Language Ecology, Language Evolution, and the Actuation Question

Salikoko S. Mufwene
University of Chicago
Collegium de Lyon

To appear in Language contact and change: Grammatical structure encounters the fluidity of language, ed. by Tor Åfarli & Brit Maelhum. John Benjamins.

1. Introduction

In this essay I argue for an ecological approach to language evolution as one most likely to help us answer the “actuation question” in language change, viz., why a particular change occurred where it did but not elsewhere, and why at that particular time but not at another (McMahon 1994, Labov 2001, Mufwene 2008). The approach will also help us understand under what particular evolutionary conditions various milestone phenomena such as the emergence of phonetic communication occurred during the protracted phylogenetic emergence of language.

My thesis rests on the fundamental assumption that languages are more like viral species than like organisms and that their ecologies, which are both direct and indirect and roll the dice on the specific ways in which linguistic systems evolve, are multifarious (Mufwene 2001ff). Among the most important peculiarities of this position are the following:

1) Inter-individual variation within individual languages and even within dialects and sociolects (henceforth language varieties) is too important to be ignored in any study of language evolution. The notion of SPEECH CONTINUUM inadequately associated only with (post-) creole settings since the early 1970s (e.g., DeCamp 1971, Bickerton 1973; but see Mufwene 2004) applies to all language situations, especially when the reality of IDIOLECT is given more than lip service. We must explain why there is such variation and what role it plays in the evolution process.
2. Consistent with Darwinian evolutionary theory, I also assume that language evolution is variational, like biological evolution, proceeding by competition and selection among variants under particular ecological conditions. That is, it does not proceed by successive and parallel transformations of the kind \( A \rightarrow B \) in the environment \( E \), as traditionally hypothesized in historical linguistics. Rather, it involves competition\(^1\) between alternatives \( A \) and \( B \) (and \( C \)), with \( A \) or \( B \) (or \( C \), or \( A \) and \( B \), or \( A \) and \( C \), or \( B \) and \( C \)) prevailing because it/they was/were favored by particular ecological factors. The selection may also result in some sort of allopatric speciation, whereby variant \( A \) thrives and prevails in one geographical environment or part of the population, while variant \( B \) does in another.

3) Although languages are typically individuated as if they were organisms, especially in historical linguistics and in studies of the phylogenetic emergence of language, they are actually collective creations reflecting the parallel contributions of various individuals that have used them at different times and on different occasions and have contributed differentially to their architectures and feature pools. Because they are internally-variable and are indeed extrapolations from idiolects of particular populations of speakers/signers, they are non-uniform constructs reflecting family-resemblance properties (Mufwene 2001, 2008).

4) I also conceive of languages some sort of folk technologies, albeit as unplanned cultural artifacts, developed and developing incrementally without foresight of the outcome of general architecture. Thus, the innovations they incorporate and the modifications they undergo, typically in ad-hoc ways, reflect particular ways in which their users/creators respond

\(^{1}\) As explained in Mufwene (2008), two or more variants are in competition when one or more of them may be (dis)favored in relation to the other(s) for a variety of reasons intrinsic or extrinsic to them, which may be captured by markedness values. For instance, a lexical item may be disadvantaged because it is less transparent or more difficult to pronounce than an alternative, or because it is stigmatized. Competition applies to any situation where variants for the same function in a system are not equally weighted by the users.
to particular communicative pressures under specific ecological conditions. These are probably better captured by Wallerstein’s (2004) notion of TIMESPACE, which can apply to the *hic et nunc* circumstances of interaction, the immediate ecology of speech acts.

5) Languages are emergent phenomena, in constant states of flux, in constant search of equilibrium, owing to changing patterns of interaction among their users. The phenomenon is also especially evident after the introduction of new elements and/or the elimination of some older ones. This emergentist perspective does not deny the significance of phonological, morphological and syntactic rules. It just invites us to assume that rules are not immutable nor anterior to, or drivers of, language change, despite the established working assumption that classes of the same items behave alike and undergo the same changes. Nor are rules anterior to various phenomena that have shown up during the emergence of language. Instead, they are emergent patterns, as understood in complexity theory and construction/usage-based grammar (see e.g., Croft 2001, Tomasello 2003). They are consequences of processes driven more likely by partial analogies between similar items than by sweeping abstract rules (McCawley 1976). The rules are similar but yet show variation across individuals, because the individual speakers’/signers’ mental and anatomical structures are not identical (despite the prevalence of Cartesianism in modern linguistics), because the producers and interpreters of utterances are not influenced by identical interactional and developmental histories, and thus because the primary linguistic data generating the patterns are not identical from one speaker/signer to the other (Mufwene 2008). Speakers/signers understand each other not because they use identical systems but because similar minds deriving similar patterns from similar data can “read” each other (Mufwene 1989), in the sense of Sperber & Wilson’s (2002)
“mind-reading” theory, which of course makes allowance for divergence or interpretation errors not necessarily due to Chomsky’s (1957) performance factors.

6) It is important to distinguish between, on the one hand, the direct language ecologies consisting of individual speakers and signers and, on the other, the indirect ecologies, which consist first of the “timespace” of utterances, then of population structures which help shape the interactants’ reactions to particular events in specific “timespaces” (Mufwene & Vigouroux, in press), and ultimately of the particular economic practices/systems that produce the relevant population structures. The latter are in turn largely consequences of particular geographical ecologies, consisting of climate and kind of soil, among many other factors which restricted the particular kind of economies that populations inhabiting or relocating to them could develop. Indeed we can speak of a “cascade of ecological determinisms” ranging from the most indirect ecological factors to direct ones.²

7) The phylogenetic emergence of language need not be conceived of as monogenetic. Geographical considerations about the emergence of our hominin ancestors and modern human linguistic behavior militate in favor of polygenesis. We must also bear in mind the evolving, thus diachronically variable, mental and anatomical structures of hominins and humans, which played a very significant role in determining which particular kinds of

---
² As shown by Chaudenson (1992, 2001, 2003) and Mufwene (2001, 2005, 2008, 2009), these considerations go a long way toward explaining the differential evolution of European languages in colonial conditions over the past half millennium, especially why slave and contract labor populations produced creoles only in some (parts of the) colonies but not in others and why, non-creole varieties did not evolve in any more identical ways than creoles did. Besides, the competition-and-selection mechanisms that produced creoles apply in kind also to the emergence of the non-creole varieties, bearing in mind that the massive shift to the dominant language did not take place at the same stage of the development of the colony for all the relevant populations. As shown in Mufwene (2009), the populations who shifted early to the dominant language and were also segregated early from the economically dominant population were more likely to produce a separate divergent variety, at least one that would survive to date, than those who shifted later and were integrated into the economically dominant group.
language(s) could emerge out of the increasingly richer information exchanges, both in contents and in topics, they were engaging in (Mufwene in press).

2. Some consequences of thinking of languages as species

The position that languages are species rather than organisms was first stated by Paul (1880/1991). It was resurrected by Mufwene (2001), who proposed the following justifications:

1) Languages as communal phenomena are not homogeneous; they vary across idiolects, dialects, and sociolects. Although idiolects are internally variable, reflecting the interactional histories of the relevant speakers/signers (having to do with the options available to them in the expression of their thoughts and feelings), languages exhibit the kind of variation that is more typical of species. They can be broken down not only into idiolects (the counterparts of organisms) but also into dialects and sociolects (the counterparts of subspecies).

2) Languages do not change uniformly, as every idiolect evolves at its own pace and every dialect or sociolect qua collection of idiolects also according to its own internal dynamics. Changes affecting some dialects or sociolects may not affect others. Modifications undergone by a particular idiolect do not necessarily change the relevant properties of the dialect, sociolect, and language it is associated with, unless they are shared by many other idiolects.

3) Language and species boundaries (and also those of dialects, sociolects, and subspecies) are fuzzy, unlike those of idiolects. In boundary areas, it is hard to tell one language variety from the other, because interactions in these areas are not consistent with the stipulated boundaries. There are thus some idiolects whose affiliation cannot be determined unequivocally.
4) Indeed, like biological species, language varieties as communal phenomena exist as constructs, as they are extrapolations from the idiolects of their speakers/signers. They have no material existence similar to that of entities or material groupings that can be pointed at.

5) Thus, the structural/architectural characteristics of language varieties are extrapolations from those of idiolects sharing them, on the family resemblance model, just like the genetic and behavioral properties associated with species are extrapolations from those of their member organisms. Phenomena having to do with their vitality, such as sustainability, endangerment, and death, reflect convergent patterns arising from the behaviors of individual members of the relevant populations.

6) Members of language communities do not behave like players in sports teams, with explicit rules of engagement. They produce convergent results because they are endowed with mental capacities that enable them to learn to respond often in similar ways to similar ecological conditions and even to copy from each other. Speakers/Signers of a language variety share with members of a biological species the capacity to respond in similar fashions to similar ecological conditions, though language involves the acquisition of the capacity by learning rather than by genetic inheritance.³

7) Being parasitic on humans, languages are more like viruses than like animals. Their vitality and structural characteristics reflect what the relevant speakers/signers, like hosts of viruses, do as they respond to various ecological pressures, including specific communicative needs under particular pragmatic conditions.

³ For practical reasons, I omit the prerequisite that the language learner be biologically endowed with the capacity to learn human-specific cultural inventions such as language and other technologies that facilitate our species’ adequate responses to diverse ecological conditions.
On the other hand, it is necessary to distinguish communal properties of languages, such as norms and inter-idolectal variation, from individual ones, such as phonetic and lexical idiosyncrasies, as well as grammatical singularities. By mutual accommodation (involving learning from each other), the idiolects of individuals who interact regularly with each other converge toward the same norms, becoming more like each other. Norms must thus be understood as emergent communal patterns rather than as stipulations by an institution (such as an academy) that is vested with authority over the members of a population. Beyond communicative networks, where idiolects are the most similar to each other, dialects and sociolects as emergent phenomena reflect particular population structures, which determine who is most likely to interact with whom and who is most likely to accommodate whom if they stepped outside their subgroups. These considerations apply as much to the evolution of particular structures as to the vitality of a particular variety. From an evolutionary perspective, intent to be easily understood and to collaborate with other members of one’s group (described by Tomasello 2008 and others as “cooperation”) is an important ecological factor prompting speakers/signers to converge in the ways they express their thoughts and feelings.

Regarding all these processes, innovators must be distinguished from followers/copiers (Tomasello et al. 2005; Croft 2000; Mufwene 2001, 2008), who spread features or practices that the former introduce into their interaction networks or communities. This observation highlights the fact that the agency of evolution lies in communicative activities of individuals, as some of them innovate new features or uses of extant ones during specific communicative acts and others copy and spread some that may subsequently be characterized as successful,
because the copiers found them useful. Multiple copying or spreading lays the evolutionary trajectories of features.

Since not every member of a community witnesses the same innovations and since alternative features can be innovated by different individuals for the same communicative needs, the copiers often face options while innovations spread. Depending on various criteria within the population, the options find themselves in competition in the feature pool shared by the overall population. This simply means that the alternative features produced by different individuals are valued/rated differently by different users and/or on different occasions.

Selection, a mechanism of self-organization at the population level, rolls the dice on the fate of the variants/alternatives. It determines which variant(s) prevail(s) over which other(s), or it establishes a division of labor between them, which is a form of specialization known in linguistics as complementary distribution.

These dynamics make of languages what complexity theorists identify as “emergent phenomena,” always in search of equilibrium (Dooley 1997; Heylighen 2008) or “coherence” (Steels 2000), with patterns emerging that are likely to change afterwards in response to changing ecological pressures. There are two main reasons why languages share these emergence characteristics:

1) There is no transmission comparable to the process in biology whereby genes are passed on by the carrier to a passive offspring or receiver/heir at conception, i.e., without any agency on the part of the heir (Fracchia & Lewontin 1999). Cultural, hence linguistic, transmission is thus a misnomer for learning by inference (Mufwene 2008), which is prone to perception and/or interpretation errors, part of what has been identified in diachronic studies as “imperfect
replication” (Lass 1997: 112), “transmission error” (Deacon 1997: 114), or more often as “transmission with modification.” Unlike in biology where, barring mutations, trans-generational changes are consequences of how genes are recombined and weighted within the new recombinations, linguistic materials that recombine are typically not faithfully reproduced in the first place. Aboh’s (2009) hybridity hypothesis captures this phenomenon adequately, showing how, in the formation of creoles, materials from both the lexifier and the relevant substrates that are selected into the emergent systems are seldom faithfully reproduced with all nuances and constraints associated with them in the source language(s). Since individual learners do not operate identically and must learn from each other during typically changing dyadic or triadic interactions, the constant adjustments to one’s interlocutors can both produce convergence and disrupt the current equilibrium in the communal system.

2) The other reason is that speakers’/signers’ communicative needs are not always exactly the same. However, communicators cannot always innovate new forms or structures to convey new meanings. They often have to exapt current forms or structures to do the new job, thus innovating in a rather abstract way by expanding the domain of use of current words or rules. These are mechanisms that keep linguistic systems in a state of flux.

In language communities, speakers/signers also alternate their roles as innovators or copiers and spreaders, as nobody has the privilege of innovating for all the others and nobody is lazy enough to wait for others to dictate to them how they must express what’s on their minds. Copiers also act as agents of the selection process, though they do not do this with foresight of the ensuing structures or architecture. Their choices are always in the present “timespace,” although they are informed by their past experiences. As they may change several times,
depending on varying pragmatic conditions, this is when we have to invoke self-organization as the producer of emergent patterns and/or structures in a language variety.

The way in which intergroup variation has arisen in modern languages, especially between those that emerged recently such as creoles and pidgins, suggests that there is no particular reason for preferring a monogenetic account of the phylogenetic emergence of languages over a polygenetic account. We can very well speak of the evolution or emergence of languages, in the plural, instead of the evolution or emergence of language, in the singular or semantically non-individuated delimitation, bearing in mind that in the former case the emphasis is on individual languages but in the latter it is on the common properties of languages or the essence that they share. Every language variety has emerged locally in response to local ecological conditions, although these are probably also responses to other ecological conditions on a larger scale, such as climates and soil in a colony conditioning a particular economic system and the latter leading to a specific population structure, as explained above regarding the proposed “cascade of ecological determinisms.”

Reasons for speaking of the evolution of languages (in the plural!), and of the evolution of language only when we want to focus on their common architectural characteristics, are probably as good as, if not better than, for speaking of the evolution of an ancestral language that speciated after the exodus of modern Homo sapiens out of East Africa. An account of how linguistic diversity arose from a polygenetic emergence of languages is probably easier and more plausible than the alternative of explaining it from a monogenetic account. The available paleontological evidence does not suggest that modern humans emerged in one single specific location in East Africa, nor that they waited until they had invented a common language before
dispersing, contrary to what is suggested in the Christian book of *Genesis* (Mufwene, in press).

Although this observation makes it more challenging to explain why they all developed similar technologies for communication, it does not make it compelling to adopt monogenesis to account for the common properties of languages. The following discussions will shed light on the direction of these conjectures.

3. **Constraints on innovations and exaptations**

Communicators are not free to innovate any kind of form or structure, nor are they free to package information in any conceivable chunks. They are severely constrained by their current command of the means they use to communicate and by what they know already. Generally, speakers/signers are constrained by their command of the current structures of their language variety (an internal ecological constraint) and by their cognitive capacity (a direct external ecological constraint). For instance, infants are not capable of communicating the same cognitive contents as their caretakers (at least in quantity and range of possible information), because their cognition is not as mature as those of their caretakers, they do not have as much experience with their environment as the latter (which is very useful), and the set of communicative pressures on them are not the same as on adults around them. Their inaccurate reproductions of the input systems are initially determined by the stage of their cognitive and anatomical maturation (which bear on dexterity), not by differences between their mental capacity cum endowment and that associated with the modern Homo sapiens. (The relationship between ontogeny and phylogeny is precisely what has led to emergence of the Evo-Devo theory in biology!) The faithfulness of the reproductions is also determined by
another important ecological factor, viz., the children’s interactional histories (Mufwene 2008, Dor & Jablonka 2010). Familiarity with particular signals just makes it easier to reproduce them. On the other hand, the populations who produced creoles and pidgins were, collectively, largely constrained by the language variety they intended to speak, which, depending on how much exposure they had had to it, limited the vocabulary and the related rules of grammar they could target. Another important ecological constraint stems from the languages they knew already, which not only affected their perceptions and interpretations of utterances in the target language but also provided alternative forms or structures to use instead of those of the target language, especially those to which they had not had sufficient exposure to learn. The ratio of the constraints varied within each population, depending on whether the interactions between the different ethnolinguistic groups in the contact setting were regular or sporadic, whether the learners were young or adult, who the model speakers were, etc. However, the changes that produced creoles and pidgins had nothing to do at all with, from a phylogenetic perspective, (contact-induced) changes in the mental capacity of their producers (Mufwene 2005, 2008). Slavery and economic destitution did not reduce the mental capacity that all adult specimens of modern Homo sapiens are endowed with to learn a second language, subject to degree of exposure, individual motivation, and personal facility for language.⁴

As can also be learned from the literature on grammaticization, the modifications made in response to the new communicative pressures are constrained not only by what the

---

⁴ These considerations should cast doubt on Bickerton’s (1990, 2010) claim that (incipient) pidgins are among the fossils of his phylogenetic “protolanguage.” However, the topic will not be pursued here, as it has been dealt with in Mufwene (2010).
communicator intends to express but also by what the grammar being exapted allows. For instance, the grammaticization of the adjectives *able* as a mood marker is constrained by the fact that it is an adjective; thus, while it can be exapted to function as a deontic modal, it should not violate the norm that it must combine with a copula when it is used predicatively. It continues to behave syntactically as an adjective even in its new semantic function. The same is true of the syntax of the future auxiliary *going*, which still behaves like a main verb in negative and/or interrogative constructions, largely because it is in the progressive form and is therefore less verby. It needs its own auxiliary qua copula to carry the negation marker and make up for the constraint against its participation in the Subject/Auxiliary Inversion convention. In other words, there is a significant amount of self-scaffolding (Wimsatt & Griesemer 2007) in the exaptations that speakers/signers make in their languages.

Such ecological considerations put the phylogenetic emergence of language in a perspective that is especially informative. Throughout the protracted and incremental emergence of languages as communicative technologies, our hominin ancestors could not produce anything that was not permitted by their current mental and anatomical structures and by the means they were already using. The emergence of language is a protracted history of successive exaptations. Though we do not know for sure when the mind that was capable of driving language evolution emerged, paleontological evidence suggests that the emergence of specific anatomical constraints probably delayed that of the particular linguistic technologies that modern Homo sapiens now uses, particularly the supralaryngeal vocal track (made possible by the descent of the tongue root, Lieberman 2007: 46, Fitch 2010: 3120), which made possible the production of large inventories of phonetic sounds and called for combinatorial constraints
known as phonology and syntax (Mufwene, in press). As pointed out by Maddieson (2006), there must be good ecological reasons why only the supralaryngeal structures are used to produce sounds (but not the fingers or feet) and why, although there is significant typological variation across languages, the sounds produced are distributed only within a definite range of formant frequencies and variants of some core vowels (such as /i, a, u/, see also Fitch 2010) and some stop and nasal consonants tend to occur in all languages.

4. An ecological perspective on the phylogenetic emergence of language

This discussion cannot proceed without determining, especially at this juncture, what are the ecologies that bore on the phylogenetic emergence of language conceived of in this essay as hybrid, physical (phonetic or manual) and mental, technology architected modularly (according to some universal and community-specific principles) for transmitting information from one individual or party to another. Focusing on speech, it appears that it would have been difficult for hominins not endowed with modern Homo sapiens’ anatomical structure to develop any variant of modern spoken languages. This is not to say that no language would have evolved at all. As is evident from signed language, alternative forms of modern languages could have emerged. A relevant question is why just these two, spoken and signed languages as autonomous means for articulate and elaborate communication. Another is why speech has prevailed over signing in all modern human populations. These questions are made more significant by Corballis’ (2002) hypothesis that communication with gestures, though probably not in any form similar to signed language, may have started as early as Homo habilis, who had hands similar to modern humans and could use primitive tools but whose buccopharyngeal structure was not the same as that of modern humans. If that gestural language was not in any
way close to modern signed language, an important question is whether Homo habilis also had the same mental capacity and lived under the same kinds of social pressures as modern Homo sapiens. According to several scholars, including Lieberman (2006), Tomasello (2008), Tomasello et al. (2005), and Fitch (2010), these were critical prerequisites to the emergence of modern language.

Because communication transmits some contents (information about the world or about some condition as constructed by a particular kind of mind) and to an individual or group thereof sharing with the communicator the ability to construct knowledge of the world in similar ways, modern mental and social structures have been treated as prerequisite ecologies to modern language. Corballis’ (2010) and Lieberman’s (2010) self-corrections that modern spoken language probably emerged only 50,000-30,000 years ago are based on such considerations. If the modern human’s anatomical structure was in place since the divergence from Homo erectus, then the emergence of modern language must be the consequence of the emergence of the modern mind around the same time or later.

One may thus also assume that the complexity of the architecture of modern language reflects that of the mind that produced it. The same mind must also account for the emergence of more complex social life that called for the particular kind of complex communication to regulate the interactions between family members and between the larger community of people sharing the same habitat. The literature has rightly focused on joint attention and cooperation, which fosters negotiations through explicit exchanges of thoughts and points of views, rather than competitiveness, which privileges use of brute force (and trickery among nonhuman primates) for particular individuals to prevail. Thus, the structural simplicity
associated with incipient pidgins and child language (McWhorter 2001, Bickerton 1990) has to do not with the basic modular architecture of language which they share with other languages but with the limited frequency of social interactions and topics to discuss in the case of pidgins and with the limited extent of cognitive maturation and limited number of topics of communication in the case of children. Studies such as Bakker (2009) now show that the phonetic inventories of pidgins fall within the average-size range among the world’s languages and are generally not significantly reduced relative to their lexifiers. We return to the complexity aspect of language below.

A noteworthy observation was made by Darwin (1871) regarding the parrot’s ability to mimic phonetic communication without being able to produce original utterances. As Mufwene (in press) states it,

(...) parrots cannot speak (i.e. produce original spoken messages intentionally), although they can imitate human speech fairly accurately. Showing what an important driver role the human mind has played in the evolution of language, [Darwin] argued that it was for the same reason that other primates do not use their buccopharyngeal structure to speak.

This observation supports the hypothesis that the modern mental capacity, which a modern human child is born with but must mature (according to a biological blueprint) before performing some of the tasks, was an important ecological factor in the emergence of language. Small though the genetic difference is between chimpanzees and modern humans (less than 2% according to the literature), the implications from the differences are exponentially big when it comes to human languages and other aspects of human cultures. The
differences are comparable to several, incremental and minor modifications of computer designs that have improved the task capacity and performance of each new generation of computers so significantly compared to their predecessors (Mufwene 2010). Evolutionarily, the modifications that our minds and anatomies have undergone over the past 6-4 million years, as opposed to those undergone by other primates, account ecologically for differences between human communication systems and the means of communication used by the others. Modern human minds have created more complex social lives and more diversified social pressures for communication, which triggered the emergence of more complex communicative systems, all endowed with the capacity to adapt to new situations. The cultures that the other primates have developed from their respective social life styles, products of their particular mental structures, led them to produce the specific means that they have produced, which, according to some experts, are limited to finite sets of messages (such as alarm calls, intimidation grunts, expression of fear or submission, signaling the location of food, and invitation to play or to mate, among primates).

Bearing the above considerations in mind, we can focus on particular ecological pressures that hominins must have experienced toward the emergence of human language(s). The parrot example also highlights the fact that speech need not be produced by the kind of buccopharyngeal structure that modern humans are endowed with. In particular, the (African) parrot has no teeth and no lips, but it has a syrinx sunk deep in its throat in lieu of the mammalian larynx and can mimic speech, although other birds likewise endowed with a similar anatomical structure cannot do the same. This suggests that the parrot is endowed with some mental capacity for phonetic imitation that other birds do not have. This also means that
humans equipped with the same minds but with a different anatomical structure for producing phonetic sounds would have developed the same kinds of language(s).

On the other hand, we cannot use our armpits to produce speech, contrary to what some African myths report about putative strange cannibalistic populations that children may encounter if they venture too far in the forest. Neither the story-tellers nor those who grew up with me who often ventured deep into the forest or in foreign territories never encountered people who used organs other than their buccopharyngeal structures to speak. Maddieson (2006) is right on the mark in pointing out that hominins/humans co-opted the supralaryngeal or vocal tract to develop the phonetic sound technology because they could thus produce a much wider range of distinct sounds (which can also be modulated by prosody!) than they could with foot stamping, finger clicking, hand clapping, or air squeezing under armpits. These considerations prevail over other practical advantages such as freeing the hands for other activities (e.g., carrying tools or pointing) while communicating, although they are significant too.

Evidently, hominins/humans domesticated the organs that were the best suited for the phonetic technology. As much as they can vocalize or use their limbs, the other primates did not evolve the kinds of minds and social cultures that would prompt them to domesticate the same organs toward similar communicative technologies (Darwin 1871). Thus, since our hominin ancestors’ anatomical and, arguably, mental structures were not like ours, we cannot assume that they could, at any stage of our phylogenetic evolution, have developed a means of communication close or similar to modern human language(s). The same minds that invented phonetic technology also produced representational systems (especially in the form of cave
drawings/paintings) that would suggest that Homo habilis or Homo erectus, like Homo habilis before him, was not capable of holding complex communication through signing. No such evidence has been found, unlike that which, combined with bone fossils, can be used to infer the probable time of the emergence of phonetic language. May sign language itself have emerged about the same time as speech?

The distinction between the evolution of the hominin-to-human mind and that of the anatomical structure as separate evolutionary trajectories and ecological prerequisites is an important heuristic dimension missing from the earlier speculations on the emergence of human language from natural cries, from antiquity to the 19th century. Darwin’s (1859, 1871) theory of evolution by successive mutations favored by changing ecologies and by competition and selection among competing variants has helped redirect the scholarship on the subject matter of the evolution of language, which has inspired the present recasting of the scenarios from an ecological perspective. Accordingly, I submit that the human anatomy and mind are the most fundamental ecological factors that have borne on language evolution, phylogenetically, historically, and developmentally. They filter, through the interpretation that every individual makes of the pressures coming from the social environment and how he/she responds to them, all other external ecological factors that bear on language evolution.

Bipedalism has been invoked typically for the greater terrestrial mobility it allowed hominins as hunter-gatherers to have and the more extensive usage they could make of their forelimbs (arms/hands) then freed to use tools, gather food, and carry it. These advantages apparently became more pronounced with Homo erectus, who could travel longer distances and even migrated out of East Africa. As suggested above, bipedalism bore as an ecological
factor in favoring the selection of phonetic technology as the preferred means of communication. However, little has been said about the anatomical reconfiguration it triggered in the Homo genus (compared to other animals), especially in the basicranial and the pulmonary structures that would eventually make speech possible by the time of modern Homo sapiens 50,000-30,000 years ago (Lienerman 2002, 2010). It is not sure when the descent of the tongue root and larynx occurred, and whether the structural change occurred in one phase or in stages, as in the human ontogenetic development. It is, nonetheless, worthwhile determining why modern speech (different from language) did not emerge until so recently in our phylogeny. What was missing? I conjecture the mind; but could there have been other reasons?

It is also commonly assumed that some sort of social life, perhaps initially comparable to that of nonhuman primates such as chimpanzees, must have been in place since those early beginnings about 3 million years ago. It would have exerted pressure to communicate, in order to organize communal life (e.g., Corballis 2002, Lieberman 2006). This has led Tomasello (2008) to underscore the role of joint attention. One can safely conclude that as social life became more sophisticated, communicative systems that evolved to meet its increasing demands for more explicit, elaborate, and/or effective communication (Rousseau 1775) also evolved more sophistication or architectural complexity. However, this conclusion says nothing about the particular technology, manual or phonetic, that was developed in response to this ecological pressure. As pointed out above, scholars such as Corballis (2002), MacWhinney (2002), and McNeill (2005) have argued that the earliest forms of hominin communication could not have been phonetic, because, according to them, Homo habilis’ anatomical buccopharyngeal
structure, which was similar to that of today’s nonhuman primates, could not produce the phonetic technology required for it.\textsuperscript{5} It is necessary here to distinguish vocalizations, which are continuous but not digital or discrete, from phonetic sounds, which are just the opposite and allow what MacNeilage (2008), anticipated by Rousseau (1775), calls “syllabic variegation.”

On the other hand, along with Lieberman (2002), Corballis had argued earlier that phonetic communication, though not necessarily equipped with modern syntax, could have started as early as the late Homo erectus, about 500,000 years ago. Both have recently retracted this position, arguing that the neck of Homo erectus and early Homo sapiens was too short (Corballis 2010, Lieberman 2010). According to the Lieberman,

neck lengths that would support a fully human SVT are not apparent in the fossil record until the Upper Paleolithic, some 50,000 years ago (...) the sudden appearance of an array of advanced artifacts has been taken to be a sign of cognitive advance. [...] The presence of a human SVT in a fossil hominid can be regarded as an index for the reiterative neural substrate that makes voluntary speech possible (Liberman 2010: 175).

As pointed out above, the capacity to produce speech was in itself a consequence of the descent of the tongue root (Lieberman 2007: 46, Fitch 2010: 3120), which, in earlier literature, had been assumed to have happened much earlier in the hominin phylogenetic evolution, even 2 million years ago according to some estimates. Assuming now that the descent occurred later, as suggested by Lieberman’s observation, nothing could really have happened before this anatomical change.

\textsuperscript{5} In this context, what underscores the significance of the modern human mind as an important ecological factor is the fact that nonhuman primates have not developed a signed language comparable with those developed by humans.
On the other hand, it has also been observed that there is a flaw in this reasoning. The larynx is deeper in adult than in pre-adolescent humans, and it descends deeper in male than in female adults. Yet these anatomical differences cannot be correlated with richer phonetic inventories in those who have a more deeply-descended larynx. This counter-observation also suggests that the “neural substrate” to which Lieberman also draws our attention had a more significant role to play. It also highlights the ecological significance of social life. Men interact with women too. There would be no reason for one of the genders to develop a richer phonetic inventory for communication within a mixed-gender community. Overall, as pointed out by Maddieson (2006), most languages have phonetic inventories that fall within the middle range (15-25 consonants and 5 oral vowels), as opposed to those that have small inventories and those that have very large ones.

The role of the mind should not be underestimated. It accounts for the emergence of complex human cultures, emerging from complex human interactions, which linguistic communication must match. One can argue that complex vocabularies reflect the complexity and diversity of information to be conveyed by users of the language technologies. The richness of information calls for complex but economical ways of distinguishing one symbol from another (thanks to recursion and constrained combinations of limited numbers of units), without making the system too difficult to learn and use. Thus, by the convergence of the systematic and mutually accommodating behaviors of individual speakers/signers, norms emerge which produce community-specific selections of phonemes or manual units and principles (phonological, morphological, and syntactic) for combining them into larger meaningful sequences: words, phrases, sentences. Although combinatorics, which allow
recursion, are a consequence of linearity of the phonetic technology, it took a special mind to
coop our basic ability to vocalize and gesture (which we share with nonhuman primates) and
develop more complex communicative technologies.

Typological variation suggests that as communal systems linguistic technologies of the
world need not have originated monogenetically. Chances are that, given the wide range of
phonetic options that can effectively be used by humans, each hominin colony that produced a
language (in response to the range of its communicative pressures) could have selected any
subset of phonetic or manual units. It so happens that there are some core vowels, consonants,
and manual units that are attested in all languages. May this suggest that phonetic inventories
evolved protractedly but no population could escape the fundamental evolutionary trajectories
that were constrained by the evolution of our anatomical structures? Are there fundamental
sounds that every population just had to start with? Assuming that, as in child language
acquisition, phonetic acquisition is concurrent with lexical acquisition, could it also be that the
phonetic inventories of various populations increased concurrently with increasing needs for
richer communicative systems possessing larger vocabularies? Typological variation would have
arisen from different populations responding to communicative pressures in alternative ways
past the basic set of core sounds, like the cardinal vowels ([i], [u], [ɛ], [o], [a]) and the most
common consonants. While the emergent systems all increased their phonetic inventories, they
need not have innovated the same additional sounds nor in the same numbers. As pointed out
by Maddieson (2006), variation in the size of the phonetic inventories has no bearing on the
size of the vocabularies that different populations can produce. Different populations did not
have to settle on the same norms for combining the sounds either, though there are
combinations that must be favored as opposed to those that populations tend to avoid because they require more energy or are found too difficult.

When morphology emerged is another question altogether, though we can assume safely that richer communication must have entailed more than continuing to name novel entities and states of affairs (events and conditions). Various populations must have found it helpful to identify some minimal meaning-bearing combinations of sounds that can be used over and over again to generate new words that can be easily understood by hearers that were not familiar with them. Eventually, pressures had to arise to go beyond naming and to predicate about entities and conditions, thus to communicate in sentences, and to distinguish between ways of referring and situating states of affairs in time and space. The reality of typological variation among the world’s languages also suggests that there were variable options; different populations settled arbitrarily on subsets of those. An interesting question from the perspective of Universal Grammar, regardless of whether or not one assumes that it drove the emergence of language (Mufwene in press) is to what extent the human mental/cognitive infrastructure constrained the kinds of syntactic phenomena that occur in the world’s languages. What role did it play in facilitating the emergence of complex sentences: what did it rule out as unworkable?

5. Conclusions XXX

6 In this respect, language communities do differ on whether they prefer derivations (by prefixation, suffixation, or infixation), compounds, or a combination of both strategies. The basic principle is still the same: forming larger meaningful units from smaller meaning-bearing ones.

7 For an idea of how part of this evolution may have happened, see Heine & Kuteva (2007), though it is debatable whether the whole grammar emerged by repeated processes of grammaticization. It is difficult to invoke this process to account for, e.g., clause subordination and predicate serialization to form complex sentences, or for the use of adjectives and relative clauses to modify head nouns.
The main purpose of this essay was less to explain how language(s) may have emerged phylogenetically in mankind than to focus on anatomical, mental, and social ecological factors that bore on this evolutionary process as facilitators or constraints. We thus focused on the actuation aspect of the protracted evolutionary process. A fundamental assumption that has threaded throughout is that communication preceded language; it would just complexify in response to pressures emanating from hominins’ greater mental capacity and more complex social life. Domesticating their own anatomies, hominins/humans developed phonetic and manual communicative technologies that would eventually evolve into modern languages. Spoken languages could only incorporate sounds and combinations thereof that hominins/humans could produce and perceive/process easily. Similar considerations must bear on the particular selections of manual shapes and movements that make up signed languages. We must bear in mind that combinatorial constraints (including phonology and syntax) are largely consequences of the linearity imposed (to different degrees) by the physical properties of the media used (Mufwene, in press).\(^8\)

Evolutionarily, we may argue that the components of the linguistic technologies did not all emerge at the same time, not even the phonetic inventories that made spoken languages possible. They must have emerged at different times, depending on the specific stages of hominin/human anatomical and mental evolution that facilitated the production of particular features in the technology, for instance, predication, mode of reference, modification of the head noun or of the predicate, and production of complex sentences, all of them consequences\(^8\)

---

\(^8\) Signing is less constrained by linearity than speech, because several components can be used concurrently, such as hand shape, motion, direction, and plane. This also makes up for the speed of articulation which, according to Corballis (2010) would have favored speech (Mufwene, in press).
of the symbolic nature of phonetic or manual “signifier,” to use a convenient Saussurean term. The phonetic inventories themselves may have evolved incrementally, to facilitate the creation of larger vocabularies, as human populations became more adept with the emergent technologies.

One can conclude that languages are the kinds of technologies that reflect what human anatomies and mental structures are capable of in response to all the pressures for communication that increasingly more complex social structures, phylogenetically speaking, exerted on hominins and humans. If it is true that modern human language(s) did not emerge until 50,000-30,000 years ago, then can we not conclude that, because the modern anatomy has remained almost the same since the emergence of Homo Sapiens some 200,000 years ago, humans had to wait until the late emergence of a modern human mind for modern language to emerge?

References:


Corballis, Michael C. 2010. Did language evolve before speech?, In Larson et al. eds., 115-123.


Lieberman, Philip. 2010. The creative capacity of language, in what manner is it unique, and who had it? In Larson et al., eds, 163-175.


