

IN DEFENCE OF THE MULTIMODAL HYPOTHESIS OF LANGUAGE ORIGINS AND EVOLUTION

* Svetlana T. Davidova

ABSTRACT

Attempts to understand the current dominance of speech as the universally preferred modality, have produced models which assume unimodal beginnings of language, either in non-human vocalizations or gestures. These stem from linguistic paradigms featuring language as a system designed and evolved for codifying the complexity of human thought in elaborate and intricate grammatical structures, demonstrated mainly in written texts. In this context spontaneous linguistic exchanges in dialogues, universally demonstrated by all humans at all times are excluded from consideration both by linguists and evolutionary linguists. As a result the road from non-human communication to the complexities and intricacies of modern languages used in service of advanced civilization has proven to be a major stumbling stone for evolutionary linguistics and often referred to in terms of miracles or just-so stories.

In breaking with the tradition, I argue that the language system is best represented by spontaneous dialogues of the average adult normal human. Language is a system for exchange of information by continuous communicative interactions by dialogues.

Spontaneous dialogues are subconsciously accompanied by non-linguistic communicative body movements, which are multimodal, mostly instinctive, reactions as active participants in spoken dialogues by adding and/or clarifying the linguistic message. Thus, linguistic communication, most often articulated by speech, is part, although a major part, of a human communicative complex, or ecosystem which includes multimodal non-linguistics signals e.g. facial expressions, body movements, manual gesticulations, e.g. pointing, etc. Human communication is a multimodal complex and language is explicable only as part of it.

Given the latest findings that multimodality in non-human communication is the norm in species and given the assumption of continuity, genetic, epigenetic, developmental, cognitive, communicative, of humans as a life form, despite the prevalent attitudes of human superiority based on unmatched linguistic abilities, I argue that the most realistic explanation of the evolution of language as behaviour and innate abilities, is in terms of restructuring of the non-human multimodal communicative complex by giving priority to vocalizations and the evolution of speech, as behavioural evolution, and subsequent adaptation of human cognition and physiology as evolutionary target aiming to facilitate this behaviour. Hypotheses of language evolution featuring multimodality are to preferred to unimodal explanations.

Keywords: language system, linguistic communication, Language Capacity, language origins, multimodality, primate communication, language evolution

1. INTRODUCTION

In the literature the human brain as a language processor is at the centre of interest. Competing theoretical alternatives understand the role of the human brain in different terms. The generative/mentalist perspective argues for Universal Grammar (Chomsky, 1972, 1986, 2000, among many other publications and lectures) as innate property of the human brain, which

shapes language, pinpointing the Broca's and Wernicke's regions as the most prominent locations of language processing activities. The usage-based perspective argues that various properties of the human brain used in wide range of cognitive functions, e.g. memory storage and cognitive processes e.g. categorization, chunking, analogy, reduction, and other mechanisms of general intelligence influence the shape of language. (Christiansen M., Chater, N., 2008; Kirby, S., 1998). In both perspectives inquiries into language origins and evolution are centred predominantly, although not exclusively, on the cognitive aspects of language.

Recent studies of sign languages (Poizner, H., Bellugi , U., Klima, E., 1990)have demonstrated that sign languages have properties highly similar to those of spoken languages which has been taken as a confirmation that the physiological properties of the channel by which the message is delivered are of no consequence in shaping the structure of language . That said, the human organism is an integrated complex, where the brain and the body function in tandem and influence one another. The human organism as a whole is a language user.

From a different angle, language as a system of symbols, has an abstract dimension and a material dimension. It is only through its material expression that we can know about its abstract architecture. One cannot exist without the other. Its communicative function is only possible as externalization of ideas by material venues. And these two aspects of language interact and influence one another.

Nevertheless, the perceptual-motor dimension of language use is underestimated and almost completely replaced by focus on the cognitive functions of language, as the role of modality is deemed irrelevant to the language system defined in purely cognitive terms. In addition, the conceptualization of the language user in terms of idealized model of a human makes the biological body irrelevant.

That said, language is a bio-cognitive behaviour made possible by the human organism as a whole, i.e. ideas are made explicit by physiology. Moreover, experience shows that real human speakers differ significantly from the idealization as they use the body as a whole, both the cognitive and physiological resources as they communicate. Moreover, speakers use differently human physiology in different spheres of language use. The average normal adult human individual as a typical representative of a language speaker organizes one's communicative interactions depending on the communicative circumstances and communicative goals.

1.1.On competence and performance

The tradition of the generative paradigm from its inception is based on Chomsky's famous rejection of Skinner's behaviourism leading to a complete rejection of the role of observable behaviour, stipulating the leading role of innate factors in cognition. In this context linguistic behaviour was deemed as unreliable indication of the properties of the Language Faculty, conceived as inward-looking systems designed to function in isolation from the rest of cognition and the human organism and its interaction with the external environment in terms

of perception and general intelligence.

Nevertheless, in life sciences it is a truism that that in all life forms biology and behaviour are closely interconnected and interdependent given that in all species the purpose of innate traits is to guide behaviour and in this way facilitate survival. Moreover, the only way to detect biological and cognitive capacities is by monitoring and/or provoking, their use in behaviour, usually by performing tests. From the muscles to the nervous system to the brain cells, one can detect their function, suggestive of their biological properties and, therefore, their evolutionary *raison d'être*, by triggering a behavioural response. That is, behaviour is the clearest indication of biological and cognitive resources in any biological form. And if one subscribes to the idea of innate Universal Grammar (UG) one would apply the same logic and assume that language use in communication is indicative of the Language Faculty.

That said, although language as a behaviour unique to humans, must rely on some aspects of the human organism, designed by evolution specifically to support this behaviour, scholars are increasingly convinced that these are not in the form of innate UG. Moreover, contrary to the assumption by biolinguists that the application of the Language Faculty in communication is an automatic reaction, i.e. an instinct, the use of language is an intentional behaviour and a choice, among non-linguistic alternative avenues, and involves planning with a purpose in mind. Some have argued that language must be defined in terms of communicative actions, referencing Austin's theory of speech acts (Kendon 2014). In this sense the prominence of linguistic behaviour, i.e. performance, as conscious application of innate potential in the formation of the abstract language system in the human mind, i.e. competence, advocated by the usage-based perspectives, is reasonable.

Thus, there are two alternative visions of language as abstractions vs. actions. Should we define language as abstractions or as actions? The most pertinent question is how are the language system and linguistic behaviour related? Does the language system have existence independent of linguistic behaviour?

And as language is only one of various uniquely human behaviours, e.g. music, dancing, building construction, etc., the same question can be extended to these with broad philosophical implications. Does a music score have independent existence from concert performances? Do plans of an architect have independent existence from building construction?

As in language, competence in any of these unique human activities, being uniquely human, must rely on some innate propensities, e.g. infants demonstrate innate propensities for rhythm, which are the innate seeds of music, toddlers build castles with lego, demonstrating budding abilities for creative thought and manual dexterity, etc. That said, although the seeds are innate, these activities are largely learned and intentionally put to use, improved by extensive education and practice with planning and a goal in mind.

Abstractions are ideas, i.e. they are by definition non-material. The act of attaching a material form to ideas by representing them in signs, be it as stream of speech sounds, alphabetic characters, pictures or other signs, is done with a purpose, e.g. to disseminate them by teaching or implementing them in practice. In addition, marking ideas with signs makes them explicit even to the author, e.g. verbalizing one's thoughts in speech and/or writing adds precision as it

helps clarify their shape and mark their boundaries. What is the purpose of abstractions unless they are used for something?

Returning to language and its use, lately labeled as languaging, there is ample evidence that the language system can only be demonstrated by its use through its material form. Given that programming languages, which have influenced linguistic theorizing and the very definition of natural language, are designed with a specific application in mind and, thus, have no independent existence beyond that, it is reasonable to extrapolate the same dependency to language and languaging.

Moreover, the language system is tailored to its utility and influenced by its material form for the practical purposes of languaging.

In written discourse the language system consists of linguistic primitives as lexicon members are one-to-one stable associations of a form and a meaning, i.e. literal meanings with clearly defined and fixed membership in discrete grammatical categories. These are organized into sentences with a multilevel internal architecture with multiple embedding of phrases, assembled according to fixed predetermined rules of grammaticality. The meaning of a sentence is the sum total of the meanings of the component units and their position in the architecture of the sentence, thus, interpretation is strictly structure-dependent and unambiguous. Here the language system is tailored to its functions in written texts of disseminating universal and timeless ideas by monologues addressed to audiences separated by space and time. Thus, written texts are produced and consumed by learned and linguistically trained individuals.

Spontaneous dialogues have different semantic and structural properties, e.g. they are information-based, not structure-based, organized around information structure based on the distinction of topic vs. focus and equally systematic. The building blocks of the system are flexible associations of form and meaning where codified standard meanings receive context-dependent interpretation. These are organized into utterances which are usually fragmented phrases and sentences. Small clauses, almost complete lack of embedding of phrases abound and sentence coordination is preferred while subordination is rare. Elliptical and abbreviated forms abound. Formulaic phrases are often used. Grammatically defective elements of unclear syntactic features, unclear morphological class and irregular phonology, or in Jackendoff's terms 'defective items' (Jackendoff 2002), 'mm', 'wow', 'sht' also abound. There is systematic omission of grammatical markers which do not contribute to the overall meaning and have only structural values, e.g. definite and indefinite articles in English. Semantically vague words and phrases, e.g. 'that fellow', 'that thing', 'people' are often used.

The vocal-auditory channel influences the language system both in form and in meaning as follows: the linguistic units are packaged in intonation contour. The sentence is organized to fit in a single prosodic contour. The boundary between a main clause and a complement clause is marked by a pause. As sounds are ephemeral, the rapid speed of processing limits the complexity of the utterance both in meaning and in structure. (Mithun, 2009, in Givón, Shibatani, 2009 p. 67).

Dialogues are conducted among people with close social ties in casual, relaxed interactions, with non-linguistic communication having major contribution to the understanding of the message. It is universally used by all speakers, regardless of age, gender, education, social

status or profession.

The language system of written texts and spontaneous dialogues is differently organized because it has different communicative purposes. The tool is tailored for the task.

2. LANGUAGE AS A UNIQUE COMMUNICATION SYSTEM: CODE AND INFERENCE

Language is usually defined in dichotomies either as a code or as an inferential communication.

Code-like communication is one where all meanings are explicitly represented in a codified form, thus, each meaning is represented by only one form, (there is no synonymy, homonymy, irony, etc.), the message is recoverable in its totality by decoding, the same associations of form-meaning are represented in the minds of all speakers, so there is no ambiguity. Code-like systems are very stable, resistant to change, so, the introduction of new form-meaning associations is very difficult. Code-like communication is 100% rule-based. Most animal communication and the artificial languages are code-like. In linguistics the generative theory's model of natural language is fashioned by artificial languages. The classical generative theory defines natural language as a code. Code-like communication is most clearly externalized in writing.

In inferential communication the meaning of an utterance is more than the meaning of a sentence: it includes the information supplied by linguistic forms plus additional information about the speaker's intentions, the listener's extralinguistic knowledge, contextual information as well as encyclopedic and cultural knowledge of each communicator. The interpretation of the message, in addition to the meaning of the signal, coded in stable form-meaning pairings, is sensitive to contextual information, which includes extralinguistic, cultural, pragmatic knowledge. Some part of the meaning of the message is not codable, thus, not recoverable from lexical or/and grammatical information alone, and not fully explicit.

The complete and correct recovery of the message depends on the combination of all these factors. The meaning of the message derived from the meanings of the individual words and their syntactic combinations is underdetermined and can create ambiguity in the interpretation. The same coded information in a sentence can be interpreted differently by different listeners in different contexts. So, normally the sent and the received messages are almost never identical, but similar.

Language, as both coded and inferential communication is unique to humans.

Moreover, the meaning of some linguistic forms is more code-like, that is, less dependent on interpretation, e.g. book, grass, sand, etc. In other cases the meaning of a word is open to interpretation, e.g. write/wrong, easy/difficult, justice, fairness, etc. vary with the interpreter and his/her individual experiences. Thus, language is a unique communication system which combines both a code and inferential components. The communicative success depends not only on decoding of the message but on its interpretation in a specific situation. (for more see D. Sperber, G. Origgi, 2010).

This hybrid nature of language is most clearly demonstrated in spontaneous dialogues as a universal of human communicative behaviour. The human mind is innately predisposed for participation in dialogues by a theory of mind and the uniquely human expectation for

relevance (Wilson, D. , Sperber,D. 2004 and elsewhere) which helps infer the specifics of the message from the vague guidelines supplied by the code.

In sum, natural language must be defined as a hybrid system with a code and inferential component.

3. HUMAN PHYSIOLOGY INFLUENCES THE LANGUAGE SYSTEM

Language is shaped by the human individual and its biological, cognitive and social dimensions. And given that the mentalistic perspective on language has been dominant in recent decades, the focus has been on the influence of the brain and mind in learning and processing language and its role in the formation of the language system . The cognitive aspects of the human organism are well-researched by all linguistic paradigms. That said, the human organism is an integrated complex, where the brain and the body influence one another as they interact. The human organism as a whole is a language user. In this sense the physiological aspects of language use and their influence on the language system are treated as an afterthought.

From a different perspective, language as a system of signs , has an abstract dimension and a material dimension. It can only function as communication system through its externalization by material signs. It is only through its material expression that we can access its abstract architecture . And these two aspects of language use interact and influence one another. Language use is a highly energy-consuming activity, both cognitive and physiological. This is reflected in its spheres of use as the human organism will naturally look for ways to minimize cost. Simple observation and experience shows a persistent pattern of correspondence between the organization of the language system and the modality which makes possible its externalization. Humans spend about 20% of all waking time engaging in casual face-to-face conversations which are predominantly spoken, although the entire body is a participant and a contributor. Spontaneous spoken dialogues among individuals with close social ties is the predominant form of language use both in space and in time. It is also universal despite language diversity. (J.L.Dessalles, 2007).

The processing of spoken language includes both production and comprehension and the biological organism imposes limitations on each as each participant is both a speaker and a listener and language processing places energy demands on both physiology and cognition. The spontaneous spoken dialogue imposes relatively low cognitive and articulatory demands as it tolerates grammatical inaccuracies, conceptual vagueness, incomplete phrases , incomplete and often inaccurate pronunciation. In addition, in dialogues the participants, often more than two, alternate in their turn taking which means that the cognitive and articulatory effort is shared and thus, minimized even further. The cognitive and articulatory demands in this case are minimal and inaccuracies and mistakes are of little consequence. Moreover, the demands for energy efficiency in articulation explains the fact that shorter forms are universally preferred (see Zipf's law of word frequency). Thus, the physiology of speech influences language use which, in turn, influences the language system.

Although spoken dialogues are considered of low energy cost, experience shows that spoken narratives have much higher energy demands, including kinetic energy. This makes long speeches, lectures, etc. rare events.

Casual conversations place limited informative demands on the language system as communicators are people with close social ties who share common beliefs, assumptions, world knowledge, i.e. common ground, which means that much information remains unspoken as it is implied. In addition, in face-to-face conversations information is supplied by non-linguistic means, i.e. non-linguistic body signals, e.g. manual gesticulations, facial expressions, head movements, etc. making language one among various contributing venues of communication.

On the other hand, communication among individuals outside the immediate circle in the absence of such common ground demands that most meanings must be made explicit by linguistic forms. This results in increase of vocabulary size, abstract linguistic forms and sophisticated grammatical organization. These trigger increased processing demands, mainly cognitive, for retention of larger vocabulary of lexical and grammatical forms. Extended cognitive efforts are demanded both by the production, as careful selection of the most appropriate forms among multiple options in the effort to make explicit complex ideas, as well as by the comprehension in effort to evaluate these. If externalized in spoken form, this elevated complexity of the language system requires, in addition, physiological demands for clear articulation on the speech organs. This restricts the usability of speech for mass dissemination of information in a large society marked by information inequality.

The increase of the number of potential communicators in large open “societies of strangers” and the demand for language complexity demonstrates, among other things, the limitations of the vocal channel and prompts the invention and proliferation of writing. As letters are stable signs, with their replacement of sound as ephemeral medium, the cognitive pressure of immediacy on the brain by allowing time for accurate encoding the ideas of the sender and the accurate decoding by the receiver are minimized. In addition, the physiological demands on the eyes by reading and writing are less than these for speech.

On the other hand, the physiological demands of speech production and written texts are minimal, the energy demands of the production of sign gestures on physiology are much higher. And although I am not familiar with kinesiological studies which calculate precisely the number of calories involved in gestural sign production, it is reasonable to speculate that physiological demands of sign language are comparable to those of everyday physical activities, e.g. gardening, running etc. As the biological body imposes limitations on all physiological activities, the same limitations apply to sign language. This explains why the use of sign language is limited to face-to-face conversations, prohibiting its use in long narration as the high energy demands on the hands, arms and the entire body would drive significant simplification of sign production which would create obstacles for comprehension. This also explains the fact that sign languages are not the default channel and are only used as a compensation when the vocal-auditory channel is damaged, not as a choice.

Thus, language structure is influenced by language use in communication. The human individual, as body and mind, is the language user and, although language use is defined

primarily as a cognitive behaviour, it is made possible by the organism as a whole, with the active participation of human physiological body which influences the language system as a tool for its materialization. In this sense the persistent co-occurrence of properties of the language system, spheres of language use and the material venues of externalization is clearly discernible. The visual modality and written texts fit well with the code system. Face-to-face dialogues are predominantly spoken, although the entire body is a participant. Sign languages are face-to-face dialogues where the role of speech is substituted by activating the kinetic energy of the whole body.

3. 1. Spoken and sign languages : meaning and structure

The generative tradition focusses on structure and understands both sign and spoken languages as representation of the same human Language Faculty.

That said, one can talk about similarities of meaning and structure between the two, explicable with the similarity of function. Notable but underestimated by theoreticians similarities, are in meaning representation, e.g. in the semiotic diversity of the linguistic signs as both types of language systems use iconic, indexical and symbolic elements at both levels of lexicon and structure. Although sign languages contain more iconic and indexical elements than spoken languages, (Kendon 2014) various scholars argue for a significant role of iconicity in spoken language, on par with arbitrariness. (Haiman, 1983; Langacker, 2008; Evans, Green 2006; Haspelmath 2005; Perniss, Thompson, Vigliocco, 2010; Kendon 2014).

Motivated signs, i.e. icons and indexes, persist in modern language in all its typological and modality diverse versions. Thus, in theory of language, irrespective of the modality one has to acknowledge the semiotic diversity of linguistic elements as a language universal. Moreover, one can look for similarities in structure, e.g. in both spoken and sign languages discrete and holistic elements are participant in building the architecture of the system.

Thus, there are similarities in the structure of spoken and signed dialogues, given the similarities in their functions, i.e. face-to-face communicative interactions, most often by dialogue among people with close social ties.

To sum up, language as a unique human behaviour is a multifaceted phenomenon. It is a system of signs, predominantly, symbols, although iconic and indexical signs are universally used. As such it is a communication system, a hybrid of code and inferential components. It is materialize predominantly by speech, with the contributive role of the human body as a whole, although in specific circumstances writing and manual signs are alternative avenues of material representation. Its universal purpose is exchange of information through continuous interaction of participants in spontaneous dialogues. Spontaneous spoken dialogues are the quintessential representation of natural language, accompanied and complemented by impulsive participation of the body as a whole. This is, in my mind, an accurate description of natural language the evolution of which must be explained.

4. MODALITY IN EVOLUTIONARY LINGUISTICS

So far I have argued that human physiology influences the shape of modern language. It is logical to extrapolate that this influence was equally significant at earlier stages of language evolution. This has been recognized by students of language evolution with expertise in biology which has led to debates as scholars diverge in their views on the original modality. Two major criteria are considered: a. compatibility with the design features of language as determined by the dominant theoretical perspectives, b. compatibility with the fundamental principles of biology focusing on evolutionary continuity. Naturally, writing as a modern invention is excluded from inquiries in language origins and the focus of debates is the choice between speech and manual gestures.

4.1. Debates on the original modality : sound vs. gesture

From linguistics standpoint the role of human physiology can be discerned from its role in language use in communication. Students of biology and anthropology have a different perspective: they study the evolution of the human organism as a life form in accordance with the Darwinian principles of evolution. And despite this underlying common ground, scholars diverge in their estimation of the role of human physiology and the preferred modality at the initial stages of language evolution, although most envision unimodal origins of language. For some, the vocal-auditory capacities were the evolutionary target resulting in the modern human capacity for speech from non-human vocalizations inferred from the fact of the dominance of speech in modern languages. Others offer arguments for the primacy of manual gestures influenced by studies of sign languages. Recently an argument for multimodal beginnings is made. In the following segment a short summary with focus on the strengths and misgivings of each is pertinent.

4.1.1. The argument for evolution of speech from non-human vocalizations

Speech is the universal channel for linguistic communication in the overwhelming majority of language varieties. It is a very specific type of vocal signal. It is unusual in various ways: acoustically, physiologically, cognitively, developmentally. Acoustically, it is different from non-speech sounds as detected by acoustic equipment (Pinker, Jackendoff, 2009). Physiologically, it is produced by the rapid and coordinated movements of the organs in the vocal tract. Its most defining and universal feature is the syllable, a repeated alternation of a vowel, or other sonorant sound and a consonant. It is represented in the human organism in a unique way. Anatomically, the ability to produce speech signals requires not only a specific type of anatomy, but also a specific type of neural connectivity. Lenneberg (1967) shows that various traits in the human anatomy and physiology are perfectly fit for use in speech. These include the anatomy and configuration of the vocal tract, e.g. the shape and position of the tongue, the morphology of the human larynx and glottis, conscious control over respiration. In addition, there is coordination among multiple traits in the human organism for the sole purpose of speech production. In sum, the capacity for speech is meticulously coordinated and exquisitely orchestrated interaction of anatomy, physiology and neuronal connectivity. Liebermann (2006, 2007, 2008, 2016)

In addition, there is ample evidence for the deep evolutionary roots of speech as continuity both in sound production, learning and processing in various species, some quite distantly

related to sapience. The songs of some birds are combinatorial and hierarchically organized prompting scholars to talk about phonology of bird songs and phonological syntax. (Hilliard, White, 2009). Vocal learning is demonstrated by change in the structure of the sound signal of a youngster in attempt to imitate that of the parents' is found in birds, some marine animals and even bats and elephants. (Hilliard, White, 2009 ; Schraff, Petri, 2011, p. 2125). Vocal communication and learning imply adequate hearing capacities, socialization and critical period for learning, found in song birds (Hilliard, White, 2009). Speech-capable anatomical and physiological machinery and cognitive resources are found in various species. Macaques (Fitch et al. 2016) and even bats have demonstrated a complex system of vocalizations , echo-detection. Bats learn new vocalization throughout their lives, their vocalizations have dialects . Interestingly, their communicative interactions resemble human dialogues in turn-taking (Vernes, 2017). Moreover, combinatoriality in vocalizations of monkeys and apes is demonstrated in addition to diversity of dialects in populations. Their vocalizations are also found to be referential and situation-specific, reflecting the audience (Arbib, Liebal, Pika, 2008). Hurford (2012) quoting studies by Gemba et al. shows that the voluntary control of vocalizations also has a long evolutionary history and is not a uniquely human trait stating that “ ...facts suggest that the neocortical area homologues to the human speech area takes part in the generation and control of monkey vocalizations ...” (Hurford, ibid. p. 106). What is unique in humans is “ fine control of articulation” (Hurford, ibid. p. 107). In addition, Sherwood et al. , referenced by Hurford (2012) show that great apes have voluntary control also over their facial movements. Liebermann has consistently argued that Neanderthals had speech capacities similar, although not identical to humans' as they weren't able to pronounce the so called quantal vowels (Liebermann (2006, 2007, 2008). Others argue that Neanderthals were “... fully articulate beings..” and had all the necessary attributes of human language speakers. (Dediu, Levinson, 2013, Abstract).

The argument for evolutionary continuity of speech capacities is enriched by evoking exaptation. MacNeilage (1998) and Davis, MacNeilage (2004) authored the frame-content theory of speech evolution where they argued that lip smacking gestures in primates were co-opted for the formation of the syllable and its most basic structure as CV. There is also a claim by Ralph-Axel Muller (1996) that the syllable is an instance of a “ general motor alphabet”, a unit of body motion , exapted for a new role in speech production. He has also suggested that the emergence of the syllable as a linguistic unit can be traced to the rhythmic jaw oscillation and the rhythmic nature of the movements of the body-extremities, termed as “general motor alphabet “ (not specific for any group of muscles and comparable across species) which, when applied to the human vocal tract, yields the set of phonemes found in human languages.” (Muller, ibid.). In a broader sense this also suggests that the syllable, one of the hallmarks of language, is an example of the general cyclicity observed everywhere in nature, e.g. alternation of day and night, the alternation of seasons, or the movements of the planets. This places us not only as a part of living creatures, but also as a part of the broader nature, not a departure from it or an exception. The inventory of speech sounds depends on the ability of the articulatory apparatus to pronounce them with sufficient speed and clarity. The formation of speech organs is interpreted as the classical example of exaptation: the original

function if the vocal tract is breathing and mastication, was reconfigured for speech which makes the rapid articulation of speech sounds and especially the quantal vowels possible. In modern human's anatomy and physiology the biological function is shared with the communicative one, demonstrating congruence with the evolutionary principles which prefer slight gradual changes in old structures to evolutionary novelties.

That said, the danger of choking when eating and speaking simultaneously is a valid counterpoint against the assumed survival advantages of spoken language from Darwinian perspective as the both activities are performed by the same biological organs.

Although this line of inquiry focuses on the aspects of externalization of the language system, Liebermann and other experts on human physiology and cognition have argued extensively for the connection between the brain regions responsible for motor control of the vocal tract and those responsible for linguistic structure. He states that "...The FOXP2 gene is clearly implicated in the formation of neural circuits that regulate the human cognitive and motor capacities (in Lieberman 2007, p.52). Liebermann (2016) argues that the FOXP2 human version of the gene is involved in the phenotypic formation of the human basal ganglia, responsible for rule-governed complex structures both in cognition and praxis, e.g. the ability to dance, produce tools and use language. This reminds that language use is activity integrated with non-linguistic activities by the organism as a whole.

4.1.2. The argument for gesture first

Various scholars have advanced the argument that linguistic communication originated in the gestural modality and only later became spoken. The recent invention of sign languages, comparable with spoken languages in form and function, meaning and structure, lead to the conclusion of the irrelevance of modality and the dominance of cognitive aspects of language. In addition, the linguistic achievements of modern apes demonstrate relative ease in learning of symbolic gestures and their use in structured combinations in situation-independent way and even occasional invention of new signs (Arbib, Liebal, Pika 2008 and elsewhere). And although apes' linguistic abilities are comparable to those of young humans, suggesting much lesser cognitive and communicative sophistication, this also suggests deep evolutionary roots, cognitive and communicative, of the essentials of human language in gestural communication, furnishing the argument for gestural origins of language. Studies in human brain architecture and functions are interpreted as support for this hypothesis, e.g. there is spacial proximity between Broca's area, assumed by many to be the language area of the brain responsible for grammar, and the area controlling the movements of the right hand, suggesting a natural connection between linguistic structure and manual gesture. Moreover, in modern humans speech is normally accompanied by spontaneous manual gesticulations suggesting neurological link between control of hand and mouth (Corballis, 2003). Comparative studies which show that area F5 in the monkey's brain, an evolutionary homologue of Broca's region, which controls manual as well as orofacial movements, are interpreted to support continuity in human brain architecture and functions of controlling the coordination of manual and facial movements.

On the other hand, primate vocalizations are viewed as much less likely precursor to modern spoken language as primate vocal signals are holistic, fixed in number innate emotional responses to specific situations, new vocalizations are never learned, even after intense training, suggesting lack of intentionality and conscious control over vocalizations (Fitch, 2009, Tomasello 2008).

Criticism of the argument focusses on the various disadvantages in the communicative utility of gestural communication in comparison with spoken language:

- a. It is useless in the dark and over long distances.
- b. Gestures are not practical as they occupy the hands and interfere with effective tool use or other activities, while vocal communication frees the hands, people can speak and work with their hands simultaneously. On the other hand, the same argument can be used against the advantages of speech pointing at the fact that speech interferes with chewing food and drinking water and the danger of choking while speaking.
- c. Gestures are much less energy-efficient than speech as the kinetic demands of speech are minimal.

Arguments against the hypothesis of gesture first (Dessaless 2006, p. 143 , P. MacNeilage, (1998) point at the fact that people, deaf or not, never spontaneously begin using sign language before learning spoken language. Dessalles (in 2006, p. 143) also points at the difference between gesticulations accompanying speech and sign language, the first bearing resemblance to an instinct, the second having all the characteristics of language as compositional, intentional , conscious communication. Moreover, the overwhelming majority of natural languages are spoken while sign languages are an exception in that , although they are natural languages, equivalent in expressive power and structural sophistication to spoken languages, they emerge in very specific circumstances when the speech capacities are damaged as a solution of last resort. There is no case when a human community physiologically capable of speech has chosen to invent and use sign language. Sign language is learned only as a substitute for speech, and not by choice.

In addition, recent findings by Perlman challenge the standard assertion of lack of voluntary control over vocalizations and vocal learning in apes and derived from that suggestions of inherent limitations in vocalizations to instinctive responses to limited environmental triggers and , thus, lack of potential for usability in language (Perlman 2017). Ample evidence is offered for vocal control as well as ability for vocal learning by imitation in various species. Tomasello (2008) Corballis (2009 , 2009a etc.) attribute the current dominance of spoken language to a genetic mutation in the FOXP2 gene at about 200,000years ago (ya) which produced its human version of the gene and resulting in transfer of control over the vocal tract to the Broca's area and opening the way for evolution of modern speech.

Rizzolatti , Arbib, 1998 , Arbib 2004, 2007, 2017 and elsewhere) argue for origin of language from manual praxis and attribute a prominent role to mirror neurons, credited to have a role in the capacity to interpret gestures as signs, i.e. intentionally emitted signals directed to a receiver. In this context the origin of language is envisioned as transformation from involuntary imitation of manual praxis routines , to voluntary imitation of much more complex events and performance of pantomime to gestural protolanguage. The argument builds upon

M.Donald's hypothesis featuring emergence of compositionality and generativity of gestures in pantomime as a precursor to language as transformation of praxis into communication. (M. Donald, 1991).

The most language-relevant aspect of this evolutionary process is the evolution of compositionality and generativity in manual gestures followed by further evolutionary developments involving Broca's area and FOXP2 gene resulted in incorporating vocalizations and the formation of a mixed manual-vocal imitation and the “ language-ready brain” i.e. capable of compositionality and generativity of signs by a mixture of gestures and vocalizations (Arbib 2002). This, the argument goes, has facilitated the formation of vocabulary and propelled the development of spoken lexical protolanguage, which as per Bickerton (1984 and elsewhere) have preceded the evolution of grammar (Arbib. 2007, 2017). Further the direct representation of physical reality in communication by action-object frame was transformed with the evolution of symbols into indirect representation by noun-verb frame and grammar, the crucial step to modern language. In sum, the mirror neurons are credited with propelling the evolution of modern grammar being the foundations of syntactic structure in motor routines or “ action grammar”(Arbib, 2004).

In this context the evolutionary transformation of the mirror neurons in monkeys area F5 into Broca's area is argued to be the possible explanation for the assumed transition from gestural protolanguage into speech, in essence solving the problem of the “ missing link” between languageless primates and language-capable humans. The capacity for copying observed simple gestural routines is, thus, projected to have been present in the common ancestor of monkeys and humans. Thus, the proponents of the argument see the roots of language in gestures. (Arbib, 2004, Arbib, Liebal, Pika 2008, to name a few).

Consistent with this are earlier findings by P. Greenfield (1991) of structural commonalities between manual object manipulation and linguistic abilities of modern apes and human toddlers, thereby demonstrating common neurological foundation of both behaviours in Broca's region, suggesting homology as evolutionary explanation.

In a similar vain scholars evoke archeological findings of Oldowan and Achewlian technologies suggesting impressive manual dexterity of pre-human ancestral species revealing capacities for hierarchical organization demonstrated in praxis, hypothesized to have been co-opted for communicative use in mimetic communication in Homo Habilis and Erectus as a precursor to modern hierarchically organized language. The transition to speech from a holistic expression of emotions , participant in the pantomime, is explained with the evolutionary principle of energy efficiency. (Sterelny, K.2012 ; Arbib, M. 2011).

The suggested role of mirror neurons in the evolution of language is not without critics. Chaney and Seyfarth (referenced by Fitch, 2009, p. 457 -) object as they point at the fact that chimps have a very limited capacity for mind reading which is used for selfish purposes in competition and not for social cooperation which is expected to coexist with the capacity for imitation and mirror neurons if they are to be considered as a precursor to language. Further objections argue that the parity connection of seeing and performing an action by a primate does not imply any connection with meaning or intention which makes it difficult to infer that

manual gestures are treated as signs and view mirror neurons as a precursor to language. Further, Fitch (2009, p. 456 -) referencing Hurford (2002) points at the fact that the mirror system can represent only a limited number of concepts of body movements, in stark contrast with the open-ended vocabulary of human language. In addition he argues that imitation produces iconic signs and cannot lead to arbitrary connection of signal and meaning found in symbols. Thus, the iconic imitation of manual movements is “ ... a poor precursor to Saussurean signs whose key properties are its arbitrariness and openness to any and all concepts...” (Fitch, 2009, p. 456). Similarly, Hurford (ibid.) argues in continuation that the structure of motor gesture combinations does not lead to abstract structures found in modern syntax.

That said, newly emerged sign languages, in their short histories initially dominated by iconic gestures, demonstrate fairly rapid change from iconicity to symbolism (Sandler, et all, 2005). One can anticipate a similar development in hypothesized protosign systems, although the high energy requirements of manual gestures would have restricted the expansion of a gestural lexicon . In sum, the role of mirror systems in phylogenesis of language is unclear.

The hypothesis for gesture first has difficulties explaining the sudden how the switch in modalities by transfer of motor control from hands to vocal tract as such biological development is not reflected in the anthropological record as a sudden switch in behaviour.

5. THE LANGUAGE SYSTEM AND MULTIMODALITY

Multimodal communication is defined as a composite integrated signal containing simultaneous emission of signals from multiple modalities, e.g. vocalizations, gesturing, body posture, facial expression. Linguistic communication is universally accompanied by simultaneous non-linguistic body signals which have a significant contributing role in the utterance meaning. This begs the question should these co-occurring body signals be identified as part of language or outside of it. In other words, is language multimodal?

5.1. Multimodality in face-to-face dialogues

To begin with, given that the brain is a multipurpose and multifunctional processor, it is to be expected that the processing of language would involve multiple functions.

Most scholars agree that language is a system of and for communication. It is exemplified most clearly by the linguistic behaviour of the average normal adult human, who spends most of his/her waking time engaging in of face-to-face dialogues. Participation in dialogues is the universal of linguistic behaviour regardless of modality, i.e. dialoguing is the common denominator of language, spoken and signed.

Sign languages are by definition multimodal and the linguistic signal in spoken and signed dialogues is constantly accompanied by multimodal co-speech body signals. In sum, dialogues, spoken and signed, use multimodality.

5.2. Non -linguistic body signals, sign and the language system

Both spoken and signed dialogues use the human body for communication, although differently. Importantly, both spoken and sign languages are used in coordination with co-linguistic body signals. Thus, co-linguistic body signals are universal aspect of human communication.

Sign languages are multimodal by nature and spoken dialogues include elements of multimodality as the linguistic message is coordinated and complemented with non-linguistic signals, e.g. non-verbal vocalizations, face expressions, body posture, eye movements etc.

In this context it is pertinent to ask are non-linguistic body signals accompanying speech part of the language system or contributing extralinguistic layers of communication. Co-speech body movements form integral part of conversations by aiding the processing of the linguistic message, both in production and comprehension, a universal phenomenon formalized in the Information Packaging Hypothesis by Goldin-Meadow and Alibali (2013).

More complex is the relation of co-linguistic body signals and linguistic signs in sign languages, currently a matter of debate. A. Kendon (2000 and elsewhere) argues for a continuum of linguistic sign and co-linguistic body signals accompanying signed utterances, known as “Kendon continuum”. Alternatively, D. McNeil finds clear discontinuity and focusses on the non-linguistic properties of body signals. For comparison see C. Muller (2018).

To note, co-linguistic body signals are diverse and authors differ in their terminological identification. Even the definition of “gesture” differs for different authors. The most clear distinction is between spontaneous singular gesticulations and conventionalized gesticulations, labeled as “gesticulations” and “emblems”(C. Muller, 2018).

Although both sign languages and co-linguistic body signals participate in human communication, I find major differences between the two. On the other hand, although non-linguistic signals differ somewhat in their communicative functions and their relation to the human body, these are most accurately defined along a continuum.

a. Most co-linguistic body signals, or gesticulations, are spontaneous, involuntary body reactions. Some are universal, e.g. a lie is accompanied by subconscious alterations in facial expression and voice modulation. Others are unique to each individual. This is why actors must learn the body reactions of the characters they portray as a signature of their individualities. It is reasonable to attribute a prominent role of nature, not nurture, as speakers perform gesticulations while speaking on the phone and blind speakers perform co-speech gesticulations while talking to blind speakers. (E. Cartmill et al. 2012). One can suspect some kind of instinct.

Emblems are culturally motivated, and thus, learned, but equally automatic body signals e.g. hand shake as greeting, head shake in some cultures is a sign of disagreement and head nod a sign of agreement. Emblems are practiced uniformly by all community members and are markers of cultural identity.

In contrast, gestures in sign language are intentional, learned and conventionalized and

demonstrate social and cultural diversity, suggesting a prominent role of nurture.

b. Body signals like non-linguistic vocalizations are mostly expressions of emotions, e.g. one subconsciously smiles when talking about pleasant things, an expression of disgust accompanies a conversation about unpleasant things, although emblems may convey propositional meaning in a holistic signal. In contrast sign languages express mostly propositional thought.

c. Given their instinctive nature, gesticulations are unchangeable vs. sign languages are susceptible to historical change.

d. Gesticulations are iconic and indexical signs, although as per Levinson, Holler, (2014) finger pointing has a symbolic component, while sign languages are predominantly symbolic systems.

e. Co-linguistic body signals are holistic signs, i.e. pairs of holistic forms and holistic meanings while sign language is combinatorial, hierarchically organized system. Consequently the meaning of gesticulations is vague and ambiguous, which is why they are produced in coordination with linguistic signals, while the message in sign language is precise.

f. Co-linguistic body signals have a complementary role in communication although emblems can be used on their own e.g. when speech is either inconvenient, as when interrupted by loud noise, or in silent films. Signs in sign languages are always components of a system.

Thus, co-linguistic body signals and sign language differ in : a. bio-cognitive aspects , b. in semiotic aspect, c. structurally, d. functionally.

In sum, the language system externalized in speech and sign has demonstrated characteristics different from non-linguistic body signals accompanying linguistic signs, suggesting that they do not belong to language. That said, their contribution to linguistic communication is significant. Thus, understanding what language is, is not possible without understanding the role of non-linguistic communicative behaviour complementing language use. Language use in spontaneous dialogues cannot be understood without non-linguistic body signals.

6. MULTIMODAL APPROACH IN EVOLUTIONARY LINGUISTICS

Inquiries in evolutionary linguistics rely on theories of language as premises in formulating their arguments. Modern linguistic theorizing dominated by the generative vision most clearly outlined by Chomsky's writings, too numerous to list here, has as its object of study an idealized version of the human organism and of human communities, resulting in idealization of natural language defined as a code. In this context the linguistic output of a small number of language experts and their highly restricted use of language as a professional tool mainly in writing is taken to represent this ideal. Thus, linguistic theorizing is based on highly atypical examples of linguistic communication in isolated circumstances, in no way reflecting the universal nature of linguistic communication and, thus, presenting a distorted picture of language and the language user. The conceptualization of language as an algorithm for processing a hierarchy of symbols espoused by the biolinguistic approach assumes innate

Universal Grammar and aim to explain its supposed evolutionary history. And although in biolinguistic context the cognitive function is viewed as primary in processing, acquisition and evolution, the presumption of language as symbolic system implicitly suggests communicative function. In this context the role of modality is deemed irrelevant to the language system, although implicitly assuming the vocal modality by attaching a phonological component as part of the multileveled structure of the formalism and of the language processing abilities of the brain, thus, assuming uni-modality.

Alternatively, communication and meaning features prominently as fundamental for the formation of the language system in usage-based/functionalist context, although there the focus is on the language system as a technology and the role of communities and cultures in its formation and change, while the contribution of the individual organism and, therefore modality, and its evolutionary aspects are ignored. The argument for gesture first relies on the conception of language as structure, while the argument for speech first is informed by the Saussurean view of language as a symbolic system.

That said, a compelling argument can be made that limitations in linguistic theorizing translate into challenges in evolutionary linguistics. (Jackendoff, R. 2010).

Given the active participation of multimodality in direct face-to-face spontaneous dialogues, a unimodal approach to language evolution is at best incomplete.

In sum, the real experience with language in spoken dialogues and the supporting role of human physiology in communicative interactions is generally ignored by theorists or, alternatively, limited to unimodal understanding of language.

Lately there are proposals of redefining language as a multimodal communication in recognition that face-to-face dialogues, the universal demonstration of multimodality in human communication and its role in linguistic communication must be acknowledged in designing theories of the language system. (Perniss, 2018; Perlman, 2017; Kendon, 2017 and elsewhere). Crucially, multimodality of modern human communication offers a plausible explanation for the evolution of language.

It is argued that the understanding of non-linguistic body signals and sign languages in terms of a continuum (Kendon, 2000 and elsewhere) could plausibly explain the formation of sign languages with historical changes in pre-linguistic communicative signals by their transformation into language systems via emergence of discreteness and compositionality, lexicalization and grammaticalization, processes mirroring the formation of spoken languages. See Muller 2018 for discussion. Thus, human communication as a multimodal complex makes use of linguistic signs and the spoken and sign modes are taken as interchangeable alternatives. That said, people, deaf or not, never spontaneously begin using sign language before learning spoken language. (Dessaless 2006, p. 143, P. MacNeilage, (1998). Sign languages emerge in very specific circumstances as a solution of last resort only when the biological equipment enabling speech production and perception is damaged, as a substitute for speech, and not by choice. There is no case when a human community physiologically capable of speech, has chosen to invent and use sign language. In this sense the evolution of spoken and sign languages have different evolutionary histories.

Moreover, given that multimodality by definition implies the participation of both body movements and vocalizations, the evolution of language is plausibly explicable as transformation in their respective roles in communication. As per Levinson, Holler “...together speech and gesture have mutually and maximally adapted to one another, resulting in a default mode of modern human communication characterized by highly efficient and informationally rich visual-verbal utterances” (Levinson, Holler 2014, p. 6). In this adaptive process the leading role of speech as vocal expressions of linguistic signals has emerged as a final stage.

6.1. Multimodal roots of language in primate multimodal communication

Multimodal signals are demonstrated by many species as distant from us as insects, frogs, birds, squirrels, etc. as they combine and coordinate signals from multiple modalities which complement one another to compose a convincing message (Slocombe et al. 2011). Thus, one could extrapolate that across the living world multimodality in communication is the rule rather than the exception.

Not unexpectedly, multimodality is at the heart of primate signalling. The communication of primates, both in natural habitats and in captivity, is a composite signal, i.e., a combination of various components which act in concert to deliver the complete message (Slocombe, Waller and Liebal 2011). Chimps in captivity communicate with signals emitted simultaneously from a variety of modalities. They recognize and produce visual, tactile, vocal signs in their communication with caregivers. Chimps are known to combine alarm calls with gaze alternations and are shown to have voluntary control over manual gesticulations as well as control over individual vocal organs, e.g. control of breathing, larynx, supra-laryngeal articulators and capacities for vocal learning (Perlman 2017). Macaques's communication is cross-modal as well suggesting that “communicative complexity may be less about how each single modality is used, and more about signal integration.” (Slocombe Waller, Liebal, 2011, p. 540). Given these findings the unimodal approach to primate communication is inadequate for understanding a complex and integrated system as studying each component in isolation ignores the influence of the rest. For example, “When a slap is paired with a playface it leads to play. Isolating the slap from the playface will not, therefore, help us understand the signal better.” (Slocombe, Waller, Liebal 2011, p. 540). See also Boe, L.J. et al 2017)

Communication in general is multimodal and a multimodal approach is to be preferred to isolationist one as the function of a single modality may not be clear without considering its use as a component of a multimodal communicative ecology. And given that language as part of the human multimodal communicative complex, human linguistic communication can be better understood as emergent in the context of a multimodal primate communication. On the other hand, inadequate understanding of primate communication is clearly an obstacle for evolutionary linguistics. Moreover, by understanding the communicative functions of the multimodal complex as a whole, we can understand the phylogenesis of the mechanisms which make this multifaceted behaviour possible.

6.2 . Language evolution as restructuring of primate multimodal communication

Under the definition of language by the code theory, evolutionary linguistics defines language evolution language as a code as the final point of language evolution. In this context evolutionary linguistics aims to explain the road from non-human multimodal communication to language as symbols and multileveled hierarchical structures, suited for representation in writing and tailored for its use in modern civilization.

That said, language is a communication system universally made explicit in spontaneous spoken dialogues, which is organized very differently from the linguistic code and writing as it has different functions. Spontaneous dialogues exist as a component of the multimodal communicative complex humans use to communicate. Thus, spoken language can only be understood as part, although very significant part, of multimodal human communication. In this context the evolution of spoken language is better understood in the broader context of human communication.

And under the assumption of continuity of humans with pre-human ancestral species not only genetically, but also developmentally, physiologically, cognitively, although the predominant view attributes to us superb linguistic abilities, justifying the perception of human superiority and a break-away from the rest of life forms on Darwin's tree of life, I find it reasonable to suspect continuity in communication as the most reasonable explanation for the rise of language as spontaneous spoken dialogues.

In this context one could anticipate pre-human multimodal communication systems as a mixture of holistic and combinatorial signs. The fact that modern languages preserve some holistic structural elements suggests that there never was a complete displacement of one system by another, but a gradual transformation where some elements assumed a leading role while others remained as a contributor. The fact that modern languages preserve some indexical, iconic and symbolic signs, inherited from an older and more primitive and less effective multimodal stage, suggests a gradual shift giving priority to symbols.

From the early starting point of multimodality speech is likely to have gradually evolved as the major component in this multimodal complex, while relegating to body posture, non-linguistic vocalizations, a contributing role. The leading role of speech is explicable from both physiological and cognitive perspective, as the most energy efficient, as well as from communicative perspective as allowing for human thoughts, encased in a symbolic and combinatorial system, to be materialized effectively with speed and precision.

And given the adaptive advantages of effective communication, it is reasonable to suspect that these structural alterations in the language system have triggered internal restructuring in the multimodal participation of the human body by coordinated adjustments among the relevant aspects of physiology ending up in the dominance of speech. Most importantly, restructuring in human communicative behaviour would have triggered adaptations in the human organism to the new kind of cognitive and physiological demands of language processing and learning, culminating in the evolution of a Language Faculty. This perspective suggest a new angle to the continuity debate in language evolution.

6.3. Multimodality and the evolution of the Language Faculty

In humans spontaneous spoken dialogues linguistic signals are co-articulated with multimodal body signals.

The processing of linguistic signs is standardly attributed to Broca's area, where language is processed by integrating phonological semantic, grammatical, extralinguistic information in the formation of individual words and their further integration into larger structures, phrases and sentences.

That said, Broca's area as part of the Inferior Prefrontal Gyrus (IPG) in humans is also known to integrate processing of vocal and gestural signs. Moreover, speech, sign language and co-speech body signals are processed by the human brain in a coordinated way (Frohlich, M. et al. 2019). Thus, the brain acknowledges the multimodal human communication, linguistic and not, as a unified complex.

The evolutionary roots of human multimodal communicative behaviour are found in primate communication. Multimodality in non-human communication is made possible by brain functions at neuronal level as various neuronal populations at various brain regions are capable of processing signals from multiple modalities simultaneously (Slokombe et al. 2011). Thus, non-human communication is multimodal both in terms of behaviour and brain functions. Significantly, Broca's homologue in the primate brain has very similar functions as integrating centre for processing of various communicative signals. A spacial and functional overlap is found between neurons processing vocal, visual and gestural signals. (Frohlich, M. Et al. 2019). It also has similar functions of integration of perception and motor functions essential in observation, imitation, planning, in macaques and humans (C. Sherwood et al. 2008) Consequently, deficits and/or damages affecting this part of the brain would impair a number of functions.

6.3.1. The mirror neurons and multimodality of language origins

To remind, the role of mirror neurons in the phylogenesis of the Language faculty is viewed in terms of transformation from automatic and subconscious copying of observed praxis routines into compositional expression of meaning with the dominant participation of speech. The discovery of mirror neurons in monkeys has been interpreted as suggesting that mirror neurons as an evolutionary innovation first evolved in monkey species. The behavioural representation of mirror neurons is argued to be demonstrated by copying of manual gesticulations.

That said, a fairly recent and less popular study finds behaviour suggestive of mirror neurons in avian species, e.g. neurons in the forebrain of sparrows become activated both by performance and perception of songs, involved in vocal learning. (R.Garcia et al. 2014)

Similarly, copying of lip smacking and feeding activities are detected in area F5 in monkeys, suggesting the presence of oral mirror neurons. Moreover, rhesus macaques are able to learn

hierarchically organized auditory sequences similar to bird songs. And although more studies are required for the definitive confirmation, this suggests that mirror neurons and their involvement in learning mechanisms have a much older evolutionary history and have functions not restricted for manual gestures but including vocal imitation.

These similarities in both behaviour and brain functions suggest that the evolution of the Language Faculty should be defined in terms of multimodal communicative interactions, suggesting that evolutionary models in terms of multimodality should be preferred to unimodal models currently dominating evolutionary linguistics.

The fact that language is embedded in multimodal communicative behaviour which shares similarities of function with pre-human communicative behaviours suggests that language as a unique human behaviour and the language capacity, which makes it possible, is explicable with continuity in multimodal communication rather than an event of switch from gesture to sound.

From a different but related perspective, it is commonly stated (mistakenly) that all and any thought can be communicated through language. On the other hand, Dor and Jablonka, (2000) point out that only selected fractions of the human semiosis is communicable through language, i.e., language is a system designed to communicate a broad although limited range of concepts. Certain scenes, events and their participants are universally easier to express in linguistic forms than others. As per Dor and Jablonka “Natural language is a communication tool, which is structurally designed for the communication of a constrained set of meanings.” (ibid. p.41). The universal patterns of categorization are universally reflected in the grammatical systems of all languages

“...semantic categories which are reflected by grammatical complexities in natural languages belong to a very constrained subset of all the categories we use to think, feel and conceptualize the world” and “...some semantic categories turn out to be grammatically marked in language after language, whereas some others consistently do not participate in the grammatical game” (Dor, Jablonka, ibid. p.39).

An example of the former is the categorial distinction of animate/inanimate objects, and of the later the lack of categorial differentiation between interesting/boring experiences. The universality of such linguistic patterns is explained with innate propensities for effortless learning of these as part of the Language Capacity.

In this context the participation of non-linguistic signs in face-to-face conversations is explicable as a compensation in the externalization of meanings for which language is not designed. Moreover, in modern societies where long distance communication among individuals separated by space, time and communicative context is prevalent, pockets of the human conceptual universe are found to be difficult or impossible to express in linguistic forms, e.g. playing basketball, spatial orientation, some scientific concepts, etc. As compensation for these deficiencies civilization has invented additional, non-linguistic technologies e.g. maps, graphs, formulas, drawings, etc..

In sum, the evolution of the Language Faculty is better explained in the context of multimodal

communication in terms of continuity with pre-human multimodal communicative abilities given the benefits of language use as fast and effective tool for encapsulating the universals of human semiosis for the purposes of communication.

6.4 . Multimodality in child language development

The argument that ontogeny represents a condensed version of phylogeny is well known in the literature and ,although it has its supporters and critics, it provides a possibility to infer information on evolution permanently lost and non recoverable from observation of development . In this sense the multimodal nature of the first communicative engagements at early age could shed light at initial stages in language evolution.

In the first weeks and months after birth the infant displays affinity for socializing and participating in communicative behaviour by visible body actions, e.g. smiling in response to a smile, imitate facial expressions, turn-taking in non-verbal sound making and playful interactions ,facial expression, gesticulations, body posture, directing attention by sound making, looking, pointing. The youngster's first experimentation with language begins with babbling , followed by one-word utterances where children 's communicative interactions are predominantly by body movements with language having a limited and supporting role which broadens with the next stages of language development with two-word utterances and beyond. As development progresses, although the child's linguistic abilities become more sophisticated, his/her linguistic output continues to be complemented by non-linguistic body signals. During the development of syntactic competence, as children often omit arguments or elements of constructions which compromises the overall understanding of the message, the lack of full linguistic competence is mediated by body actions, e.g. pointing, etc. Linguistic behaviour is constantly synchronized with communicative use of other modalities as a component of a multimodal unified system (Morgenstern A.,2014). And although the use of body gesticulations as compensation for linguistic immaturity gradually decreases as the child's linguistic skills develop in full , to paraphrase Morgenstern A.(2014), body signals continue to illustrate, specify, reinforce, modify the meanings of linguistic production (ibid. p. 18)

In sum, as early language development is viewed as providing insights into the evolution of the language, the multimodal participation in early linguistic behaviour in young children suggests that early language also emerged as a component of a multimodal communicative ecology.

SUMMARY AND CONCLUSIONS

Human communication is a multimodal complex of various components at the centre of which is spoken language as a representation of underlying language system. Its communicative use is universally complemented by various non-linguistic layers of communication as demonstrated by face-to-face dialogues in spoken and sign languages.

On the other hand, unimodal externalization is always partial and incomplete, e.g. writing is unable to convey some components of the intended message, e.g. tone of voice as expression of

irony, negative attitude veiled by affirmative linguistic forms, in spatial orientation linguistic explanations alone are often insufficient and are complemented by gesticulations and maps, many concepts in science are made explicit with the combined use of language and formulas, etc. Thus, language is a component, although a major component of human multimodal communicative ecology.

Communication in life forms is multimodal both at a behavioural and neuronal levels. Human communication, and language as a central part of it, although highly unusual, is also multimodal, demonstrating cognitive and communicative continuity. The language system has evolved as a reflection of its use in direct face-to-face dialogues and a vital component of human multimodal communicative behaviour. These behavioural changes are reflected in the evolution of the human body by prioritizing adaptation to spoken dialogues. The active participation of multimodality in spoken dialogues suggests that multimodality must have a prominent place in the study of language evolution.

REFERENCES

Aaltonen, O., Uusipaikka, E. 2005, Speech is Special: What is Special about It?
<https://web.stanford.edu/group/csli/publications/.../18-aaltonen-uusipaikka.pdf>

Arbib, M., 2007, The mirror system hypothesis, how did protolanguage evolve, in Tallerman, M. ed. Language origins, Perspectives on Evolution, Oxford University Press, p. 21-47)

Arbib, M. 2002, The mirror systems, imitation and the evolution of language, in Nehaniv, C., Dautenhahn, K. Eds. Imitation in animals and artifacts, MIT

Arbib, M. 2004, From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics, in Behavioural and Brain sciences , 28: 105-167

Arbib, M., Liebal, K., Pika, S., 2008, Primate vocalizations, gesture and the evolution of human language, in Current anthropology, 49/6: 1053-1076

Arbib, M. 2011, From mirror neurons to complex imitation in the evolution of language and tool use, Annual review of Anthropology, vol.40. p. 257-273

Arbib, M. 2017, Towards language ready brain: biological evolution and primate comparisons, Psychon Bull rev (2017) 24: 142-150

Aronoff M., Meir, I., Padden, C., Sandler, W., 2008, The roots of linguistic communication in a new language, Interaction Studies, 9(1):133-153, doi: 10.1075/is.9.1.10aro

Bickerton, D. 1984, The Language Bioprogram Hypothesis, in Behaviour and Brain Sciences, vol.7, p. 173 -221

Boe, J.L, Fagot, J. Perrier, P., Schartz, J.L. Eds. 2017, Speech production and perception, vol.4 Peter Lang Edition

Cartmill, E. , Beilock, S., Goldin-Meadow, S., 2012, A word in the hand: gesture and mental representation in humans and non-human primates, *Philosophical Transactions of the Royal Society*, 367, p.129-143

Chomsky,N ,1972 *Language and mind*, Cambridge Univ. Press

Chomsky, N. 1986 *Knowledge of language, Its nature, origin and use*, Greenwood Publishing

Chomsky, N. 2000, *New horizons in the study of language and mind*, Cambridge University Press

Corballis, M. 2003, From hand to mouth, the gestural origins of language, in Christiansen, M., Kirby, S. Eds. *Language evolution*, p.3201-218, Oxford University Press

Corballis , M. 2009, *The recursive mind, The Origins of Human Language, Thought and Civilization*, Princeton University Press

Corballis, M. 2009 a, The evolution of language, in *The Year in Cognitive Neuroscience*, *Ann.NY Acad. Sci.* 1156, p. 19- 43
doi: 10.1111/j.1749.6632.2009.04423.x

Christiansen, M. , Chater, N.,2008, *Language as shaped by the brain*, *Behaviour and brain sciences*, 31, 459-558

Dabrowska E. , 1997, *The LAD goes to School, a cautionary tale for nativists*, *Linguistics* 35 p. 735-766

Davis, B. , MacNeilage, P. 2004, *The frame-content theory of speech evolution: from lim smacks to syllables*, *Primates*, 6: 305-328

Dediu, D.,Levinson,S., 2018, *Neanderthal language revisited, not only us*, *Current opinion in behavioural sciences*, 21: 49-55, Abstract

Dediu,D., Levinson, S. 2013 *On the antiquity of language , the reinterpretation of Neanderthal linguistic capacities and its consequences*, *Frontiers in Psychology*,4:397
doi: 10.3389/fpsyg.2013.00397

Donald, M.1991 *Origins of the modern mind* , Harvard University Press

Dessalles, J. L., 2007, *Why we talk, evolutionary origins of language*, Oxford University Press

- Dor, D., Jablonka, E., 2000 From cultural selection to genetic selection, a framework for the evolution of language, *Selection*, 1, 1-3, 33-55
- Evans, V. , Green, M., 2006, *Cognitive linguistics, Introduction*, Edinburgh University Press
- Fitch T. 2002, *The evolution of language*, Oxford University Press
- Fitch ,T. Matur, N.,de Boer,B., Ghazanfar, A., 2016, Monkey vocal tracts are speech-ready, *Science advances*,2(12) DOI: 10.1126/sciadv.1600723
- Frohlich, M.et all. 2019, Multimodal communication and language origins: integrating gestures and vocalizations, *Biological Reviews*,Cambridge Philosophical Society, doi: 10.1111/brv.12535
- Garcia, R., Zamorano, F.,Aboitiz, F. 2014, From imitation to meaning: circuit plasticity and the acquisition of conventionalized semantics, *Frontiers in human neuroscience*, vol 8 art.605, doi: 10.3389/fnhum.2014.00605
- Greenfiels. P. 1991, Language, tools and brain, the ontogeny and phylogeny of hierarchically organized sequential behaviour, *Behaviour and Brain sciences*, (14) p.531-595
- Goldin-Meadow, S., Alibali, M. 2013, Gesture's role in speaking, learning and creating language, *Anu. Rev.Psyciol.* 64, p.257-283
- Haiman, J. 1983, Iconic and economic motivation, *Language* ,Vol.59, N0.4. p. 781-819
- Haspelmath, M., 2005, Iconicity vs.Frequency in explaining grammatical asymmetries [www.academia.edu/.../iconicity_vs_frequency_in_explaining_grammatical_ assime...](http://www.academia.edu/.../iconicity_vs_frequency_in_explaining_grammatical_assime...) DOI: <https://doi.org/10.1515/COG.2008.001>
- Hilliard A.,White S.,2009, Possible precursors of syntactic components in other species, in Bickerton D., Szathmary E. eds. *Biological Foundations and Origin of Syntax*, MIT Press, p.161-185
- Hurford, J., 2002 ,Language beyond our grasp, what mirror neurons can and cannot do for language evolution, in Oller, Griebel, Plunkett, eds. *The evolution of communication systems, a comparative approach*, MIT
- Hurford, J. 2012, *The origins of grammar*, Oxford University Press
- Jackendoff , R. 2002, *Foundations of language*, Oxford University Press
-

Jackendoff, R. 2010, Your theory of language evolution depends on your theory of language, in Larson, R., Deprez, V., Yamakido, H., eds. The evolution of human language, biolinguistic perspective, Cambridge University Press

Kendon, A. 2000, Language and gesture, unity or duality? , in McNeil, D. ed. Language and Gesture, Cambridge University Press, p. 47-63

Kendon, A., 2014, Semiotic diversity in utterance production and the concept of “language”, Philosophical Transactions of the Royal Society, B, 369, doi: 10.1098/rstb.2013.0293

Kendon, A. 2017, Reflections on “gesture first” hypothesis of language origins, Psychonomic Bulletin and Review, 24, p.163-170

Kirby, S. 1998, Language evolution without natural selection, Edinburgh Occasional papers in Linguistics, www.courses.csail.mit.edu

Langacker, R. 2008, Cognitive Grammar, A Basic Introduction, Oxford University Press

Lenneberg , E. 1967 Biological foundations of language, Wiley

Levinson ,S. Holler, J., 2014, The origin of human multimodal communication , Philosophical transactions of the royal society, B,vol. 369, issue 1651

Lieberman, Ph. 2000, Human Language and our Reptilian Brain, The Subcortical Bases of Speech, Language and Thought, Harvard University Press

Lieberman, P. 2002, On the Nature and Evolution of the Neural Bases of Human Language, in Yearbook of physical anthropology, 45, p. 36-62

Lieberman, Ph. 2006, Towards and Evolutionary Biology of Language, Harvard University Press

Lieberman, Ph. ,2016, The evolution of language and thought, in Journal anthropological sciences,vol 94, p.124-146

MacNeilage, P. 1998, The Frame-Content Theory of Evolution of Speech Production, BBS, 21, p. 499-546Frame-content theory of evolution of speech production

Morgenstern ,A., 2014, Children's multimodal language development, in Facke,C. ed. Manual of language acquisition, De Gruyter, p. 124-142

Mueller R. A., 1996, Innateness, autonomy, universality? Neurobiological approaches to

language, Behaviour and brain sciences Vol. 19 ,issue 4, p. 611-675

Muller, C. 2018, Gesture and sign, cataclysmic break or dynamic relations? *Frontiers in Psychology*, vol.9 art.1651

Perlman ,M. 2017 Debunking two myths against vocal origins of language, *Language is iconic and multimodal to the core*, *Interaction studies , Social behaviour and communication in biological and artificial systems*, 18(3)376-401,

Perniss, P., Thompson,R., Vigliocco, G.,2010, Iconicity as a general property of language , evidence from spoken and signed languages in *Frontiers of psychology*, 1:227,

Perniss P. , 2018, Why we should study multimodal language; *Frontiers in psychology*, 9:1109

Pinker, S. R. Jackendoff, 2005, *The faculty of Language, What is special about it?* , *Cognition* 95 , p. 201-236

Pinker S.,Jackendoff , R.,2009, *The components of language*, in Christiansen, Collins, Edelman, *Language universals*, OUP

Poizner, H., Bellugi,U., Klima, E.,1990, *Biological foundations of language, clues from sign language*; *Annual Review Neuroscience* ,13, 283-307

Rizzolatti, G., Arbib , M.1998, “, *Language within our grasp* “, *Trends in neuroscience*,vol. 21Issue 5, p. 188-194

Sandler, W. et all, 2005, *The emergence of grammar, systematic structure in a new language* , *Proceedings of the national academy of sciences*, vol.102, N. 7., p. 2661-2665

Schraff, C.,Petri,J., 2011,Evo-devo, deep homology and FOXP2: implications for the evolution of speech and language ,*Philosophical Transactions of Royal Society*, 366, 2124-2140, doi: 10.1098/rstb.2011.0001

Slocombe, K., Waller, B., Liebal , K. , 2011, *The language void: the need for multimodality in primate communication research*, *Animal behaviour*, 81 (2011) 919-924

Sherwood, C., Subiaul, F., Zawidzki, T., 2008, *A natural history of the human mind, Tracing evolutionary changes in brain and cognition*, *Journal of Anatomy*, Apr. 212 (4) p. 426-454, doi: 10.1111/j.1469-7580.2008.00868.x.

Sperber, D., G. Origgi 2010, *A pragmatic perspective on the evolution of language*, in Larson, R., Deprez, V., Yamakido, H. Eds. *The evolution of human language, a biolinguistic*

perspective, Cambridge University Press, chap.8, p.124-132

Sterelny, K. 2012, Language, gesture, skill, coevolutionary foundations of language, in *Philosophical Transactions of the Royal Society, B*, 367, p.2141-2151, doi:10.1098.rstb.2012.0116

Tomasello, M. 2008 *Origins of human communication*, MIT Press

Vernes, S., 2017, What bats have to say about speech and language, *Psychonomic bulletin and review*, vol. 24, issue 1, p. 111-117

Waller, B., Burrows, A., Liebal, K., Slocombe, K., 2013, How can a multimodal approach to primate communication help us understand the origins of communication, *Evolutionary psychology*, 11(3) 538-549

Wilson, D., Sperber, D., 2004, Relevance Theory, in L.Horn, G. Ward, eds. *The Handbook of pragmatics*, Oxford, Blackwell, p. 607-632

* Svetlana T. Davidova is a linguist, unaffiliated researcher based in Toronto, Canada
address for correspondence: svetlana.t.davidova@gmail.com
