

Review of Peter MacNeilage's "The Origin of Speech"

"The Origin of Speech" is a book that reflects Peter MacNeilage's ideas about the nature and the origin of speech and language. It is at times an exposition of MacNeilage's theory of the origin of speech, at times an overview of behavioral and neurological organization underlying speech and at times a rather strong critique of mainstream phonology. It presents an unorthodox view of speech, but one that in my opinion meshes better with biology than most other existing theories of speech. MacNeilage's attempt to work out a detailed biological perspective on speech make this book worthwhile, even though both the book and MacNeilage's theory have their share of weaknesses.

The book's theme is the biological evolution of speech. Its aim is to present a plausible scenario of the evolution of the first complex utterances. It focuses on Peter MacNeilage's frame/content theory of the origin of speech, but it also contains a large number of other observations that the author finds relevant for the evolution of speech. Among these are the presentation of the postural origin theory of hemispherical specialization, a comparison of the structure of signed language and spoken language, and a thorough deconstruction of generative phonology.

MacNeilage's (and Davis', 2000) frame/content theory proposes that speech is based on two types of movement. One type is the cyclical motion of the jaw. This results in repeated opening and closing of the mouth. When combined with vocal fold vibration, this produces syllable-like utterances, where the closed part of the cycle sounds like a consonant and the open part sounds like a vowel. One opening-closing cycle is called a frame. On these frames motions of other articulators, such as the tongue or the lips, can be superimposed. This is the content. According to MacNeilage, the combination of these two behaviors produces syllables.

The frame/content theory makes a number of predictions about speech. Three of these are: certain consonants and vowels tend to co-occur (namely those with similar settings of the content-related articulators), vowels and consonants are different entities, and distinctive features are artifacts of description, but have no reality in speech. MacNeilage presents evidence from cross-linguistic studies and studies of infant babbling that shows that the co-occurrences predicted by the theory (labial consonants with mid vowels, lingual consonants with front vowels and dorsal consonants with back vowels), do indeed occur preferentially. He also presents evidence from speech errors showing that consonants are interchanged with consonants and vowels with vowels but that consonants and vowels are never interchanged. Interchanges at the level of the distinctive feature do not happen, either.

In the book the frame/content theory is extended to an evolutionary scenario. MacNeilage makes an effort to do this within the neodarwinian biological framework. He bases his investigation on the four questions that, according to the ethologist Niko Tinbergen, should be asked about any animal behavior (Tinbergen, 1963). These questions are: how does the system work, what does it do for the organism, how does it develop in ontology

and how did it evolve. MacNeilage adds a fifth question of his own: how does it work in the brain?

Using the frame/content theory, the first three of these questions can be answered according to MacNeilage. It provides an answer to the question of how speech works – as a combination of frames with content. It also provides a scenario of how speech develops in infants. They start producing pure frames, and add content motions later in development. As for the function of language, its adaptive value is clear in modern humans, but it cannot be assumed that at earlier stages of evolution it was able to fulfill the complex communicative function that it fulfills today. MacNeilage favors the theories that the earliest forms of speech-like communication were used for vocal grooming and parent-child interactions, as developed by Dunbar (1996) and Falk (2004) respectively.

Using these elements, MacNeilage presents an evolutionary scenario to answer Tinbergen's fourth question. He proposes that the frames of speech developed from lip smacks such as those used by several species of macaques and baboons. These lip smacks would make use of even older mammalian motor programs used for sucking, licking and chewing. In combination with phonation, they result in a small repertoire of signals. These signals cannot be considered words of a language yet, as they have no learned and therefore culturally transmitted meaning. The crucial step of attaching learned meaning to these signals occurred in MacNeilage's scenario when mothers interpreted their infants' first pure frame utterances as a word for themselves: [mama] means "mother".

The discovery that arbitrary meanings could be attached to vocal utterances, causes a pressure for creating more and more word forms. This development was made possible by humans' ability to learn through imitation. In order to accommodate more and more different words, syllable variegation became necessary. By using motions of the other articulators, variation of the pure frame could be made, thus creating a set of possible building blocks. As the frame itself is a repetitive motion, this almost automatically leads to a combinatorial system, where a small number of building blocks is chained together to form a much larger set of longer utterances. At this stage, speech consists of combinatorial utterances with arbitrary meaning, and thus can be considered equivalent to modern speech.

MacNeilage also devotes three chapters in his book to the organization of the brain underlying speech and the frame/content theory. He presents evidence from studies of aphasia indicating that the frame and its content are controlled by different brain areas. He also uses evidence from the study of mirror neurons to support the hypothesis that they are crucial in the imitative abilities necessary for learning and culturally transmitting speech. Mirror neurons are neurons that fire both when a certain action is observed and when the same action is performed. They could therefore help in mapping observed actions to motor behavior needed to imitate such actions. Finally, MacNeilage presents evidence that hemispherical specialization of the brain is not unique to humans, but exists already in all primates. He advocates the postural origin of hemispherical specialization in the brain, where arboreal living made it necessary for one half of primate's bodies to specialize in holding on to a tree, while the other half was used for quickly grasping

insects. This lead to specialization of the left hemisphere for routine tasks, while the right hemisphere became more specialized in risk-related emergency behavior. Language, having to do with routine behavior, therefore uses left-hemisphere parts of the brain, while brain areas having to do with emotion are located more in the right hemisphere.

Throughout the book, MacNeilage stresses the importance to focus on *action*. In order to understand evolution of speech, it is necessary to understand the physical action on which it is based, and not just the mental mechanisms that underlie it. He argues that even though the interest in embodied cognition is increasing, action is still neglected in both linguistics and cognitive science, using the lack of attention to action in many reference works on cognitive science as an illustrative example. MacNeilage also argues that in most modern theories of phonology, attention is only paid to mental aspects of speech, while aspects of physics and action are ignored. In such theories, the focus is on an abstract notion of competence, while actual performance is ignored.

According to MacNeilage, this focus on the abstract notion of competence in linguistics makes it impossible to build a biological account of speech (or language) based on Tinbergen's four questions. This is part of another theme that occurs throughout the book: a thorough deconstruction of generative phonology and related theories of speech, such as optimality theory. He devotes numerous sections of the book to arguing why generative phonology does not provide correct answers to Tinbergen's four questions and why it is fundamentally unable to do so for (at least) the question about evolution. Although MacNeilage provides arguments for why a number of basic constructs of generative phonology (such as distinctive features, markedness and the distinction between competence and performance) are deeply flawed, his most fundamental objection is that generative phonology is in his view an essentialist theory, based on the assumption that universal grammar is an abstract ability that exists in the human mind and that is separate and independent from the body. He maintains that such a view is static, and cannot be the basis of a model of descent by gradual modification required by an evolutionary scenario.

In order to illustrate that there is no mental, modality-independent component devoted to phonology, MacNeilage compares the organization of sign language with that of spoken language. He concludes that spoken language has a clear structure that is derived from frame/content organization and that this structure is shared with bird song. The structure consists of modified closed-open cycles. Bird song and human speech evolved separately, and therefore their similar organization must be due to physical constraints related to how speech is produced. Sign language has its own unique organization, using location, movement and hand shape, all of which occur simultaneously. This structure makes it ideally tailored to the visual/gestural modality, and is therefore also explicable as the result of physical and action related constraints.

In his chapter about sign language, MacNeilage also devotes some space to the theory that sign language is actually an evolutionary precursor of spoken language. Although he allows for the possibility that simple gestures have always been part of human communication, he argues against the possibility that a fully combinatorial system of

signs first developed and was later replaced by a spoken language. He does this on the ground of continuity between the structure of spoken language and that of lip smacks and other cyclical oral behavior, but also on the more theoretical consideration that evolution tends to stick with a given solution. It is very unlikely that in evolution one behavior is replaced by another, but it is much more likely that the existing behavior is developed and modified through evolution. A switch from a fully functional signed language to a spoken language would therefore be extremely unlikely, according to MacNeilage.

If this seems like a lot of information, that impression is correct. The book covers a lot of ground, and brings together many different topics from linguistics, biology, neuroscience and psychology. Many of the non-linguistic topics are not well-known within linguistics, but are nevertheless relevant for understanding the nature and the origins of linguistic behavior. Conversely, the linguistic topics are covered in a way that is unusual (and perhaps controversial) in linguistics. This treatment is interesting because it puts many assumptions and abstractions that linguists tend to take for granted in a different light. Broadness of subject matter and an original perspective are good for a book on a topic that is as interdisciplinary and as new and exciting as the study of the evolution of language. On the other hand, I have the impression that MacNeilage's tone is at times too polemical. This causes his coverage to be broad and original, but at the same time rather one-sided.

As theories of the evolution of language go, the frame/content scenario is rather good. It is well defined in subject matter, in that it is about the origin of combinatorial, learned speech in populations of already highly social and cooperative early hominids. The pre-existing behaviors it assumes (vocal grooming, lip smacks, mandibular oscillations and a budding ability for imitation) have been attested in our primate relatives and are therefore not implausible for early hominids. The evolutionary changes that the theory supposes to have happened are relatively small and biologically and neurologically plausible. Furthermore, MacNeilage makes a real effort to present his scenario in a biological framework by using Tinbergen's four questions, and by trying to keep in mind what a Darwinian explanation should be like. Such detailed reference to biological theory is actually too rare in papers and books on the evolution of language, and MacNeilage's attempt is therefore important.

However, it seems as if in places MacNeilage gets carried away with his enthusiasm for his own theory on the one hand and with his dislike of generative phonology on the other. Even though the frame/content theory is elegant and not implausible, it is nevertheless based on a rather narrow base of evidence. There is the evidence from child babbling, which does appear to support a frame/content like organization of modern speech. There is also some evidence from aphasia that points to frame/content organization of speech in the brain. However, an argument against the theory is that frame-like vocalizations are not made by apes, and MacNeilage's reliance on lip smacks (that appear to have been studied in macaques and baboons, rather than apes) is less convincing. Furthermore, MacNeilage states on page 92 that "Primate vocalization systems typically consist of a relatively small number of calls", and cites a number of 22 vocal patterns in gelada baboons. This number is not much lower than the 36 bisyllabic patterns that can be

produced with the simplest variegated frame/content syllables and actually an order of magnitude higher than the two unvariegated pure frame words [māmā] and [baba] that MacNeilage proposes as the first words. Using MacNeilage's own argument against the primacy of sign language, one could ask why a system of holistic utterances would be replaced by an initially much less powerful system of combinatorial frame/content utterances. Although the above issue might be resolved, the example illustrates that the frame/content scenario still has unresolved weaknesses.

Although MacNeilage makes an important effort to propose an evolutionary scenario that is in line with biological theory, in the end his scenario still falls short of a full evolutionary explanation. His use of Tinbergen's four questions is a step in the direction of a more biologically grounded theory of speech and language. However, a biologically grounded description of a phenomenon and a set of likely steps that could have occurred in its evolution do not constitute an evolutionary explanation. A more appropriate list of criteria for understanding evolutionary adaptations has been proposed by Parker and Maynard-Smith (1990) and was probably first applied to language evolution by Zuidema (2005). This list contains five elements that an evolutionary explanation should have: a biological question about evolutionary adaptation, a range of possible strategies, an optimization criterion, a relation between different strategies and fitness, and finally experimentally testable predictions based on the choice of strategies, optimization criterion and fitness. Of course, it is very difficult for any theory about a phenomenon as complex as speech to be explicit about all these elements. However, consideration of Parker and Maynard-Smith's criteria can help to point out where exactly the weaknesses of a given scenario lay. MacNeilage does a good job in posing a biological question, proposing a fitness criterion (which has mostly to do with being able to produce an unlimited number of distinguishable utterances) and in posing experimentally testable questions. However, he fails to analyze the effect of the possible alternative strategies, and how these would interfere with optimization of fitness. As has been pointed out above, holistic utterances are apparently a viable alternative strategy for many monkey species, and it is not clear how and why these could have been replaced by the (initially very limited set of) combinatorial utterances that are produced by the frame/content mechanism. Furthermore, he does not explain how associating a learned meaning with a *single* infant vocalization would influence fitness, and how this could spread through a population.

The weaknesses of MacNeilage's scenario would not be so much of a problem if the tone of the book was less polemic. However, as it is, MacNeilage devotes quite a lot of space in the book to deconstructing generative phonology, and other theories that he perceives to be in the same paradigm, including optimality theory and Juliette Blevins' (2004) evolutionary phonology. Now I agree with most of what MacNeilage says, but I cannot help to feel that he is attacking mostly straw men. Even though generative linguists work in a paradigm that is not conducive to evolutionary explanation, there are more than a few people working in that field who have given thought to how it can be incorporated in an evolutionary framework (although, it needs to be said, phonology has received less attention than other aspects of language). The extremely static and essentialist picture that MacNeilage paints of generativist phonology gives the impression of being a caricature

and its deconstruction is therefore less convincing than a more nuanced critique could have been. I at least could not help thinking: “Surely there must be more nuanced positions in generative phonology than this.”

In the end, I think that both the strengths and the weaknesses of the book are due to the fact that it is not just on the origin of speech, but also the reflection of an important part of MacNeilage’s life’s work. This consists of the understanding of complex actions (amongst which speech) with models that are less based on innate, mental, static abstractions, but more on concrete actions, concrete neurological data and biological continuity with other species. Hence one can understand his insistence on an action-based theory of the origin of speech, his inclusion of the postural origin model of the hemispherical specialization, and his impatience with extreme generativists. Read in this light the book turns out to be not so much about a theory on the origin of speech, as well as about the expression of a scientific career’s worth of views and observations by an original and very interdisciplinary cognitive scientist, and therefore all the more worth reading.

References

- Blevins, J. (2004). *Evolutionary phonology*. Cambridge: Cambridge University Press.
- Dunbar, R. I. M. (1996). *Grooming, gossip and the evolution of language*. London: Faber and Faber.
- Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, 27(4), 491-541.
- MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288, 527–531.
- Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, 348, 27-33.
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410-433.
- Zuidema, W. (2005). *The major transitions in the evolution of language*. Unpublished PhD., University of Edinburgh, Edinburgh.