IS POINTING THE ROOT OF THE FOOT?
GROUNDING THE «*PROSODIC WORD*» AS A *POINTING WORD*

CHRISTIAN ABRY

Language Sciences Department, Stendhal, BP 25 FRANCE-38040 Grenoble Cédex

VIRGINIE DUCEY

Virginie Ducey, GIPSA-Lab
Stendhal, BP 25 FRANCE-38040 Grenoble Cédex

Recently in the *Vocalize-to-Localize* framework (a functional stance just started in the *Interaction Studies* 2004-2005 issues we edited, Abry et al., 2004), we addressed the unification of two grounding attempts concerning the *syllable* and the *foot* in language ontogeny. Can the movement time of the pointing strokes of a child be predicted from her babbling rhythm? The answer for 6 babies (6-18 months) was a 2:1 pointing-to-syllable ratio. Implications for the grounding of the first words within this *Pointing Frame* will be examined. More tentatively we will suggest that babbling for protophonology together with pointing for protosyntax pave the way to language.

1. Introduction

While the main scientific endeavour is *fission*, say first break already known units, as in physics typically, the afterthought of formal constructions is to restart from primitives, e.g. building blocks. This is the foundational Chomsky & Schützenberger's *free monoid* for computational linguistics, then *Move and/or Merge* in the Minimalist Programme (MP). In physiological behavior the degrees-of-freedom problem is rather seen developmentally as a problem of breaking early given coordinations (e.g. thumb-sucking *in utero*, Babkin's reflex, etc.) in order to elaborate new couplings for new skills (hand-to-mouth feeding... piano playing).

2. Emergence as mergence: *Sign+Sign=>Sign* and *Foot+Foot=>Foot*

Regarding the emergence of phonology, some students like Lindblom and ourselves have considered that features, particles, primes, etc., are just *by-products* of other mechanisms (for a recent tentative reconciliation with the use

of features within our *Perception-for-Action-Control-Theory*, see Schwartz, Boë & Abry, 2007). But what are the unit of the system you start from? The number of segments? The possible onsets and offsets of syllables...? In computational evolutionary phonology, the issue is still between a holistic-formulaic starting point, or a yet undefined layman word unit. This in spite of our linguistic state-of-the art, since «we still do not have strict definitions of even the most basic units, such as segment, syllable, morpheme, and word», as complained by Joan Bybee (2003, p. 2).

Now instead of fission, can *fusion* help? In other words can the compositional making of larger units from smaller bricks, be replaced by the blending of already more or less large units, typically two into one unit of the *same level* (an idea taken earlier in the *categorial grammar* formalism, still compatible with MP)? Which of course leaves open the evolutionary issue about where they could come from.

Let us take an example from a still-on-the-making phonology. In Sign Language, where no stable consensus does exist about phonological units, can one use semantic blending and morphological fusion to evidence these components? In ASL, *MIND+DROP=>FAINT* (we are indebted to Wendy Sandler for this videoclip example). If *Sign+Sign=>Sign* is semantic blending (*snowman*), what are the corresponding phonological units? Is there a sign-language specific «syllable conspiracy», as Sandler claims: *Syll+Syll=>Syll*? Or a more common foot isochrony *Foot+Foot=>Foot*? Like one-foot *music, musical, musically*? *Snowman* is obviously shorter than *snow+man* duration. In fact, once measured, the downstroke phase of *FAINT* (which starts from the head for *MIND*, with the finger point erased) is just a videoframe longer than the one for *DROP* (starting lower from the waist). Which is a strong cue of isochrony control for compression in one unit (chunk, template, etc.).

Is that just emergence-supervenience of units due to informational constraints, just language-use, the war of attrition on constructions as form-meaning pairings, in cognitive construction grammars? Said otherwise: data compression for *sparse coding*? Are there no macroscopic units corresponding to universal *control* units, macroscopic primitives for making morphogenetic «language bubbles», not acquired simply by perceptuo-motor statistical pattern-finding? Are there phonologically universal babble-syllable constraints in speech acquisition, and more, signs and words in both speech and sign language (even if syllables could be not ubiquitous in both media)? In other words, when in evo-development do you get a tuner for tuning? Who could attune what, along language attunement-imitation, without a specific what-tuner to capture the preferred radiostation among the buzzy broascasting landscape of speakers?

3. The *syllable*, then the *point*: whence the *word*?

Recently in the *Vocalize-to-Localize* framework (a functional stance just started in the *Interaction Studies* 2004-2005 issues we edited, see Abry, Vilain & Schwartz, 2004), we addressed the unification of two grounding attempts concerning the *syllable* and the *foot* in language ontogeny. Both units are highly disputed among phonologists and psycholinguists. But the proposal of a root for proto-syllables in canonical babbling can now be neurally evaluated on the basis of a motor control platform: MacNeilage's *Frame/Content* theory starting from the control of the mandible as the carrier articulator. We proposed the same ground of evaluation for the *foot* as the basic control unit for the phonology of the *proto-word*. We predicted that, if we would measure the babbling rhythm of a baby from the burst of canonical babbling around 6-7 months, we could calculate the range of duration of her pointing arm-strokes, from 9 months upwards. Tested on 6 French children in a longitudinal study, each fortnight between 6 and 18 months, this «astonishing» hypothesis was globally successful (Ducey, 2007), with a mean 2.18 pointing/babbling ratio. Moreover each child had at her disposal in her repertoire a sufficiently long point to cover a disyllabic utterance.

Like for linguistic demonstratives, the semantics, pragmatics, and even the syntax of pointing have all deserved valuable attention and brought out results in related fields. And Sign Language phonology too, which meets ubiquitously pointing. But nothing was said about the proper phonological integrative links of the pointing gesture with speech phonological units, smaller or larger than the point, like the syllable, the foot, and the so-called «prosodic word».

We can now consider that the phonology of the point with the arm-index could give for free the template of the ubiquitous one/two-syllable word foot (instead of an arbitrary FOOTBIN in Optimality Theory, where a one-syllable/moraic foot is considered as «degenerated» or «subminimal»). Grounding the phonology of the point motorically, in the neural arm-index control, gives thus for free the template of the two-syllable word as a coordination of the hand and the mouth in language semiotics and phonetics. This result offers in addition considerable insights in line with the parallel development of syntax use of THat-demonstratives and WHat-interrogatives through the grammatization process in the world's languages (Diessel, 1999). It is in favor of an early *demonstrative site*, later attuned to language specific morphonology: see English *(the) house* vs. Swedish *huset*, French *la maison* vs. Rumanian *domul*; and even more elaborated compounding, with what could be

tagged «double filled sites»: French *cette maison-ci vs.* Swedish *det här huset,* or Afrikaans *hierdie huis,* etc.

This is just one of the issues, the developmental framework reminded below (Fig.1), allowed us to address up to now, in between the *Vocalize-to-Localize* (2003) seminar and the 2007 *VOCOID (VOcalization, COmmunication, Imitation, and Deixis, in infant and adult human and non-human primates)*, both international meetings we organized in Grenoble.

4. Beyond the presented Framework (Fig.1)

Beyond reinforcing the very general claim that «pointing is the royal road to language for babies» (as recalled by the late George Butterworth in Kita, *Pointing,* 2003), we can add to our prediction of pointing stroke duration distributions from individual babbling rhythm distributions another replicated prediction: namely the prediction that two-word utterance emergence can be calculated from the beginning of the coproduction of a word together with a non redundant pointing (a result found in Susan Goldin-Meadow's group, and replicated with Jana Iverson in Iverson & Goldin-Meadow, 2005). Since this is not a pure slot-grammar story (POINT+Word gives Word+Word, but the POINT is still there in the predicate-argument structure), the rationale behind this development beyond the first year word, remains still a lot mysterious (personal conversation with Susan Goldin-Meadow and Elena Lieven).

Finally we will add work in progress on two possible neural circuits found in adults, which could be relevant for language acquisition of the word-foot metric unit, namely the one we dubbed the THAT-PATH, for pointing with the eye, the arm and the voice (Loevenbruck *et al.,* 2005, 2007). And ultimately the verbal working memory network, we dubbed the STABIL-LOOP (Abry, Vilain & Schwartz, 2004), for stabilizing the linguistic word forms (Abry *et al.,* 2003, Sato *et al.,* 2004, 2006). Working memory was already proposed by Francisco Aboitiz and Ricardo Garcia (1997) as a masterpiece in the primate evolution toward language, but with little concern about language (universal) preferred forms before matching for recall. We will insist here on the fact that, in our view, this STABIL-LOOP system can stabilize both word order (basic syntax and compounds) and word form structure (morphonology).

5. Summary

Beyond the fission/fusion metaphors, several of these empirical findings from ontogeny could help in building an evo-devo story of language with *caveats*:

(i) Syllables are definitely not built from segments; but segments are a late by-product of new degrees of freedom, making the carried lip and tongue articulator more and more independent from the carrier jaw (rhythm control).

(ii) Words are neither built from syllables; but chunked from the babbling flow, in the pointing frame (discrete stroke control).

(iii) Syntax does not emerge with 2-word utterances; but syntactic demonstrative (argumentative-referencing) pointing is there from the first word; and still there when 2 words appear, depending on the preceding date of emergence of the skill of pointing to the argument while predicating about a different referent from the pointed one (e.g. saying «Daddy», while pointing to his shoes... be Daddy's feet inside or not, from a strictly logical point of view).
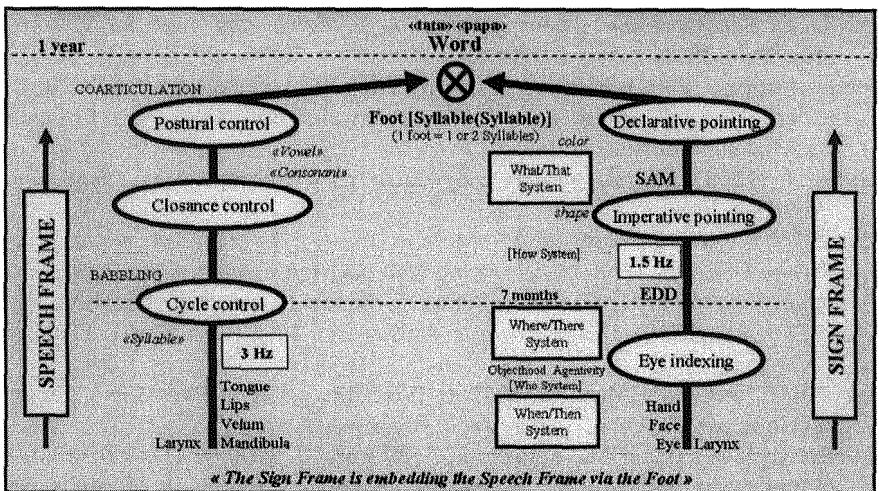

Figure 1. **A Framework for two Frames**. At about one year, the *Speech Frame* will be embedded into the *Sign Frame*: one-two... *Syllables* in a *Foot* template for the first *«Prosodic Words»*. For the *Speech Frame*, after *Canonical Babbling*, say «Syllable» rhythm emergence, two additional controls have to be mastered: *Closance* control for the «Consonant», and *Coarticulation* (*Coproduction*) for the «Vowel» *Postural* control, within the «Consonant». For the *Sign Frame*, three maturating brain streams become recruited: occipito-parietal event detection (*When*), which enters the dorsal (*Where*) and ventral (*What*) paths. Their outcomes are *Objecthood* and *Agentivity* (*Who* system), while the ventro-parietal *How* system affords *Shape Affordance*, before the objecthood *Color What* system. Among the corresponding «answers» (*Then/There/That*) to these *Wh*-systems, the most relevant stream for linguistic pointing (imperative, declarative, cooperative) is the fronto-parietal *That-Path* (Broca-SMG), together with the *Stabil-Loop*, the verbal working memory under articulatory gesture phasing control. Classically the *Sharing Attention-Intention cooperative Mechanisms* (*SAM-SIM*) develops later than *Eye Direction Detection* (*EDD*). Then, given 2-syllable first words, and once measured a mean of 3Hz for Babbling cycles, the prediction of this framework is a 2:1 Babbling/Pointing ratio. More empirically the outcome is that, knowing the distribution of the

babbling cycles of a child, one can predict the range of durations of her pointing strokes: in-between 2-3 syllables in a point, that is a universal trend for the word... point.

References

Aboitiz F., & García, R. (1997). The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Research Reviews*, 25, 381–396.

Abry, C., Sato, M., Schwartz, J-L., Lœvenbruck, H. & Cathiard, M-A. (2003). Attention-based maintenance of speech forms in memory: The case of verbal transformations. *Behavioral and Brain Sciences*, 26:6, 728-729.

Abry, C., Vilain, A. & Schwartz, J.-L. (2004). «Vocalize to Localize»? A call for better crosstalk between auditory and visual communication systems researchers: From meerkats to humans. In C. Abry, A. Vilain, & J.-L. Schwartz (Eds.) Special issue: «Vocalize to Localize». *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 5(3), 313-325.

Butterworth G. (2003). Pointing is the royal road to language for babies. In S. Kita (Eds.), *Pointing* (pp. 9-33). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Bybee J. (2003). *Phonology and Language Use*. Cambridge, UK: Cambridge University Press.

Ducey, V. (2007). *Le cadre de la parole et le cadre du signe: un rendez-vous développemental [The Speech Frame and the Sign Frame: A developmental rendez-vous]*. PhD in Language Sciences, Stendhal University (Grenoble 3).

Holger Diessel (1999). *Demonstratives*. Amsterdam: John Benjamins Publishing Co.

Iverson, J., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*, 367-371.

Lœvenbruck H., Baciu M., Segebarth C. & Abry C. (2005). The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics*, 18, 237-258.

Lœvenbruck, H., Vilain, C., Carota, F., Baciu, M., Abry, C., Lamalle, L., Pichat, C., Segebarth, C. (2007). Cerebral correlates of multimodal pointing: An fMRI Study of prosodic focus, syntactic extraction, digital- and Ocular-pointing. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (Saarbrücken, Germany, 6-10 August), 1861-1864.

Sato, M., Baciu, M., Lœvenbruck, H., Schwartz, J-L., Cathiard, M-A., Segebarth, C. & Abry, C. (2004). Multistable representation of speech forms: A functional MRI study of verbal transformations. *NeuroImage, 23*, 1143-1151.

Sato, M., Schwartz, J-L., Cathiard, M-A., Abry, C. & Lœvenbruck, H. (2006). Multistable syllables as enacted percept: A source of an asymmetric bias in

the verbal transformation effect. *Perception & Psychophysics*, *68:3*, 458-474.

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* (pp. 104-124). Oxford: Oxford University Press.