

INCREMENTAL EVOLUTION OF SYNTAX: FROM ‘BIGRAMS’ TO ‘BRANCHING’

Dieter Hillert^{*1}

^{*}Corresponding Author: dhillert@gmail.com

¹School of Speech, Language, and Hearing Sciences,
San Diego State University, USA

Here, we explore how syntax might have evolved after the hominin lineage branched off from our closest living relatives, monkeys and nonhuman great apes. Our approach focuses on ‘bigrams’ and ‘branching’ in terms of unordered, external and internal Merge in association with workspace and memory storage resources. The syntactic abilities of monkeys and nonhumans great apes but also various typologies of human syntax in children and adults indicate that the language capacity incrementally involved in the hominin lineage. Indirect paleoanthropological records let us moreover predict that already *Homo erectus* was equipped with a language capacity which was premodern or even modern.

1. Introduction

All organisms are adapted to communicate with one another (or sometimes across species). Cultural transmissions, however, can be found only in a subset of life-forms, including fish, insects, birds, monkeys, and apes. The development and acquisition of these behavioral cultures are based on species-species genetic endowments. However, we should not ignore that modern humans share with other primates or mammals most neurobiology (e.g., neuronal morphology, neurotransmitter, subcortical structures) and the vocal tract of monkeys are considered as ‘speech-ready’. Thus, the neurobiological differences between nonhuman primates and modern humans can be considered as a ‘needle in a haystack’ (Fitch, 2011; Fitch et al., 2016).

Modern humans (or possibly their closest extinct ancestors), however, added another dimension by inventing a language system to share subjective experiences or to express thoughts. Already, Lenneberg (1967) recognized that this dimension is based on an innate human-specific capacity, and Chomsky (1968) described the human language capacity as an innate ‘universal grammar’ underlying all human languages. Speakers make ‘infinite use of finite means’, that is, a finite set of syntactic rules are used to create an infinite number of sentences or text structures, aka “super-sentences” (Watumull et al., 2013). Our closest living relatives, the nonhuman great apes, clearly display superior cognitive abilities compared to monkeys or other animals. We explore for this reason how this innate capacity

evolved in context of the expansion and refinement of (sub-) cortical circuits after the hominin lineage branched off from monkeys and nonhuman great apes.

2. Bigrams

The strong ‘minimalist program’ of generative linguistics explicitly claims that syntactic recursion is an innate property of UG (Chomsky, 1995; Hauser et al., 2002; Berwick & Chomsky, 2016). This recursion concept is based on the operation ‘unbounded Merge’, that is, it recursively merges words to create bottom-up larger phrases in a binary fashion. The simplest form of Merge is to take two distinct elements of the lexicon such as (X, Y), and combine them to an unordered binary set unit {X, Y}. It is a binary unit without a syntactic head which determines the syntactic category of a phrase. Headless binary sets are still marked by an unspecified (neutral) label. However, the pre-Merge conditions involve the ability to form and recall discrete elements and create bigrams.

Bigrams can be found not only in our closest living relatives, monkeys and nonhuman great apes, but they also (re-) surface in varieties of modern spoken and sign languages. Bigrams are rooted in non-linguistic cognition across primate species and represent the basic linear layer of human syntax across different domains (e.g., Jackendoff, 2003; Miyagawa, & Clarke, 2019). In addition to long-term ‘memory storage’ (MS) for lexical elements, our account considers the increase of temporary ‘workspace’ (WS), which is required for syntactic operations (e.g., Rizzi, 2016). We do not exactly know which behavioral and/or environmental changes triggered the expansion of (sub-)cortical regions and circuits in the hominin lineage, but the increase of brain structures goes hand-in-hand with an increase of WS capacity required for more complex intrinsic structure-building processes.

2.1. Monkey Syntax

Monkeys primarily communicate by means of vocalizations. Their call repertoire consists of a small set of discrete alarms calls. One or more calls can be combined with another call, either discrete or suffixed. Female Diana monkeys (*Cercopithecus diana*), for example, emit four calls whereas call A, which refers to caller identity, is systematically combined with the calls L, H, R to function as a social context modifier (Candiotti et al., 2012). Male putty-nosed monkeys (*Cercopithecus nictitans*) produce the general alarm call P (pyow) and the specific eagle-related alarm call H (hack). Both calls combined refer to a non-predator situation, whereas H follows always P. The number of Hs and Ps in a sequence informs about the distance of group movement (Schlenker et al., 2016). Cotton-top tamarins’ (*Saguinus oedipus*) long-call combinations are emitted by socially

isolated individuals to get in touch with their conspecifics. These contact calls consist of one or more chirp-like syllables followed by two or more whistle-like syllables. Again, two call types are combined in a fixed order and the variation is the number of call type repetitions (Cleveland, & Snowdon, 1982). Black-fronted titi monkeys (*Callicebus nigrifrons*) produce A-calls for predators in the canopy and B-calls for predators on the ground. The combination of both call types informs conspecifics not only about predator type but also about the location of the threat and the number of continuous B-calls might indicate probabilistic rather than categorial meaning (Casar et al., 2013; Berthet et al., 2019). These examples show that the syntax of monkeys is restricted to the combination of maximal two elements in a fixed order and their WS = 1 (e.g., Pasternak, & Greenlee, 2005; Wakita, 2020).

2.2. Nonhuman Ape Syntax

Old World monkeys and apes branched of about 30-25 million yrs. ago (mya). Thus, we expect to find qualitative differences between both clades about their abilities to structure their ‘language’. Indeed, great apes coordinate often facial expressions, postures, vocalizations, or gestures, and they outperform monkeys on many cognitive tasks in an enculturated environment. The photographic memory capacities of chimpanzees (*Pan troglodytes*) are impressive. *Ai* and her son *Ayumu* outperformed humans in tasks which required to recall number sequences (Matsuzawa, 1985). Since the cognition of great apes’ centers on visuospatial information and their communication is rather multimodal than relying on vocalization, they have been primarily trained on sign language or graphic symbols (lexigrams) to examine their language capacities.

Kanzi, the famous bonobo (*Pan paniscus*), acquired the meaning of lexigrams as an infant by observing his foster mother’s lessons. It is said that he learned about 348 of those symbols. *Kanzi* accompanied often graphic communication with an articulatory but distorted equivalent. He had the ability to communicate without difficulties EVENT, OBJECT combinations such as HIDE ICE or WATER HIDE. He also produced new combinations such as CAR TRAILER meaning he wanted to be driven to the trailer (Greenfield, & Savage-Rumbaugh, 1991). *Kanzi* acquired also some American Sign Language by watching a video of the signing gorilla *Koko* who learned thousand signs and understood about 2,000 spoken English words (Patterson, 1978).

Chantek was an orang-utan (*Pongo pygmaeus*) who mastered to learn approximately 150 signs and initiated communication to meet his needs. His vocabulary matches that of a very young child (Miles, 1993). He learned single concepts such as people names, places, food items, actions, animals, pronouns, locations, attributes, etc. and immediately began to combine his acquired signs

(e.g., He signed COKE DRINK after drinking a coke, or PULL BACK while pulling a caregiver's hair through the fence.) He began to over- or underextend the meanings of his signs (e.g., He used the sign DOG for dogs, barking noise, birds, horses, noisy helicopter, etc). *Chantek* also invented new sign combinations (e.g., EYE DRINK for contact lens solution). He clearly imaged objects or actions (e.g., He pointed to objects when he was asked WHERE HAT or WHAT WANT?). His signing was spontaneous, non-repetitious, did merely imitate his caregivers' signing, actively initiated signing to meet his needs.

Nonhuman great apes acquire conceptual representations of events and objects. In an encultured environment, they can flexibly combine two elements to express needs or ideas. Trigrams, however, are less often used. The order of the elements is in principle free but seems to follow mostly pragmatic strategies. For example, in DAVE MISSING FINGER the topic is 'Dave'; however, in HIDE ICE or WATER HIDE, the order EVENT or OBJECT seems to be subjective or random. The concepts are ordered linear, and it is unclear whether "grouping" takes place. WS of great apes is 2 ± 1 (Read et al., 2022).

3. Branching

In briefly reviewing monkey and nonhuman ape language, it is apparent that multiple components contributed to the evolution of syntax in modern humans. The continuous increase of cognitive resources in terms of WS and MS in the lineage of great apes (including humans) may be the main factors for the ability to generate complex structures between signals (cf. Zuberbühler, 2020). Moreover, complex structure-building processes can be found at all linguistic levels but also across non-linguistic domains (e.g., Fujita, 2017; Asano et al., 2021). The uniqueness of human syntax may be therefore based on an underlying universal computational capacity which includes enhanced WS/MS resources.

3.1. Human Syntax

Typically developing (human) children produce at around the age of 18 months two-word utterances such as WANT-JUICE or CAR-GO which lack inflections and function words (Bloom, 1993). Some accounts assume that this stage reflects WS limitations (Blake et al., 1994). At this point, human and encultured ape syntax are quite similar in production: (a) No significant differences were found for the type of bigrams such as AGENT-EVENT, EVENT-OBJECT, EVENT-LOCATIVE; (b) Auditory WS of 2 ± 1 matches between panins and 2-year-old toddlers (Patkowski, 2014).

Adult human language differs from nonhuman ape language in two fundamental aspects. The human child has the innate ability to acquire abstract

semantic, syntactic and phonological entries independent of immediate perceptual experiences and goals, and over time it can build or tap into additional cognitive resources to process complex sentence and discourse structures. Cowan (2001) argues therefore that only 4 ± 1 items can be held within focus of attention. Complex sentence processing ability is not enabled by WS per se but by its strategic subcomponents. However, human languages tap into these resources to different degrees.

For instance: Independent of particular language pairs, adult L2 learners acquire first a linear strategy with bigrams before they learn more complex structures. Broca's aphasic patients who suffer from agrammatism produce only telegraphic-like speech without inflections and function words. Bigrams are also the basic layer of particular sign languages. For example, the first generation of deaf children of the *Idioma de Señas de Nicaragua* combined signs to bigrams such as AGENT-EVENT, and the next generation elaborated on these structures (Senghas et al., 2004).

About ten million people speak the Malayan dialect *Riau Indonesian* which functions as a lingua franca. This language virtually has no syntactic categories or inflections, and the word order is based on semantic principles (Gil, 2005). For example, *ayam* means CHICKEN and *makan* EAT. The bigrams *makan ayam* or *ayam makan* are open to the same readings and they are unmarked for numerus, tense, aspect, semantic role, etc. Situational context is however sufficient that the speakers get the meaning across. Speakers of the indigenous, agglutinative Amazonas language *Pirahā* get by without rearranging canonical structures and clear evidence of hierarchical syntactic structures is missing (Everett, 2009). In general, the spectrum of typologies from analytical resp. isolating languages to polysynthetic languages varies strongly with respect to the use of syntactic hierarchies and recursion.

3.2. *Homo erectus*

Paleoanthropological data indicate an incremental neurobiological process. External evidence from phylogenetic and ontogenetic stages, let us hypothesize milestones of syntactic evolution. There are many more intermediary steps which need to be spelled out. Monkey's initial alarm response will be presumably replaced by a modifying response since they cannot keep more than one unit in WS. In contrast, nonhuman great apes acquire concepts rooted in perception and combine them into bigrams. Their lexicon may have an upper limit since they do not make use of compositional structures.

We assume that the next milestone broadly corresponds to *Homo erectus* since significant neurobiological and behavioral changes can be associated with this species. Based on indirect evidence, we predict an intermediate stage between

of structure-building processes between nonhuman great apes and *Homo sapiens*. The appearance of the C variant of the protein *SRGAP2* which involves neuronal migration and differentiation is also associated with the rise of *Homo erectus* about 2 mya (Hillert, 2015; Martins et al., 2018). The endocasts of *Homo erectus* share a pronounced Broca's cap, the bulge that appears at the level of the temporal pole in *Homo sapiens*. The specimens' brain sizes, moreover, range between 600-1,000 cc and had human-like prefrontal and temporoparietal regions. A recent endocranial morphology study of *early Homo* supports the view that frontoparietal areas did not evolve separately but in concert (Ponce de León et al., 2021). *We must, therefore, assume that the genetic endowment of a (pre-) modern language capacity preceded cultural-linguistic accumulations.* Acheulean tool industry coincides with the appearance of *Homo erectus* about 1.8 mya and Acheulean hand axes produced in Africa might fall in the category of art as their making reflects planning and abstract thinking. In addition, *late Homo erectus* produced non-functional, aesthetic engravings and decorations (Stout et al., 2015; Mcbrearty & Brooks, 2000). Accumulated cultural processes began about 1 mya in the merging species *late Homo erectus* and *pre-archaic resp. superarchaic Homo sapiens* (e.g., Bergström et al. 2021).

In sum, indirect neurobiological and cognitive-behavioral changes indicate that *Homo erectus* was equipped with a language capacity, whether it was premodern or even modern, we cannot tell (Dediu, & Levinson, 2013; Hillert, 2021). Possible WS differences can be considered as marginal. They might have used basic structures such as a linear grammar and headless binary sets (unordered Merge). Since direct evidence is lacking, we reserve asymmetric grouping (external Merge) and recursion (internal Merge) to the language of *Homo sapiens*.

4. Conclusions

In discussing the biology capacity of syntax and language in general, it is important to separate the genetic endowment of species and the extent to which a species makes use of its cognitive capacities. The evolution of syntax, from bigrams to branching, indicates that multiple factors played a crucial role, including WS and MS resources. In particular, inner speech (rehearsal) may have contributed to WS and representational resources required for complex sentence and discourse processing. Syntax incrementally emerged in the hominin lineage, and we must therefore assume that *Homo erectus* was already equipped with a premodern or even modern language capacity.

Acknowledgements

Supported by FY2020 JSPS invitational long-term fellowship at Kyoto University.

References

- Asano, R., Boeckx, C., & Seifert, U. (2021). Hierarchical control as a shared neurocognitive mechanism for language and music. *Cognition*, 216, 104847.
- Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E.M.L., & Skoglund, P. (2021). Origins of modern human ancestry. *Nature*, 590, 229-237.
- Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. *Science Advance*, 5, eaav3991.
- Berwick, R.C., & Chomsky, N. (2016). Why only us; language and evolution. Cambridge, MA: MIT Press.
- Blake, J., Austin, W., Cannon, M., Lissus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development*, 17 (1), 91-107.
- Bloom L. (1993). The transition from infancy to language: Acquiring the power of expression. Cambridge University Press.
- Candiotti, A., Zuberbühler, K., Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition*, 15, 327-339.
- Cäsar, C., Zuberbühler, K., Young, R.J., & Byrne, R.W. (2013) Titi monkey call sequences vary with predator location and type. *Biology Letters*, 9, 20130535.
- Chomsky, N. (1968). *Language and Mind*. New York: Harper and Row.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Cleveland, J., & Snowdon, C.T. (1982). The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). *Zeitschrift für Tierpsychologie*, 58 (3), 231-270.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral Brain Science*, 24, 87-185.
- Dediu, D., & Levinson, S.C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology*, 4, 397.
- Everett, D.L. (2009). Pirahã culture and grammar: A response to some criticisms. *Language*, 85, 405-442.
- Fitch, W.T. (2011). The evolution of syntax: an exaptationist perspective. *Frontiers in Evolutionary Neuroscience*, 3, 9.
- Fitch, W.T., de Boer, B., Mathur, N. & Ghazanfar, A.A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2 (12).
- Fujita, K. (2017). On the parallel evolution of syntax and lexicon: A Merge-only view. *Journal of Neurolinguistics*, 43 B, 178-192.
- Gil, D. (2005). Word order without syntactic categories: How Riau Indonesian does it. In: A. Carnie, H. Harley, and S.A. Dooley (Eds.), *Verb first: On the syntax of verb-initial languages* (pp. 243-263). Amsterdam: John Benjamins.

- Greenfield, P.M., & Savage-Rumbaugh, E.S. (1991). Imitation, grammatical development, and the invention of protogrammar by an ape. In N.A. Krasnegor, D.M. Rumbaugh, R.L. Schiefelbusch, and M. Studdert-Kennedy (Eds.), *Biological behavioral determinants of language development* (pp. 235-257). Hillsdale, NJ: Erlbaum.
- Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298 (5598), 1569-1579.
- Hillert D. G. (2015). On the evolving biology of language. *Frontiers in Psychology*, 6, 1796.
- Hillert, D. (2021). How did language evolve in the lineage of higher primates? *Lingua*, 263, 103158.
- Jackendoff, R. (2003). Précis of foundations of language: Brain, meaning, grammar, evolution. *Behavioral and Brain Sciences*, 26, 651-707.
- Lenneberg, E.H. (1967). *Biological foundations of language*. New York: John Wiley and Sons.
- Martins, P.T., Maties, M., & Boeckx, C. (2018). SRGAP2 and the gradual evolution of the modern human language faculty. *Journal of Language Evolution*, 3(1), 67-78.
- Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature* 315, 57-59.
- Mcbrearty, S., & Brooks, A.S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of Human Evolution* 39 (5), 453-563.
- Miles, L.W. (1993). Language and the orang-utan: The old 'person' of the forest. In P. Cavalieri, & P. Singer (Eds.) *The great ape project* (pp. 42-57). New York: St. Martin's Griffin.
- Miyagawa, S., & Clarke, E. (2019) Systems underlying human and old world monkey communication: one, two, or infinite. *Frontiers Psychology*, 10, 1911.
- Pasternak, T., & Greenlee, M. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6, 97-107.
- Patkowski, M. (2014). Looking for structure: Is the two-word stage of language development in apes and human children the same or different? *Studies in Second Language Learning and Teaching*, 4 (3), 507-528.
- Patterson, F.G. (1978). The gestures of a gorilla: language acquisition in another pongid. *Brain and Language*, 5 (1), 72-97.
- Ponce de León, M.S., Bienvenu, T., Marom, A., Engel, S., Tafforeau, P., et al. (2021). The primitive brain of early Homo. *Science*, 372, 165-171.
- Read, D.W., Manrique, H.M., & Walker, M.J. (2022). On the working memory of humans and great apes: Strikingly similar or remarkably different? *Neuroscience & Biobehavioral Reviews*, 134, 104496.
- Rizzi, L. (2016). Monkey morpho-syntax and merge-based systems. *Theoretical Linguistics* 42 (1-2), 139-145.

- Schlenker, P., Chemla, E., Arnold, K., & Zuberbühler, K. (2016). Pyow-hack revisited: Two analyses of Putty-nosed monkey alarm calls. *Lingua*, 171, 1-23.
- Senghas, A., Kita, S., & Ozyurek, A. (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science*, 305 (5691), 1779-1782.
- Stout, D., Hecht, E., Khreisheh, N., Bradley, B., & Chaminade, T. (2015). Cognitive demands of lower paleolithic toolmaking. *PLoS ONE* 10 (4), e0121804.
- Wakita, M. (2020). Common marmosets (*Callithrix jacchus*) cannot recognize global configurations of sound patterns but can recognize adjacent relations of sounds. *Behavioural Processes*, 176, 104136.
- Watumull, J., Hauser, M.D., Roberts, I.G., & Hornstein, N. (2013). On recursion. *Frontiers in Psychology*, 4, 1017.
- Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B*, 375, 20190062.