Human self-domestication and the evolution of prosody

Antonio Benítez-Burraco¹, Wendy Elvira-García²

¹Universidad de Sevilla
²Universitat de Barcelona

abenitez8@us.es, wendyelvira@ub.edu

Abstract

Human self-domestication refers to a new evolutionary hypothesis. According to this view, humans have experienced changes that are similar to those observed in domesticated mammals and that have provided us with many of the behavioural and perhaps cognitive pre-requisites for supporting our social practices and advanced culture. At the core of this hypothesis is the claim that self-domestication is triggered by a reduction in reactive aggression.

Since the findings of increased complexity in the communicative signals of domesticated animals compared to their wild conspecific, the human self-domestication hypothesis has been used to account for the sophistication of the grammars of human languages. Nonetheless, less research has been done in the domain of phonology. In this talk, we apply this evolutionary model to the evolution of human prosody, arguing for a progressive complexification of prosody that parallels (and is triggered by) the complexification of grammar, also in response to a reduction in reactive aggression levels. Two different types of evidence support our claim: the parallel complexification of prosody and grammar found in emerging sign languages and the parallel sophistication of prosody and grammar during language acquisition, which in turn parallels an increased control over the mechanisms involved in reactive aggression.

Index Terms: prosodic evolution, prosodic acquisition, sign language prosody, self-domestication.

1. Introduction

The mainstream view in language evolution studies has been that once our distinctive cognitive and behavioural abilities emerged, modern-like languages emerged too. Chomskyan approach to language evolution (e.g. [1]) nicely exemplifies this view. Over the years, however, it has attracted a strong criticism, not only because of its claim that language appeared suddenly as a result of one single gene mutation and has not changed since its invention, but mostly because it teases apart the evolution of language as a faculty from the evolution of languages, that are conceived just as hy-products of that faculty. Ongoing research on different fields, from paleoanthropology, to neuroscience to language typology, suggests instead that our cognition and behaviour might have changed after our inception, that language features impact on our cognition and behaviour (pretty much as language features depend on our cognitive and behavioural features), and that aspects external to language (physical, but mostly social, particularly the way in which languages are learned and used) impact on language structure [2], [3].

Overall, a more nuanced view of how language might have evolved is that it resulted from a mutually reinforcing feedback loop between changes in our cognitive architecture and behaviour, and changes in environmental and cultural factors. This is why current language evolution studies are particularly interested in approaches that consider biology and culture on a par. The hypothesis of human self-domestication (henceforth, HSD) as applied to language evolution is in line with this paradigm shift in language evolution studies, as HSD is expected to result in biological changes that favoured the creation of the niche that enables the emergence of aspects of language complexity via cultural evolution.

2. The evolution of language under the self-domestication hypothesis

In brief, HSD refers to a recent hypothesis about how our species emerged. It claims that the human distinctiveness is, to a large extent, the outcome of an evolutionary process similar to animal domestication [4]. In mammals, domestication is usually triggered by selection for tameness and in most cases, it results in a constellation of distinctive traits that are physical, cognitive, and behavioural by nature. According to some views, this is due to the fact that tameness reduces the input to the neural crest, an embryonic structure giving rise to many different body parts during development [5]. The hypothesis of HSD builds on the finding in humans of many of the traits commonly found in domesticated varieties of mammals, including reduced skulls/brains; childish facial features; reduced hair; prolonged childhood; increased playing behavior, and particularly, a less aggressive behavior (see [3] for review). Among the factors commonly cited in the literature that might have triggered our self-domestication, one finds the rise of community living, the advent of co-parenting, changes in our foraging ecology, and the increasingly harsh environments resulting from Last Glaciation. All these factors seemingly promoted a selection toward less emotionally reactive partners and toward tolerance for extra-group individuals. In turn, the physical, behavioral, and even cognitive changes brought about by HSD might have favoured the emergence of many human-specific distinctive features, including our enhanced social cognition, increased cooperation and extended social networks, and ultimately, our advanced technology and sophisticated culture [4].

However, it was the finding that in some birds, domestication results in more complex communicative signals that paved the way towards claims that HSD could be extremely valuable in capturing key aspects of the evolution of language, specifically, the aspects that are thought to emerge through a cultural mechanism [6]. Recent work [7], [8] outlines a model of how HSD might have contributed to the
evolution of language (and of languages), which encompasses four stages.

Stage One in the model corresponds to the beginnings of HSD, roughly 200 kya. Because reactive aggression was still high, communication through language could not have involved patient and cooperative turn-taking, using long utterances, but just single-word commands, threats, and exclamations, mostly aimed to convey emotions.

As features of HSD, the cultural transmission process that fuels the sophistication of linguistic structures started to be enhanced as well. It is believed that at this early period this was made possible mainly by the principal physiological and behavioural outcome of domestication: a reduced reactive aggression. This was surely crucial for establishing stronger in-group networks, involving more diverse, frequent, and prolonged contacts between members. As a consequence, both language structure and language use became more sophisticated. However, we think that changes in the management of aggression might have impacted as well on our brain structure, and ultimately, on our language processing abilities. One reason is that an increased control of aggression demands an increased connectivity in selected cortico-striatal brain networks. However, these networks are also involved in cross-modality, and language processing more generally (see [9] for details). This is the Stage Two in the model, which could have spread between 200 kya and 50 kya. For this Stage Two it is hypothesised that single word utterances started to be combined in a pair-wise fashion, basically leading to rudimentary two-slot grammars that would have used nouns and verbs to express predications. It is also stated that these early grammars might have been primarily used for creating particularly colourful derogatory expressions, and ultimately, for replacing physical (reactive) aggression by verbal aggression. This would have resulted in an accelerated feedback loop between these early forms of grammar and HSD. The ultimate reason is the existence, as discussed earlier, of a common underlying mechanism supporting three core dimensions of language and of language evolution: cross-modality, aggression, and language processing.

At some point around 50 kya, HSD might have reached its peak, as suggested by paleoanthropological evidence. This is the starting point of our Stage Three. The highly reduced reactive aggression (facilitating more frequent and more diverse contacts, resulting in enhanced teaching and learning) together with the extended juvenile period also brought about by HSD (resulting in an extended learning period, and more practising through play) gave rise to more sophisticated forms of grammar, specifically, to the first hierarchical grammars expressing transitivity. In our view, the most confident proxy of the languages spoken during Stage Three are the languages spoken by present-day hunter-gatherer societies, and more generally, what has been called esoteric languages. These languages, spoken by isolated human groups living in small, close-knit communities with high proportions of native speakers, typically exhibit larger sound inventories and complex phonotactics, opaque morphologies (with more irregularities and morpho-phonological constraints), limited semantic transparency (with abundance of idioms and idiosyncratic speech), reduced compositionality, and less sophisticated syntactic devices.

As population size increased, and seemingly also because other changes (like changes on human foraging ecology or climatic changes), inter-group contacts generalized, and extensive social networks emerged, relevant for trading and mating. As a consequence, the necessity of exchanging de-contextualised meanings and know-hows with strangers also increased. We believe that this favoured the emergence of a second type of complex languages, which are sometimes called exoteric languages. These languages feature expanded vocabularies and increased syntactic complexity (including greater reliance on recursion), as well as greater compositionality and enhanced semantic transparency. By contrast, they exhibit simpler sound combinations and more regular morphologies. This is Stage Four, whose starting point has been tentatively situated 10 kya, during the transition period from the Paleolithic to the Neolithic. The advance of exoteric-like languages can be linked as well to the advent of new forms of aggression, specifically, proactive (that is, premeditated) aggression, that became generalised during this period. Increased proactive aggression has been claimed to result from group selection in favour of risk-prone altruism. Our contention is that it might have been facilitated by the emergence of exoteric-like languages, which are especially fit for conscious planning, thus potentially contributing to large-scale hostilities and escalated battles, and ultimately, to the emergence of cultural institutions around war and peace in complex societies.

In summary, this model of language change in Prehistory ties the different stages of the evolution of languages with changes in the management of aggression, either reactive or proactive, and ultimately with the behavioural and cognitive changes brought about by HSD, with both aspects, namely, language features and HSD, being engaged in a mutually reinforcing feedback loop, as both depend on, and impact on a common neurobiological substrate. Because the prosodic bootstrapping hypothesis proposes that phrasal prosody helps infants to discover the syntactic structure of sentences (e.g., [10]), our main claim in this paper will be that prosody evolved in parallel with syntax, seemingly facilitating the acquisition and use of increasingly complex form of grammar.

3. A model for the evolution of prosody under HSD forces

In the following paragraphs, we try to give an account of what prosodic evolution would have developed considering the four stages presented in section 2 and making a parallelism with syntax development in language acquisition and emergence of sign languages when data is available.

3.1. Stage 1: Prosody in a protolanguage

For the first stage of language development, [8] proposes the existence of a rudimentary language characterized by the expression of emotions expressed through word commands, threats and exclamations possibly made of one word. We cannot produce sounds without prosody. Even the simplest vocalization will have duration, intensity, pauses imposed by the necessity of breath and arguably tone. Therefore, these acoustic features of prosody must have been present in this stage. And, therefore, we would have utterances (intonational phrases) made of prosodic words.

Bibliography suggests than at this stage both monosyllabic and multisyllabic words could have been possible. In the case that words were multisyllabic, stress patterns (lexical stress, lexical pitch contours or lexical tones, depending on language) must have arisen.
Stress patterns in language acquisition are one of the features that is perceived earlier (with timing). Children show preference by the pattern of their language as early as 6 month and we can find discrimination of stress patterns in babies 9 months old [11]. However, we cannot know if these protolanguages would have had a fixed stress pattern or variable and therefore phonological.

3.2. Stage 2: Two-slot grammars

Moving forward to two-slot grammars we find utterances made of more than one word. From a perceptual perspective, it is well known that children at the one-word stage already understand two-word commands [12]. For prosodic processing, having utterances made of more than one word means that, firstly, word segmentation is required.

The usage of prosody as a cue for word segmentation can be found in babies between 6-12 months of age [13]-[16]. And according to some research even in newborns [17].

3.3. Stage 3: transitive vP stage

First transitive constructions imply first hierarchical grammars expressing transitivity. At the same time, hierarchical grammar is known to be bootstrapped by hierarchical syntax meaning that speech would have acoustic cues for this hierarchy.

For this stage, we specifically expect to find some of the universal cues of prosodic structure (phrasal prosody), such as pauses and phrase-final lengthening given that in language acquisition those are used earlier than language-specific ones.

For example, it is known that 18 months exploit the prosodic structure of a sentence to assign syntactic constituents, they can use this info to help them assign category (e.g. noun vs. verb) to a novel word [18], [19]. Moreover, several studies have proved that babies can distinguish between languages with different word orders (head-initial languages versus head-final languages) using prosody [20].

Other than that, we cannot know which specific cues for phrasal prosody those languages would have had. At this stage, intermediate phrases are not required, meaning that in this stage phonological phrases would have been also intonational phrases.

3.4. Stage 4: Two-slot grammars

In stage four, we find a fully-fledge grammar system which implies complex sentences. From a prosodic perspective, that requires the existence of both intermediate phrases and intonational phrases, i.e. the complete prosodic hierarchy that we find in modern languages.

From an acquisition data, we know that complex sentences are developed later that simpler ones. For example, children begin to use embedded noun phrases between 4 and 6 years old [21]. Furthermore, in the emergence of signed languages specific intonational patterns used in subordination (for example, conditional intonation) has been observed in second generation speakers [22], [23].

Furthermore, given the growing in population size, it is also logical to think that some turn-taking mechanisms would have developed. And we know the 3 years old can use prosodic information order to refine their semantic/pragmatic interpretations (e.g. in ellipsis) [24], that could have been also the case for prehistoric languages.

4. Discussion

If we take into account the age of acquisition of each of the prosodic milestones reviewed in this study, we can see that with every new stage proposed, the age of acquisition increases. In other words, there is parallelism between the four stages proposed for complex syntax emergence and how prosodic features are acquired by children.

This would be consistent with the bootstrapping hypothesis [25]. In order to correctly process this increasingly complex grammar, speakers would have need to use prosodic cues.

However, acquisition studies also give to us evidence in against this hypothesis. The results by [26] for Japanese prove the independence between prosody and complex syntax acquisition showing that prosody does not help to bootstrap Japanese children’s recursive phrases.

But even if prosodic bootstrapping is not a requirement for the acquisition of complex syntax, we can still argue for a parallelism between prosodic emergence and syntax complexity if we take into account the prosodic hierarchy model usually assumed in our field [27]. That is:

- Prosodic words contain syllables
- Phonological phrases contain one or more prosodic words.
- Intonational phrases contain one or more phonological phrases
- Utterances contain one or more intonational phrases

Following that, we will find that in each stage of the proposed model of syntax complexification under the HSD hypothesis, different prosodic strategies must have been emerged in order to mark this hierarchy.

5. Conclusions

Research in language acquisition and disorders has proved that prosody is essential to acquire highlight syntax dependencies, therefore prosody needs to be integrated in language development models that account for syntactic emergence.

This paper has applied an evolutionary model to the evolution of human prosody, arguing for a progressive complexification of prosody that parallels (and is triggered by) the complexification of grammar, also in response to a reduction in reactive aggression levels.

6. Acknowledgements

This work was supported by grant PID2020-114516GB-I00 funded by MCIN/AEI/ 10.13039/501100011033. The second author is a Serra Hunter Fellow.

7. References
