm using externalized as a combination of sensorimotor modules and conceptual ones.

Along these lines of reasoning, there is a something to be said about the point of the authors that language was "meant" as an inner mental tool, and not driven by external communication. If

language was the result of the ability to form compound thoughts, its "origin" lies somewhere in between these two positions. Some internal brain organ "discovers" that each time it makes a sensorimotor module to produce an output (vocal cords for sound), it gets some input as well (through some other organ connected to modules for hearing). If language was started as different internal organs making such "discoveries", even though its purpose wasn't exactly external communication, it still was in some ways used to establish input-output loops between the "internal" and "external". We cannot be certain that because we hear ourselves in language in our head when being alone, language was meant as an inner mental tool. It could as well be that the externalization process (input-output loops between internal brain organs and environment) that allowed for external communication between people, is extended to the person himself; you talking to yourself is a case of you talking to other people. Language was probably "meant" for both: inner mental tool and external communication.

From reading the book, it is also not clear to me why the mechanism for Merge/compositionality has to be simple. It falls under the Strong Minimalist Thesis, but the book doesn't really justify why that thesis holds. While the mutation for Merge was simple, the caused changes may have resulted in a mechanism that is hard to understand.

**3. Future Research**

Here are some possible research topics that reading the book can inspire students to work on.

- Figuring out the mechanism of compositionality. In the Critique section I talk a bit about my personal take on it.
- How are the atoms of language (concepts) formed? It's likely that every concept is built by something like Merge/compositionality operating on basic concepts. These basic concepts have perceptual origin. Somehow, some pattern of inputs from the world becomes common or important enough to be a symbol. A symbol is just a pattern of bits, but how it is stored in the brain and retrieved, is still an open question.
- How is compositionality related to the ability for short-term and long-term prediction? The evolutionary value of having a nervous system is the ability to quickly react to situations: to take advantage of opportunities and to avoid threats. Any improvement in these abilities is directly rewarded by more efficient reproduction. But these abilities require prediction of the future. Figuring out how advanced predictive modules build on organs responsible for primitive abilities, may lead to interesting results.
- Do whales have language or something similar to Merge/compositionality? The book talks about how Merge is missing from songbirds and land animals, but it doesn't consider marine animals like whales. Whales have very sophisticated forms of communications, and they may be just as smart as primates. Furthermore, whales have songs, while primates don't. This makes them very interesting creatures to study when it comes to intelligence.