18

Speech is Special: What's Special about It?

OLLI AALTONEN AND ESA UUSIPAIKKA

From the ethologist perspective, speech is to the human being as echolocation is to the bat or song is to the bird. Thus, speech, as well as the phonological communication it underlies, is plainly a species-typical product of evolution. Speech defined as the production and perception of vowels and consonants originates from a pre-phonetic capacity to perform speech sounds and gestures. Similarly, language defined as the syntax "machine" originates from a pre-syntactic capacity to organize longer sequences of sounds and gestures. We suggest therefore that the faculty of human language is biological and thus a product of evolution. Furthermore we suggest that formal language follows from speech which is based on motion (gestures) and perception of motion (sensorymotor perception of articulatory gestures).

18.1 Biological View of Language and Speech

Language depends on 'being human'. From a scientific perspective language is neither a divine gift nor a "cultural invention", but a product of human biological evolution. Spoken language evolved to make rapid vocal communication possible, providing man with a better chance of surviving in the struggle for existence (Darwin, 1874). Every normal individual acquires language in a uniform and automatic way by going through the same stages at the same age, without requiring specific instruction (Stromswold, 1996). Once learnt, the complex processes of speech production, perception and syntactic coding become automatized and are carried out below the level of conscious awareness, allowing the semantic content of the message to be the primary concern

Inquiries into Words, Constraints and Contexts.
Antti Arppe et al. (Eds.)

of the speaker or the listener. Hence, in order to say a word the speakers need not to know what sequence of sounds it comprises but only to think of its meaning. Indeed, they do not even have to know that it has a spelling. The specialized speech system automatically converts the phonological representation of the word into the coarticulated movements of the articulators that convey it. Correspondingly, to perceive a word, listeners need not puzzle out the complex and peculiarly phonetic relation between signal and the phonological message it conveys. All we have to do to perceive speech is to listen; somehow the meanings just emerge as the sounds go by. Again, the phonetic specialization automatically parses the sound so as to recover its phonetic structure. Hence processes of speech, whether in production or perception, are not calculated to put the speaker's attention on the phonological units that those processes are specialized to manage. Thus, a complex design, functionally completely different from animal communication, must have evolved for spoken language (Pinker 1994, pg. 362).

The phonetic units of speech are the vehicles of every language on earth, and they are commanded by every neurologically normal human being. Lai et. al. (2001) found a gene, FOXP2, which seems to be involved in speech. The regulating gene, located on chromosome 7, was discovered while studying a family most of whose members had troubles controlling their lips and tongue and forming words. More recently, Enard et. al. (2002) studied FOXP2's evolutionary history by comparing versions of the gene in various primates and mice. According to these comparisons FOXP2 has remained essentially unaltered during mammalian evolution, but it changed in humans after the hominid line of descent had split off from the closely related chimpanzee one. The changes in the gene are universal in human populations. Enard et. al. suggest that the changes affected articulation and they estimate that the human version of the gene emerged only 120,000 years ago. Perhaps this mutation of the FOXP2 was the final adjustment that allowed speech to become autonomous, freeing the hands for the development of technologies. Thus, writing and reading did not evolve as part of the language faculty and, therefore, writing and reading differ biologically from speech, being intellectual achievements in a way that speech is not. Many languages do not even have a written form, and, among those that do, some competent speakers find it impossible to master. Awareness of phonological structure is obviously necessary for anyone who would make proper use of an alphabetic script, but such awareness would not normally be a consequence of having learnt to speak.

18.2 Emergence of Symbolic Species

The Victorian people must have been quite shocked when Darwin presented in his Descent of Man (1874) that man evolved from apes. According to Klein's scenario (Klein, 2000), the first primate with bipedal locomotion (Ardipthecus ramidus) lived on African savanna roughly 4.4 million years ago and it took about 2 million years of additional evolution before the first crude tools appeared in the paleontological record about 2.5 million years ago. Brain expansion in homo line begins around 1.2. million years ago and the period of most rapid brain expansion occurred between 500 and 100 thousand years ago. However, all human fossils from 30,000 years ago to today share the same modern anatomical form: a distinct skull shape, a large brain (1,350 cubic centimeters), a chin and a lightly built skeleton. Neanderthals were as human as we are but something dramatic must have happened about 30,000 years ago when Neanderthals suddenly went extinct.

Neanderthals' disappearance coincided with the arrival of the anatomically modern Homo sapiens (Homo sapiens sapiens). Genetic evidence reveals that Neanderthal DNA is distinct from that of modern humans, and it implies that the two lineages diverged perhaps 400,000 years ago. Archeological artefacts left behind show that 100,000 years ago Neanderthals and Homo sapiens were quite similar culturally. However, about 40,000 to 50,000 years ago, a massive transformation occurred (Johanson, 2001; Klein, 2000). Tools became diverse and tailored for different purposes, burials became elaborate and hunters began to target dangerous large animals. This "creative explosion" was almost exclusively limited to Homo sapiens.

Deacon (1997) suggests that symbolic communication originating from new brain adaptations in Homo sapiens sapiens made possible better cultural information transmission from one generation to another and hence better organizational skills that permitted more efficient utilization of sources. Thus, a modern man was equipped with neural prerequisities for the use of symbols in communication, while the Neanderthals were evolved differently in this respect. Neither non-human primates seem to have this adaptation. According to an alternative explanation, there is no specific adaptation for the symbolization per se but adaptation was for understanding others on analogy with the self and symbols then developed as a kind of natural consequence (Chomsky, 1991; Tomasello, 2003). From the comparative perspective, probably the potential for symbolism exists in any animal with a brain of sufficient complexity.

Studdert-Kennedy and Goldstein (2003) suggest that once gestures of distinct organs had evolved as discrete, combinable units, expansion of the phonological systems have occurred by sociocultural processes without any further genetic change. On this view, speech as a motor function evolved

from phylogenetically ancient mammalian oral capacities for sucking, licking, swallowing and chewing (MacNeilage, 1998). For example, sucking, licking, and tongue actions for swallowing might have initiated neuroanatomical differentiation of the mammalian tongue, which the evolution of speech carried further by differentiating tongue tip, tongue body, and tongue root into independent organs of phonetic action. On this view, the basic unit of phonological structure is gesture, not the sounds those gestures produce.

It may be that that the human brain and body were at time 'language-ready' in the sense that the first Homo sapiens used a form of vocal communication which was but a pale approximation of the richness of language as we know it today. The Mirror System Hypothesis (Arbib, 2003) suggests that the functional specialization of human Broca's area derives from an ancient mechanism related to the production and understanding of motor acts. The mirror system's capacity to generate and recognize a set of actions provides the evolutionary basis for language parity, in which an utterance means roughly the same for both speaker and hearer. Therefore, Arbib (2003, pg. 194) states:

> extension of the mirror system from a system for recognition of single actions to a system for recognition and imitation of compound actions was one of the key innovations in the brains of hominids relevant to language.

According to this motor theory of speech (for a review, see e.g. Liberman 1996) the gestures are specifically phonetic, having evolved solely for the purpose of phonological communication. Therefore, apprehending phonetic structures has to be managed by a distinct, language-specific system that has its own phonetic domain and its own phonetic mode of processing served by a neurobiology of its own (speech module). The motor theorists suggest that the biology of language incorporates a precognitive specialization for the production and perception of vowels and consonants, and that perception of those is therefore immediate; there is no translation from a nonphonetic (auditory) representation because there is no such representation for speech.Thus, early hominids changed by adopting for communicative use an apparatus already divided into discrete units and specialized perceptual system for the recognition of articulatory gestures from the continuously varying acoustic signals (pre- or protolanguage). These adaptations in production and perception of speech finally resulted in symbolic communication by exhaled breath (language). Thus, in addition to changes in the organization of the brain, more peripheral adaptations were also needed for spoken language to be favoured in natural selection.

During the evolution of language speaking and listening became so tightly integrated that they seem to be merely two different manifestations of a single linguistic faculty rather than two separate abilities, coordinate but distinct. The distinction between speaking and listening is clear at the pheripheral

level, because they are based on fundamentally different organs. At a more central level, however, the distinction is less clear; speaking and listening simultaneously would not be so difficult, if they were not integrated at some point. Therefore, the processes of production and perception must somehow be linked and, consequently, their sensorimotor representations must, at some point, be the same. The evolution of brain guaranteed this parity (or the 'mirror' property) between speaking and listening and, thus, speech signals became more relevant for man than other acoustic signals in the environment and became linked to units of language (Liberman, 1996).

18.3 Language and the Brain

Inhuman rationalist principles in the philosophy of science have held on for centuries also in the study of language evolution. Accordingly, it has been a tacit assumption in linguistics and psychology that the purely physical or biological aspects of language should be distinguished from the psychological aspect, and that only the latter belongs to the study of language (Chomsky, 1965). Nevertheless, human language is primarily spoken, which suggests that its evolution must have been constrained by the speech apparatus and the auditory system. In recent years alternative views based on this perspective have emerged, indicating that rather than being two independent domains, the physical and psychological aspects overlap significantly (Diehl, 1991). Therefore, theories of language must link up with theories of brain function. Otherwise the study of language degenerates into a signal-processing oriented or a formalist discipline, both perfectly possible per se but remote from the study of what actually takes place in human beings when something is articulated or perceived.

From the evolutionary perspective, the brain was not built like a computer with a special design in mind but natural selection is responsible for its development. In this process of millions of years of evolution, new anatomical structures and functions developed in succession in relatively distinct stages from existing structures (Lamendella, 1980). These changes often involved increases in the anatomical size and configuration of particular structures, qualitative changes in physiological and functional organization, and increases in the overall information processing potential as existing structures took new functions. New structures arose and carried out old functions in new ways. Consequently, all parts of the brain are functionally integrated so intimately in the course of evolution that physically distinct neural movements for spoken language cannot be shown on the basis of the gross anatomy of the human central nervous system. Therefore, there is no single site for language in the brain but it is scattered all over the distinct parts of the brain.

Biologically human language originates from earlier pre-adaptations

which pave the way for subsequent adaptive changes (Hurford, 2003). For example, bipedalism set in train anatomical changes which culminated in the human vocal tract. Similarly, changes in human mental capacities were necessary before modern man became ready for language. These cognitive pre-adaptations set forward another process of evolution which led to the appearance of syntax relatively late in the history of man. Syntax involves the stringing together of independent subunits into a larger signal. In phonological syntax in units, like the speech sounds, there is no independent meaning, while in lexical syntax in the units, such as the words, there are meanings which contribute the overall meaning of the whole signal.

It is nowadays commonly accepted that language somehow emerges gradually from highly complex neuronal events which are firmly organized on a time basis. These neuronal events can be referred to as a kind of programme to emphasize the computational character of the higher-level brain functions. The term "serial action programme" (Ingvar, 1983) has been used in neurophysiology to refer to conceptual structures, which is a term used in linguistic literature for temporally organized neuronal events pertaining to language. According to Chomsky (2004), uniquely human component of the language faculty is syntax, varying little among humans and without significant analogue elsewhere. Thus, language is biologically isolated in its essential properties, and a rather recent development in human evolution. Chomsky (1991) has argued that language is not an adaptation at all, but rather is a by-product or side effect of the tremendous growth of the human brain. His argument is that after the brain attained its current size and complexity, language simply emerged spontaneously as one of many side effects. Despite arguing that language is not a designed adaptation produced by evolution, Chomsky nevertheless has argued that the deep structure of the grammar is innate rather than acquired, and universal in all humans.

18.4 Comparative approach to evolution of symbols and syntax

The faculty of language refers to the narrow syntax "machine" (faculty of language in the narrow sense, FLN), which is a computational system operating on syntactic symbols according to specific rules of computation, and generates an infinite number of utterances from a finite set of syntactic symbols (Hauser, Chomsky & Fitch, 2002). FLN represents a "language organ" per se, which is a subsystem of a more complex structure consisting of two interfaces: the Articulatory-Perceptual and the Conceptual-Intentional (faculty of language in the broad sense, FLB). The syntax "machine" was not instantaneously inserted into a mind/brain with the rest of its architecture fully intact. Rather, it is embedded within the broader architecture of the mind/brain and it interacts with other systems. Therefore, the systems within which the lan-

guage faculty is embedded must be able to communicate the expressions of the language and use them as guidelines for thought and action. Similarly, the sensorymotor systems have to be able to read their instructions having to do with sound and the articulatory and the perceptual systems have specific design that enables them to interpret certain properties, and not others. Thus, the focus in explanations of the language faculty shifted from the study of its subcomponents to their interrelations. Hauser and Fitch (2003) suggest that animals lack the capacity of recursion implying that FLN is an adaption produced by evolution, while subsystems that mediate speech production and perception are not. Many characteristics of speech production and perception are also present either in our closest living relatives or in other, more distantly related species.

Many bird species can learn songs with phonological syntax and apes are known to show a pre-syntactic capacity to organize longer sequences of sounds. Thus, it may be that combinatory principles underlying phonology and syntax of human language emerged gradually by a gradually enlarging brain providing more available neurons and more specialized connections between neurons, not greater intelligence per se (Chomsky, 1991; Bickerton, 2003). As a result of an enlarging brain, the modular and highly domain-specific system of recursion may have become penetrable and domain-general, because human mind cannot consist solely of isolated mechanisms that are completely walled off from each other. Selection favors functionally specialized mechanisms that work well together in various combinations and permutations (Buss, 2004). If recursion evolved to solve computational problems such as navigation, number quantifications, or social relationships, then it is possible that other animals have such abilities (Hauser, Chomsky & Fitch, 2002). Under these circumstances, FLN evolved as a by-product of evolution without any survival value.

There are two features of languages, in whatever modality they are expressed, that are generally not present among the communications of other animals: symbols and syntax. Symbolic communication arose first being within the reach of a number of non-human animals, while syntax emerged later remaining beyond the reach of any other species. Thus, protolanguage, with symbolic content but no syntactical structure evolved from different genetic and neural substrate than the subsequent language with syntax (Bickerton, 1995; Pinker, 1994). Okanoya (2003) studied complex vocalizations of Bengalese finches and suggests that Bengalese finches and humans follow similar developmental path. In both species, phonological development precedes syntactical development. Bengalese finches show syntactical control of singing, which may have evolved through the process of sexual selection. Thus, the rudimentary syntax might have evolved also in humans as a by-product of sexual selection without the need for survival value. In addition, there is also

some evidence that complex syntactic rules emerge from quite simple systems of networks, which have a very small number of initial assumptions and learn from imperfect inputs (Tonkes & Wiles, 2003).

18.5 Concluding Remarks

Computers can be programmed for various purposes; in this sense the computer is a domain-general information processor. The idea that there might be some information-processing problems that the human mind was specially designed to process was missing from the cognitive revolution in psychology. For example, the information processing view on speech perception sees the perception of speech as a wholly unexceptional example of the workings of an auditory modality that deals with speech as it does with all other sounds to which the ear is sensitive. In so doing, however, this view sacrifices a more important kind of generality, since it makes speech perception a mere adjunct to language, having a connection to it no less arbitrary than that which characterizes the relation of language to the visually perceived shapes of an alphabet. However, speech is special, but neither more or less so than any other biologically coherent adaptations, including language itself. Thus, the specializations for phonetic and syntactic perception have in common that their products are deeply linguistic, and are arrived at by procedures that are similarly synthetic.

References

Arbib, M. A. 2003. The Evolving Mirror System: A Neural Basis for Language Readiness. In M. Christiansen and S. Kirby, eds., *Language Evolution*, pages 182–200. New York: Oxford University Press.

Bickerton, D. 1995. *Language and Human Behavior*. Seattle: University of Washington Press.

Buss, D. 1933. *Evolutionary Psychology: The New Science of Mind*. New York: Pearson.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: Cambridge University Press.

Chomsky, N. 1991. Linguistics and cognitive science: Problems and mysteries. In A. Kasher, ed., *The Chomskyan Turn*, pages 26–53. Cambridge, MA.: Blackwell.

Chomsky, N. 2004. Language and Mind: Current thoughts and ancient problems. In L. Jenkins, ed., *Variation and Universals in Biolinguistics*, pages 379–405. Amsterdam: Elsevier.

Darwin, C. 1874. *The Descent of Man*. New York: Hurst & Co.

Deacon, T. W. 1997. *The Symbolic Species. The Co-evolution of Language and the Brain*. New York: W. W. Norton & Company.

Diehl, R. L. 1991. The role of phonetics in the study of language. *Phonetica* 48:120–134.

Enard, et. al., W. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418:869–872.

Hauser, M., N. Chomsky, and T. Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve. *Science* 298:1569–1579.

Hauser, M. and W. Fitch. 2003. What are the Uniquely Human Components of the Language Faculty? In M. Christiansen and S. Kirby, eds., *Language Evolution*, pages 158–181. New York: Oxford University Press.

Hurford, J. R. 2003. The Language Mosaic and its Evolution. In M. Christiansen and S. Kirby, eds., *Language Evolution*, pages 38–57. New York: Oxford University Press.

Ingvar, D. H. 1983. Serial aspects of language and speech related to prefrontal activity: A selective review. *Human Neurobiology* 2:177–189.

Johanson, D. 2001. Origins of modern humans: Multiregional or out of Africa? *www.actionbiosciences.org* .

Klein, R. G. 2000. Archeology and the evolution of human behavior. *Evolutionary Anthropology* 9:17–36.

Lai *et. al.*, C. S. L. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–523.

Lamendella, J. T. 1980. Neurofunctional Foundations of Symbolic Communication. In M. L. F. . S. H. Brandes, ed., *Symbol as Sense*, pages 147–174. New York: Academic Press.

Liberman, A. 1996. *Speech: A Special Code*. London: The MIT Press.

MacNeilage, P. F. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21:499–546.

Okanoya, K. 2003. Sexual display as a syntactical vehicle: The evolution of syntax in birdsong and human language through sexual selection. In A. Wray, ed., *The Transition to Language*, pages 46–63. New York: Oxford University Press.

Pinker, S. 1994. *The Language Instinct*. Middlesex, England: Penguin Press.

Stromswold, K. 1996. The Cognitive and Neural Basis of Language Acquisation. In M. S. Cassaniga, ed., *The Cognitive Neurosciences*, pages 855–870. Cambridge, Mass.: The MIT Press.

Studdert-Kennedy, M. and Goldstein L. 2003. Launching Language. In M. Christiansen and S. Kirby, eds., *Language Evolution*, pages 235–254. New York: Oxford University Press.

Tomasello, M. 2003. On the Different Origins of Symbols and Grammar. In M. Christiansen and S. Kirby, eds., *Language Evolution*, pages 94–110. New York: Oxford University Press.

Tonkes, B. and J. Wiles. 2003. Methodological issues in simulating the emergence of language. In A. Wray, ed., *The Transition to Language*, pages 226–251. New York: Oxford University Press.