INTRODUCTION
Vocalize to Localize? A call for better crosstalk between auditory and visual communication systems researchers
From meerkats to humans

Christian Abry, Anne Vilain and Jean-Luc Schwartz
ICP-Grenoble

Vocalize to Localize: Just a speech scientist’s bias?

In the last two days of January 2003, the Vocalize-to-Localize conference was held in Grenoble, organized by the Institute of Speech Communication (ICP-Stendhal-INPG, CNRS UMR 5009) and sponsored by a European Science Foundation Project (COG-Speech: From Communication by Orofacial Gestures in Primates to Human Speech), launched in 2002 within OMLL (Origin of Man, Language, and Languages), and presented at the 5th Evolution of Language Conference (Leipzig 2004).

The aim was clearly not to suggest seamless continuity between meerkat or vervet alarm calls and full-fledged languages, or even the babbling-pointing, deictic behaviour of human infants. Who would dare tell such a “meerkat-that” story? The story of the English deictic that, demonstrative → relativizer → complementizer, has been perfectly retold by our colleague Elizabeth Traugott. It is seen as a part of a Chomskyan recursion syntactic story: the ball that hit the ball that hit the ball… and Chimpsky sees that Chompsky sees that…(see the heterogeneous joint proposal by Hauser, Chomsky, Fitch, 2000 vs Pinker & Jackendoff, in press). We know that this evolution is not at all restricted to English: e.g., jumping overseas, this story happened to ken ‘here’, a locative adverb → demonstrative → ‘relativizer’, found in the Buang language, in field work by Gillian Sankoff’s Montreal team in Papua New Guinea (pers. comm.). Elaborating on Greek deixis, deiktikos — in the tradition of Apollonios Dyskolos in Alexandria (2nd Century), and Maximos Planudes, in the late medieval Byzantium (13–14th Century) — the famous pioneer ‘localists’ of the sixties (just to mention John Lyons), who remain linguists
fond of space and scene analysis in semantic-conceptual representations, would prefer not to be hailed by meerkat calls! The same for mind philosophers, like Zenon Pylyshyn, recently reviving a long scholar tradition about indexicals — now viewed as ‘objecthood’ trackers, backed by Alan Leslie’s and others’ studies on developmental naive ontologies. And perhaps nobody would be able to tell anything more empirically fleshed out about an evolutionary link in the near future.

So we did not expect that any participant, whether or not they worked in speech science, — ethologists, developmentalists, neuropsychologists, neuroscientists, and roboticians attending the conference — could bring more to the table than anyone else in what we called in French, our Spanish albergo, a potluck, snowy winter gathering. And they brought and debated well enough, as can be read from the lines printed here and in a following issue of this new hospitable journal Interaction Studies, successor to Evolution of Communication. We extend our thanks to Harold Gouzoules and the reviewers.

Vocalizing to localize predators by conspecific calls

Emitting sounds that can be used to localize things in the world, e.g., echolocation with ultrasounds like bats, is not so widespread. The more common situation is that the emitter will be localized by others, as when distress calls of youngsters are detected by their mothers … or by predators. Emitting sounds that can be used by others to localize things in the world — apart from the emitter itself — is achieved by referential speech in all its informational complexity, which sets humans apart from the rest of the animal kingdom. In fact, among human intentional linguistic signals, including those that are the most neglected by linguists (like interjections, Ameka, 1992), some utterances like “Watch it!” or “Timber!” are not so precise — regarding its object for the former, or its directionality for the latter. But when embedded in a situation, they may nonetheless increase one’s chance to stay alive. The same is true for vervet monkeys, suricates (meerkats), and dwarf mongooses. Ground squirrels, prairie dogs and (farm) chickens also differentiate predators by calls. For chickens this is simply a convergence, as interesting as any product of morpho-functional pressure, e.g., streamlining in otherwise unrelated sharks and dolphins. But contrary to what is often said and repeated (from Cheney & Wrangham, 1987 to Tomasello, 2003a), it seems not so difficult to track this faculty throughout the primates, via Barbary macaques, Diana monkeys, ‘up to’ chimpanzees in the wild (Crockford & Boesch, 2003).

So what can comparative biology tell us about this faculty in order to better ground language evolution? Deflationists — philosophers, psychologists — have
argued against the referential, not to speak of the deictic, character of alarm calls (e.g., Tomasello, 2003a). Some of these calls simply convey — as evidenced by replicated experimental field results — information about the different spatial range of avian, mammalian, reptilian predators, and a certain proximity of the danger (emergency level). This audible augment is comparable to a wartime air raid warning — not to an all-purpose siren recruiting the firemen of my village for any type of accident — and it makes meerkats flee or scan the sky, with no more or less need to check up on conspecifics’ gazes than I would have to do in detecting the planes myself before hurrying toward a shelter. It is thus specific enough, without requiring that meerkats or monkey minds demonstrate intentionality for such localizing information about a presence.

Our meeting stance was that, without inflating alarm calls as directly relevant to human deixis and predicate faculties, one could tell something worthwhile about primate ways of pointing and human grammaticalization of deictic tools for getting words. For those who do not reject such a notion from the first call, we hope that readers of these two joint issues will find reasons to keep all these questions open, i.e. track human precursors and not insulating boundaries. Moreover, stimulated by a revival, i.e. the multi-voiced proposal that signers could be given the first place in language evolution, readers will probably enjoy a new version of this now popular debate, once they have realized what could be the core insights offered by the crosstalk between the still significant gestural medium for referencing and the now dominant verbal medium for predicating.

First gestural references with vocal/gestural predicates?

Prior to our meeting, no more than the beginning of such a story was told at the 3rd Evolution of Language conference (Paris, 2000) by Bickerton (2002, pp. 218–221), a professional linguist and pioneer in language evolution, who has long been positing a proto-language stage before language, and later an ecology-based theory of proto-language origin. (For our present argument we need not endorse his global proposal, nor his rejection of social intelligence versus foraging as a selective pressure on language emergence.)

Note that the first linguistic communications need not have been mono-modal, nor need their units have been arbitrary in the Saussurean sense. Directional gesturing with the hand, accompanied by the imitation of the noise made by a mammoth, could easily have been interpreted as meaning ‘Come this way, there's a [...] mammoth.’ [...] Although the two symbolic units (the ‘come’ gesture and the ‘mammoth’ noise) might seem disjoint — two separate, single-symbol utterances, like the one-word utterances of to 15–18-months-old humans — they could easily
have been reinterpreted (just as infant utterances at the one-word stage can often be reinterpreted predicatively, Bloom, 1973) as ‘[…] mammoth thataway’, in other words, as the first true predication. And as pointed out above [p. 210], predication — focusing on something, then making a comment about that something — is one of the most basic characteristics of human language, one that clearly distinguishes it from all other animal systems. […] with a minimum of units [hominids] could convey messages regarding the location and nature of available food supplies that would have a direct and immediate impact on the survival of those who heard and correctly interpreted those messages […] (Bickerton, 2002, pp. 219–220)

Notice that vocal or gestural predication makes no difference for Bickerton and others, whereas vocal localizing is not envisaged in the least bit, even in the case of the meerkat.

Integrating prelinguistic calls into proto-language: Vocal alarm predicates + gestural references?

Bickerton’s foraging theory ends with this transitional proposal:

Note that under [the threats of predators] (and perhaps only under these circumstances) units from pre-linguistic communication systems might have been absorbed into the proto-linguistic system; Cheney and Seyfarth (1990: 144–9) have shown that, whether or not such calls come under limbic control, their utterance is subject to voluntary modification. Assume that some ancestral species had warning calls that related to major predators, as vervet alarm calls do today. Such calls (perhaps with a different inflection), if coupled with pointing at a python-track, pawprint, bloodstain, or other indication of a possible nearby predator, could very likely have been understood as a warning that did not require immediate reaction, but rather a heightened awareness and preparedness for action. (Bickerton, 2002, pp. 220–221)

The missing link: A ‘referential and conceptual feces’?

Suricates (Manser et al., 2002), apart from general alerting calls, display one type of call specific for terrestrial predators, primarily jackals, and another specific for avian predators, including the martial and tawny eagles, and the pale chanting goshawk. In addition:

They give a third alarm call type to snakes, such as the Cape cobra, the puff adder and the mole snake. Snake alarm calls are also given to fecal, urine or hair samples of predators and/or foreign suricates [our italics]. Because snake alarm calls to all of
these stimuli cause other animals [conspecifics] to approach the caller, give alarm calls themselves, and either mob the snake or investigate the deposit, they are collectively termed recruitment alarm calls. (Manser et al., 2002, pp.55–56)

This mobbing behaviour is not unique to the suricates, converging in crows as well as in primates. What seems interesting to mention is less the generalizing scope of the call for very different stimuli, than what is empirically new. Unexpected and unpredictable was the link between a snake and various external threats, including those posed by foreign conspecifics and those from terrestrial predators, since suricates have a separate call for the latter. But even more essential is the evidence of the ‘conceptual’ similarity, among these cooperatively breeding mongooses (with sentinels and nannies), of a present snake predator and an absent terrestrial predator or foreign conspecific. This offers a first progressive, incremental, answer to the supposed gap between humans and other animals, typically deemed unable to ‘think’ beyond the hic et nunc. This is empirically remarkable if one wants to track the continuity of these calls toward any “other indication of a possible nearby predator”, long before pointing, and not just as imagined above by Bickerton (2002: 210). Traces, here olfactory ones, are the most obvious possible links between hic et nunc occurring situations, with their specificity, and animal-human neural memories, through exemplarity (see the taste of the overquoted Proust madeleine).

A Sign Language case of reference lumped with the predicate (and beyond)

Engberg-Pedersen (2003), a specialist of Danish Sign Language, begins her contribution to Pointing (Kita, 2003) with this seminal anecdote:

Once a deaf mother signed Charlotte where (‘where is Charlotte?’), Charlotte being her daughter standing right next to her. Charlotte responded by pointing energetically to herself. She did not point to the ground where she was standing as a way of answering the request for a location. Neither did she point first to herself and then to the location to indicate who was where. A point to an entity X in a location Y as a response to the question Where is X? can be seen as a condensed way of saying X is at Y; the point has the same communicative function as a simple proposition used to refer to X and predicate of X its existence at Y. But while the pointing gesture simply links two entities, X and Y, Y is predicated of X in the linguistic expression X is at Y, and in this sense Y is subordinate to X […]. When we point to entities in locations, we do exactly that: we point to the entity not the location. We focus on entities, but use space to keep track of them. The indexical aspect of a pointing gesture is its use of a location in space, but in a pointing gesture the two functions, reference and predication, are expressed by one form. (p. 269)
Philosophers are still discussing whether deixis includes location or not. Linguists know that a locative adverb can be grammaticalized as a demonstrative (see above Buang ken). What is less known is that a locative adverb can become a predicate of existence, like Gothic hiri, hirjith, hirjats (‘come!’ 2nd pers. sg, pl, dual; from hêr, compare German hier ‘here’; and see Fillmore 1966, as a pioneer of deixis in verbs). Now linguists have become clearly aware that not all languages display an overt subject-verb predication (see, among others, the illuminating example of ‘omnipredicative’ Classical Nahuatl, as exposed by Launey, 1994).

**Commands = predicates with implicit references, and ‘fossils’**

Meerkat alarm calls are not statements and they do not predicate. “Note that even commands imply a subject-predicate distinction — ‘[you] do so-and-so’ — and that in any case commands are of little use for doing what language uniquely does: transmitting (purportedly) factual information.” Bickerton (2002, p.210) reiterates here that his foraging theory fosters a survival pressure and does not need ‘socially intelligent’ commands as precursors of language, even if they do exist among apes, like the most often quoted ‘arm-raise’, an invitation–initiation gesture to chimpanzees’ rough-and-tumble play. Bickerton (2003, p.85) does not hedge: “[...] if, say, initial utterances were things like Give that to me! [...] — you wouldn’t need language to express them”; “Body language is much more reliable for most animal purposes” (p.83, and n.4).

If, for our purpose, we just take advantage of his acknowledgement that predication is implicit in some existing primate commands, one surely cannot credit Bickerton to be an inflationist on this point (unlike Bloom whom he quoted above for the child one-word stage). He simply joined language philosophers and logicians who have been interested for a long time in finding how to cope, in addition to statements, with commands, questions, etc., say speech acts.Jackendoff (2002, pp.255–256), elaborating on the proto-language/language stages, strangely puts “questions, commands and exclamations” into Bickerton’s second stage, whereas he acknowledges that they can be universally conveyed by intonation (word order, inflections, function words, depend upon further language-specific grammaticalizations). Some pages ahead (p.240), he gathered his “‘fossils’ of the one-word stage of language evolution” which contain exclamations and even the reputedly human ‘proto-command’ no. “Their semantic and pragmatic diversity suggests that they are island remnants of a larger system, superseded by true grammar.” From the beginning (pp.131–132) he has exemplified such English-specific “‘defective’ lexical items”, as oops! and goodbye!, the latter being a grammaticalization of God be with you. Hence it seems that all that can be said is that there are different
'streamlining' pressures (pace Bickerton, 2003, p. 89), not 'fossils,' but basic functions, that compact word constructions into the same integrated templates, as for interjections, exclamations, vocatives, calls... *Hail Virgin Mary* or *hail-fellow-well-met!* French *oui* (‘yes’) is such a case coming from Latin demonstratives *hoc ille* (lit. ‘this he (does)’); for *hoc* only, cf. South Gallo-Romance, so-called Occitan or Langue d’Oc). Another interesting French case — which illustrates the transfer from a command-predicate to a deictic — is the example of *voici, voilà,* (‘behold’ or ‘here is, there is’), two presentative demonstratives issued from *voi(s)* imperative, or bare-stem (‘see! look!’), with a locative adverb (locational adposition), proximal *(i)ci* (‘here’) or distal *là* (‘there’).

**Coupling with pointing, and without leaving the ‘royal road to language’ unpaved**

“[... ] although human infants vocalize and babble from soon after birth, it is gestures that for many children seem to be the first carriers of their communicative intentions. And it is gestures that seem to pave the way to early language [...]” Tomasello (2003b, p. 35) converges here on an important point with the late George Butterworth (2003), who wrote ultimately that, among gestures, pointing opens ‘the royal road’ towards language. The stories they told about the evolution of pointing do not resemble all in all the one proposed even by a functionalist linguist like Givón (1998, p. 85): going from object grabbing to emitting a ‘specific lexicalised vocal cue alone,’ say *that!* — via reaching, pointing, and adding a general, then a specific vocal cue. But once ape and human so-called ‘imperative’ pointing is superseded by the little man’s ‘declarative’ stance, both authors seem to offer no constraining device in order to derive language from non-language: hence the way is not really paved to shape words and other larger constructions. Butterworth gives general correlational evidence of pointing and language skills. Tomasello relies on *pattern-finding,* for perceptive as well as for motoric metrical templates. But why are first words one-two syllable long? And why syllables? If not a copy of motherese, this is at least an evolutionary issue. Our answer (Abry et al., forthcoming) takes the issue that, after the 7-month babbling — say, MacNeilage’s mandibular frame as the origin of proto-syllables — another frame is the 9-month (‘imperative’) pointing. Some data (including ours, but still too sparse), indicate that, while [bababa...] or [dadada...] babbling runs at a 3Hz rhythmic mode, the control of a discrete point arm-stroke is about 1.5Hz. The pointing stroke can thus chunk the babbling flow into one long or two shorter syllables. Identified with the control of the metrical unit known as the *foot* — manifest after one year for the first prosodic words — this ‘point-foot’ could ensure a crosstalk between a semiotic, symbolic
unit, the word, and phonological rhythmic units, the syllables, thus making control resources available for a template, otherwise miraculous. This is not to say that working memory has nothing to constrain into such a template/frame, but just that the precise span for one-year old children is still unknown.

**Neural ‘that-path’ and ‘stabil-loop’: Two pieces in the puzzle of language evolution**

Extrapolating backwards, from increasingly available adult neural-behavioural data, a possible developmental scenario could be the following.

1. Pointing, whether performed with the right index or the left, is a left dominance, as recently evidenced by Astafiev et al. (2003), recruiting the fronto-parietal circuits for the eye and the hand; Cantalupo and Hopkins (2001) can add that left anatomical ‘Broca’ dominance in chimps corresponds to a right-hand bias, when they vocalize and gesture.
2. This could be linguistically what we call now the left dominant ‘that-path’, fleshing out a bit more the classical and too general dorsal ‘where way’, Hurford (2004) proposed to recruit for deixis: and this is what a neuro-imaging experiment on deixis, via intonation (focus) vs. syntax (extraction), gave as a first result (Loevenbruck et al., 2002, in press), recruiting a left fronto-parietal circuit for the ‘linguistic laryngeal-oro-facial mouth’, BA47-BA40.
3. Since we finally evidenced that the stabilizing neural circuit for multi-stable verbal percepts (the asymmetric life → fly effect) was actually the phonological loop (Abry et al., 2003, Sato et al., 2004), we now view this input memory device as sensitive to motor control, favouring first long-term winning of more in-phase articulatory gestures (Sato et al., 2002). Thus, e.g., ma, mama win over am, amam, because the lowering tongue gesture for the open phase of the cycle, the vowel, can be anticipated, is in fact set, within the closing phase, the consonant gesture. And this as early as the first words, a remarkable coincidence with such a major step in speech control as coarticulation, or coproduction (Sussman et al., 1999). Therefore the two gestures can be set in-phase in ma, whereas in am the closing gesture obviously cannot be completed simultaneously, in synchrony, with the open vowel.
4. Taken together, our neuro-imaging results show that the left fronto-parietal attentional ‘that-path’ for speech deixis is part of the phonological-loop = verbal-transformation-effect stabilizing circuit: which we dubbed the ‘stabil-loop’.
5. While uniquely intonational deixis (focus: “MADELEINE did it”) activates both frontal and parietal loci, syntactic deixis (with cleft function words entrenched
from a former presentative deixis "It's Madeleine who did it") deactivates the supramarginal area. Since, meanwhile, Broca's area remains active, we interpreted this parietal deactivation as a grammaticalization step, for which the elaborated sensory parietal information corresponding to the expected feedback of one's own action goal (here feeling your voice showing, say pointing) is no longer necessary once the stored deixis packaging is being used.

6. Finally when phonology becomes fluent you just need the perisylvian cortex, without Broca (as evidenced by different groups, e.g. Wise, Ackerman, and ourselves).

Coming back to the 'that-path', which needs at least part of Broca's area, we will finally notice that it fits well with what we learned from the empirical construction approach to syntax acquisition up to 4 years, along the remarkable collaboration of Tomasello with Lieven and Diessel, in particular. The use of presentational constructions — Babybot! → That's Babybot! → That's Babybot (that is) naughty! → That's Babybot (that) sai(d) (that) you (are) naughty! — is in straight line with deixis use in syntax, in fact a perfect that-that-that story. But surely more behavioural and neural knowledge is needed on babies that point.

The ultimate lesson from meerkats: No more pure armchair stories!

Surely more behavioural and neural knowledge is needed too on animal vocalizations that localize. But aware as we are now of the bulk of tangible field work that has already been done, one should remember that armchair theorists would not have: (1) predicted that different alarm calls for different predators existed in vervets or meerkats; (2) imagined the precise acoustic shape of these calls. So why should linguists, including speech scientists and any other people, have the slightest chance to correctly posit the first word or 'language fossil', in-between animal and child data which, together with (paleo-) genetics, will probably be the main growing empirical fields for our phylogenetic quest?

The division of the conference contributions in two issues is of course just a matter of convenience. The first issue contains meerkats, monkeys, apes and humans, including Neandertal, ending with a dismissal of any acoustic charge against this related vocal-tract. The second issue deals more specifically with babies, babybot systems, ending with a 'first-in/last-out' nascent-remnant view of aphasia. It opens by offering a thread for crosstalk, with a proposal of 'interweaving proto-sign and proto-speech' (not to speak of our present introductory motto: put a foot into the arm!). And it is particularly welcome as a link with the first issue, articulating a hot debate… which rebounds as soon as the second issue opens. To be continued!
References


Loevenbruck, H., Baciu, M., Segebarth, C. & C. Abry (2002). Prosodic deixis (focus) and syntactic deixis (extraction) in LIFG and LSMG. *8th International Conference on Functional Mapping of the Human Brain*, Sendai, Japan.


About the authors

**Christian Abry**, Docteur d’État of the University Stendhal, Grenoble (France), Professor of Experimental Phonetics in 1998. Since Head of the Department of Linguistics (1998–2008). In the last 20 years he has led at the Institute of Speech Communication (ICP-INPG-Stendhal CNRS UMR 5009) three groups: Articulatory Modelling, Linguistic Anthropology of Speech, and founded *Speech, Multimodality and Development*. Principal investigator in several national and international research projects (1993–1995: Prime, with P. Badin, of ESPRIT-BR Speech Maps, coordinating 14 European Labs in speech inversion and robotics). Main interests in speech production, control and robotics, perceptuomotor interactions, speech working memory, bimodal speech, speech development, speech evolution, and narratives. Now at CRI (Center for Research on the Imaginary), Stendhal.

**Jean-Luc Schwartz**, a member of the French CNRS, lead the Speech Perception Group at ICP (Institut de la Communication Parlée) from 1987 to 2002, and has now been leading the laboratory since 2003. His main research areas involve auditory modelling, speech perception, bimodal integration, perceptuomotor interactions, speech robotics and the emergence of language. He has been involved in various national and European projects, and has authored or co-authored more than 30 publications in international journals such as IEEE SAP, JASA, *Journal of Phonetics, Computer Speech and Language, Artificial Intelligence Review, Speech Communication, Behavioural and Brain Sciences, Hearing Research, Cognition, NeuroImage*, about 20 book chapters, and 100 presentations in national and international Workshops. He is the co-editor of a book on speech communication, of a special issue of the *Speech Communication* and the *Primatologie*. 

Anne Vilain completed a PhD in Phonetics in 2000, and has been Maître de Conférence in Experimental Phonetics of the University of Grenoble since 2001. Her research is lead within the Linguistic Anthropology of Speech group at the Institut de la Communication Parlée (ICP), and her main research interests are speech motor control, the emergence of language, and the ontogeny of speech production. Now at GIPSA-LAB, INPG - Stendhal.