Emergence of a Language Through Deictic Games Within a Society of Sensori-motor Agents in Interaction: An Evo-devo Anthropomorphic Perspective

Clément Moulin-Frier[1], Jean-Luc Schwartz[1], Louis-Jean Boë[1], Julien Diard[2] & Pierre Bessière[3]
[1] GIPSA-Lab, DPC (Ex-ICP), Grenoble
[2] Laboratoire de Psychologie et de NeuroCognition (LPNC), Grenoble
[3] Laboratoire d'Informatique de Grenoble (LIG-Lab), Grenoble
France

FirstName.LastName@gipsa-lab.inpg.fr; Julien.Diard@upmf-grenoble.fr; Pierre.Bessiere@imag.fr

Since the 70's and the Lindblom's proposal to "derive language from non-language", phoneticians have developed a panel of "substance-based" theories. The starting point is Lindblom's Dispersion Theory (Lindblom, 1984) and Stevens's Quantal Theory (Stevens, 1989), which open the way to a rich tradition of works attempting to determine and possibly model how phonological systems could be shaped by the perceptuo-motor substance of speech communication. These works search to derive the universals of world languages from morphologic constraints arising from perceptual (auditory and perhaps visual) and motor (articulatory and cognitive) properties: we call them "Morphogenesis Theories".

More recently, a number of proposals were introduced in order to connect pre-linguistic primate abilities (such as vocalization, gestures, mastication or deixis) to human language. The framework that we adopt in the present work considers a double evolution grounding the birth of language in two pre-existing primate behaviours. Firstly, we assume that human language is derived from a precursor deictic function, providing an evolutionary development of the ability to "show with the voice" ("Vocalize to Localize", Abry *et al.*, 2004). Secondly, we capitalize on the "Frame-Content Theory" introduced by (MacNeilage, 1998) assuming that mastication provides a system enabling to modulate the vocalizations in a controllable way, thus allowing the birth of an efficient orofacial communication system, bootstrap by the deictic function. The Frame-Content Theory also offers a powerful developmental scheme for the mastering of serial order in speech motor control (MacNeilage & Davis, 2000). This offers an evo-devo framework for an "Origin Theory".

We propose that the principles of Morphogenesis Theories (such as dispersion principles or the quantal nature of speech) can be incorporated and to a certain extent derived from Origins Theories. While Morphogenesis Theories ask questions such as "why are vowel or consonant systems shaped the way they are?" and answer that it is to increase auditory dispersion in order to prevent confusion between them, we ask questions such as "why do humans attempt to prevent confusion between percepts?" and answer that it could be to "show with the voice", that is to improve the pre-linguistic deictic function. We illustrate this point of view Figure 1.

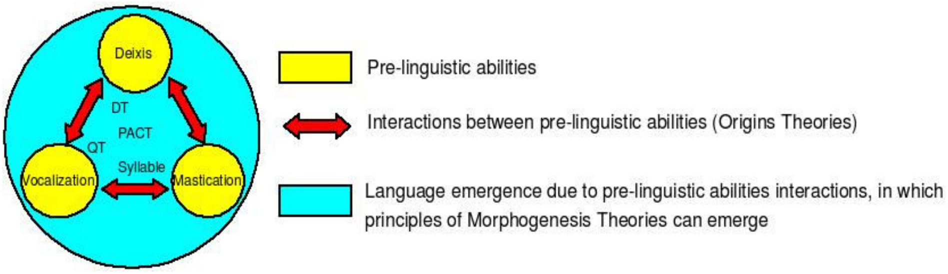

Figure 1 – *Principles of Morphogenesis Theories can emerge from Origins Theories [DT : Dispersion Theory (Lindblom, 1972); QT : Quantal Theory (Stevens, 1989) ; PACT : Perception for Action Control Theory (Schwartz & al., 2007)]*

We build a computational Bayesian model incorporating the Dispersion and Quantal Theories of speech sounds inside the Vocalise-to-Localise framework, and show how realistic simulations of vowel systems can emerge from this model. The model is based on computational sensori-motor agents able to produce orofacial motor gestures (M), perceive sounds (P), face objects (O), and associating these variables into a probabilistic rule p(O,M,P). Then, we design various interacting situations, called "deictic games", in which two agents are in front of an object O. One produces a motor command M, the other perceives the corresponding sound P, and both update its knowledge according to the interaction. (Fig. 2).

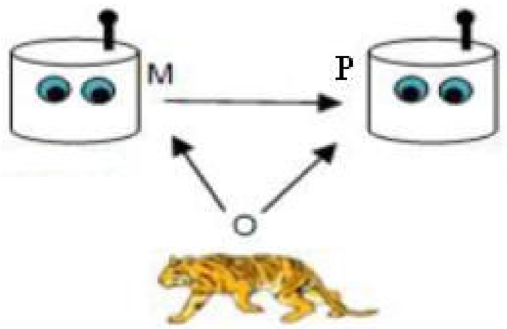

Figure 2 – *A deitic game between two interacting sensori-motor agents. The left agent sees an object O and produces a motor command M. The right agent perceives the sound P generated by the articulatory command M. Both update its knowledge according to this interaction.*

We compare various interaction configurations, called "reflex", communicative" and "hybrid", and show that in the "communicative" configurations, a common phonetic code can emerge from interactions (Fig. 3), and display a number of patterns compatible with the Dispersion Theory and with the Quantal Theory. The "hybrid" configuration associates motor and perceptual constraints on sound systems, in a way compatible with the "Perception-for-Action Control Theory" that we have developed in the last years (Schwartz *et al.*, 2007). Simulations with an elaborate model of the vocal tract (VLAM, Boë *et al.*, 1995) lead to realistic simulations of vowel systems.

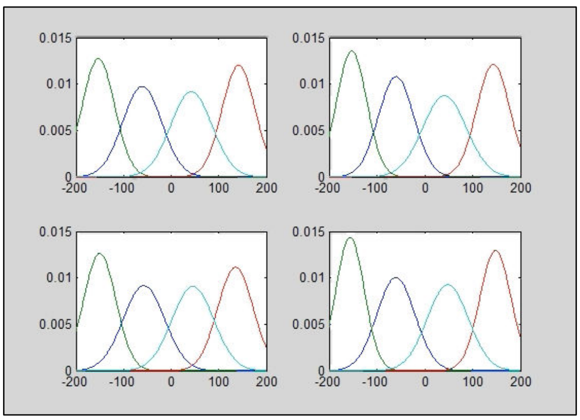

Figure 3 – *Emergence of a common phonetic code for four objects and four agents. In these*

preliminary simulations, the motor and perceptual variables M and P are merged, and monodimensional. After a series of interactions, the four agents (one in each box in the figure) have converged towards similar shapes of the distribution function for gestures associated to each of the four objects (the four bell shapes in each box). In consequence, the communication is efficient between the agents

In a further step, we derive the emergence of plosives from the generation of pseudo-syllabic alternations through jaw cycles in the Frame-Content framework. We show that acoustic/auditory nonlinearities shape the simple jaw rhythmic activity in a quantal pattern, achieving the generation of alternations of vowels and consonants in a simple way developmentally plausible and efficient. Then we show that bilabials, dentals and velars (*e.g.* [b d g]) provide an optimal system in terms of auditory dispersion, provided that they are embedded in this developmental framework, pharyngeals, though auditorily salient, being eliminated by their high jaw configuration incompatible with the Frame-Content scenario.

REFERENCES

Abry, C., Vilain, A. & Schwartz, J.-L., 2004. 'Vocalize to Localize'? A call for better crosstalk between auditory and visual communication systems researchers. "Vocalize to Vocalize", Special issue of *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 5, 313-324.

Boë, L.-J., Gabioud, B., & Perrier, P., 1995. Speech Maps Interactive Plant « SMIP ». *Proc. XIIIth International Congress of Phonetic Sciences*, Stockholm, Sweden, 426-429.

Lindblom, B., 1984. Can the models of evolutionary biology be applied to phonetic problems ?. *Proceedings of the 10th International Congress of Phonetic Sciences*, Utrecht, 67–81.

MacNeilage, P. F., 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499-511.

MacNeilage, P. F. & Davis, B. L., 2000. On the origin of internal structure of word forms. *Science*, 288, 527--531.

Schwartz, J.L., Boë, L.J., & Abry, C., 2007. Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In M.J. Solé, P. Beddor & M. Ohala (eds.) *Experimental Approaches to Phonology*. OUP (to appear).

Stevens, K.N., 1989. On the quantal nature of speech. *Journal of Phonetics*, 17, 3–45.