Friulian Journal of Science **5**. 2004, 103-111

Speculations on the origin of language

GIACOMO RIZZOLATTI* & LAILA CRAIGHERO#

Abstract. There are two main views on the origin of human language. The first is that it derives from animal's call, the other from gestures. In the present article we argue that the second view is much more plausible. We discuss first a neurophysiological mechanism, the mirror neuron mechanism, which solves the problem of direct comprehension of action meaning. We discuss then how sounds have been added to gestures and eventually became the main medium of human communication.

Peraulis clâf. Mirror neurons, language evolution, gestual comunication, action, semantic.

Introduction. The faculty of language is a cognitive ability that only humans posses. How language appeared? This is a completely open question. Yet, the discovery of a new class of neurons in the monkey, the so-called *mirror neurons*, indicates a mechanism that may give some clues on the origin of speech and its continuity with non-human primate behavior. This mechanism, in fact, is of great evolutionary importance since it is supposed to be at the basis of the way in which primates understand actions made by their conspecifics. Moreover, evidence coming from different experimental approaches have demonstrated that a mirror-neuron system is present also in humans. The most intriguing finding deriving from brain imaging studies is that one of the regions mainly involved in this system is the left inferior frontal cortex in correspondence of the Broca's region which is classically considered a language-related brain region. In the present article we will briefly describe the basic properties of mirror neurons in non-human primates and man and we will propose a theory of

---

* Department of Neuroscience, University of Parma. E-mail: giacomo.rizzolatti@unipr.it

Department S.B.T.A., Section of Human Physiology, University of Ferrara, Italy.

language evolution based on the mirror neuron properties.

Mirror neurons in monkeys. Mirror neurons are a set of neurons originally discovered in the ventral premotor cortex of the macaque monkey. Their defining property is that they became active both when the monkey *does* a particular action (like grasping an object) and when it *observes* another individual making a similar action (see Rizzolatti et al. 2001). In order to be triggered by visual stimuli, mirror neurons require an interaction between a biological effector (hand or mouth) and an object. The sights of an object alone, of an agent mimicking an action, or of an individual making intransitive (non-object directed) gestures are all ineffective. The object significance for the monkey has no obvious influence on mirror neuron response. Grasping a piece of food or a geometric solid produces responses of the same intensity. Mirror neurons show a large degree of generalization. Largely different visual stimuli, but representing the same action, are equally effective. For example, the same grasping mirror neuron that responds to a human hand grasping an object, responds also when the grasping hand is that of a monkey. Similarly, the response is, typically, not affected if the action is done near or far from the monkey, in spite of the fact that the size of the observed hand is obviously different in the two conditions. It is also of little importance for neuron activation if the observed action is eventually rewarded. The discharge is of the same intensity if the experimenter grasps the food and gives it to the recorded monkey or to another monkey, introduced in the experimental room.

An important functional aspect of mirror neurons is the relation between their visual and motor properties. Virtually, all mirror neurons show congruence between the visual actions they respond to and the motor responses they code. According to the type of congruence they exhibit, mirror neurons have been subdivided into “strictly congruent” and “broadly congruent” neurons (Gallese et al. 1996). Mirror neurons in which the effective observed and effective executed actions correspond in terms of goal (e.g. grasping) and means for reaching the goal (e.g. precision grip) have been classed as “strictly congruent”. They represent about one third of F5 mirror neurons. Mirror neurons that, in order to be triggered, do not require the observation of exactly the same action that they code motorically, have been classed as “broadly congruent”. They represent about two-third of F5 mirror neurons.

Mirror-neuron system in humans. There are no studies in which single neurons were recorded from the putative mirror-neuron areas in humans. Thus, a direct evidence of the existence of mirror neurons in humans is lacking. There is, however, a rich amount of data proving, indirectly, that a mirror-neuron system does exist also in humans. Evidence in this sense comes from neurophysiological and brain-imaging experiments (Gastaut & Bert 1954; Cochin et al. 1998, 1999; Hari et al. 1998). More direct evidence that the motor system in humans has mirror properties was provided by transcranial magnetic stimulation (TMS) studies. Fadiga et al. (1995) recorded motor evoked potentials (MEPs), elicited by stimulation of the left motor cortex, from the right hand and arm muscles in volunteers required to observe an experimenter grasping objects (transitive hand actions) or performing meaningless arm gestures (intransitive arm movements). Detection of the dimming of a small spot of light and presentation of 3-D objects were used as control conditions. The results showed that the observation of both transitive and intransitive actions determined an increase of the recorded MEPs with respect to the control conditions. The increase concerned selectively those muscles that the participants use for producing the observed movements.

The MEPs facilitation during movement observation may result from a facilitation of the primary motor cortex due to mirror activity of the premotor areas, to a direct facilitatory input to the spinal cord originating from the same areas, or from both. Support for the cortical hypothesis (see also below, brain imaging experiments) came from a study by Strafella & Paus 2000. By using a double-pulse TMS technique, they demonstrated that the duration of intracortical recurrent inhibition, occurring during action observation, closely corresponds to that occurring during action execution.

A large number of brain imaging studies showed that the observation of actions done by others activates in humans a complex network formed by occipital, temporal and parietal visual areas, and two cortical regions whose function is fundamentally or predominantly motor (Rizzolatti et al. 1996; Grafton et al. 1996; Grèzes et al. 2003; Iacoboni et al. 1999, 2001; Nishitani & Hari 2000, 2002; Buccino et al. 2001; Decety et al. 1997; Koski et al. 2002, 2003; Manthey et al. 2003). These two last regions are the rostral part of the inferior parietal lobule and the lower part of the precentral gyrus plus the posterior part of the inferior frontal gyrus (IFG). They form the core of the mirror-neuron system in humans.

It is important to stress that Broca's area is located in correspondence of the posterior part of IFG. Thus, its activation during the observation of hand actions, suggests that this region is the human homologue of area F5. In this direction point also some comparative cytoarchitectonical data by Petrides & Pandya (1997) suggesting a morphological parallel between monkey premotor area F5 with area 44 as well as fMRI data (Binkofsky et al. 1999) demonstrating that Broca's region become active during manipulation of complex objects.

A comparison between the mirror neuron system of humans and monkey shows some similarities but also some differences (Rizzolatti & Craighero 2004). Firstly, unlike monkeys where the presence of an object is necessary to activate the mirror neurons, human mirror neuron system is activated by the observation of mimed actions (Buccino et al. 2001; Grèzes et al. 2003). Secondly, it responds to intransitive, meaningless gestures (e.g., Fadiga et al. 1995; Iacoboni et al. 1999). Thirdly, the motor facilitation, determined by action observation, follows the time-course of the observed action (Gangitano et al. 2001).

These data indicate that the human mirror neuron system is able to code, besides the goal of an action, the movements necessary to achieve it. This is a prerequisite necessary for imitation learning.

Gestures versus sounds. Mirror neurons represent the neural basis of a mechanism that creates a direct link between the sender of a message and its receiver. This mechanism, by transforming an action done by an individual into a representation of the same action in the motor cortex of the observer, creates, a direct, non-arbitrary link between two communicating individuals. Actions done by other individuals become messages that are understood by an observer without any cognitive mediation. On the basis of this property, Rizzolatti and Arbib (1998) proposed that the mirror-neuron system represents the neurophysiological mechanism from which language evolved. The theory of Rizzolatti and Arbib belongs to theories that postulate that speech evolved mostly from gestural communication and that sound was a late communicative acquisition developed on the top of it, with no link with ancient calls (see Armstrong et al. 1995; Corballis 2002). Its novelty consists in the fact that it indicates a neurophysiological mechanism that creates a common (parity requirement), non-arbitrary, semantic link between communicating individuals.

Mirror-neuron system *in monkeys* is constituted of neurons coding object-directed actions. A first problem for the mirror-neuron theory of language evolution is to explain how this close, object-related system became an open system able to describe actions and objects without directly referring to them. It is likely that the great leap from a closed system to a communicative mirror system depended upon the evolution of imitation (see Arbib 2002) and the related changes of the human mirror-neuron system: the capacity to respond to pantomimes (Buccino et al. 2001; Grèzes et al. 2003) and to intransitive actions (Fadiga et al. 1995; Maeda et al. 2002) that was absent in monkeys.

The notion that communicative actions derived from object-directed actions is not new. Vygotsky (1934), for example, explained the evolution of pointing movements as due to attempts of children to grasp far objects. It is interesting to note that, although monkey mirror neurons do not discharge when the monkey observes an action that is not object-directed, they do respond when an object is hidden, but the monkey knows that the action has a purpose (Kohler et al. 2002). This indicates that breaking spatial relation between effector and target does not impair the capacity of understanding the action meaning. The pre-condition for understanding pointing – the capacity to mentally represent the action goal- is already present in monkeys.

An alternative view, based on the fact that humans mostly communicate by sounds, is that language evolved form the sound-based animal communication. In fact, human speech and animals' calls are widely different phenomena. First of all, the anatomical structures underlying speech and primates' calls are different. Primates' calls are mediated primarily by the cingulate cortex and by subcortical structures (see Jürgens 2002). In contrast, human speech network is located on the lateral cortical surface. Second, speech is not necessarily linked to emotions, whilst animals' calls are. Third, speech is endowed with combinatorial properties that are absent in animal communication. The anatomical heterogeneity of speech and animals' calls represents an enormous difficulty for theories that claim that speech derived from animals' calls. How may it have occurred that speech centers "jumped", in evolution, from one side of the brain to another? There is, however, a further, fundamental difficulty for sound-based theories of language evolution: the relation between sound and meaning. Sounds have no meaning. How is it possible, therefore, that arbitrary sounds acquired it? The defenders of the evolution of speech from animal calls postulate that animals' calls acquired progressively referential meaning. While originally they were related to an emotion in general (e.g., fear), subsequently they started to describe also what caused that emotion (e.g., a snake or an eagle). Against this view is the fact that fundamental property of human language is the possibility to use the same words in completely different emotional contexts. Fire does not convey only the message "escape ", but, according to the context, it may convey a positive message (e.g. "come here, the fire for cooking is ready"). A language based on emotion cannot "accommodate", in principle, opposite emotional meanings.

From protosigns to vocal communication. Once, in evolution, the mirror neuron system in humans acquired the possibility to code pantomimes and intransitive actions, a vocabulary of "protosigns" (immediately understandable for their intrinsic properties) developed. Signs describing directions and objects, and pantomimes of actions formed it. Protosigns were often accompanied by sounds. This fact prevented the occurrence of a full-fledged sign language, as those used by deaf people, and was at the basis of speech development.

The protosign communication was progressively substituted by speech. An interesting attempt to explain how this occurred has been advanced by Paget (1930). According to him, when the individual gesticulated, "his tongue, lips and jaw unconsciously followed suit". There was, therefore, congruence between hand action and the sounds accompanying the mouth actions. This congruence gave meaning to sounds. As far as vowels are concerned, he suggests that, in all languages, "A" (as in large) refers to anything that is large, wide open, while "I" (as in mini) to something that is small or pointed. Consonants also convey gestural symbolism. "M", for example, implies a continued closure; "DR" denotes running or walking. According to this theory, the great majority of words are pantomimic. They are built "much as the Chinese ideographs are, by addition of separately significant elements" (Paget 1930). This type of organization explains the difficulty to discover the original sound meaning in the words. Furthermore, because almost every action or idea can be pantomimed in many different ways and every gesture can be construed in many different ways, language largely differ one from another.

Although largely speculative, Paget theory provides a possible clue on how intrinsically known messages (hand gestures) were transferred to an opaque gestural system, as the oro-laryngeal system, without loosing their intrinsic (non-arbitrary) meaning. Furthermore, it allows a clear neurophysiological prediction: hand/arm and speech gestures must have a common neural substrate.

Recent studies demonstrate that this prediction is true (see for review Rizzolatti & Craighero 2004). Among them, particularly interesting are some experiments by Gentilucci et al. (2001). Participants were presented with two 3-D objects, one large, and the other small. On the visible face of the objects either two letters or a series of dots were shown. Participants were instructed to grasp the objects, and, in those trials in which the letters appeared on the object, to open their mouth. The kinematics of hand, arm, and mouth movements was recorded. The results showed that, although participants were instructed to keep the mouth aperture constant, lip aperture and the peak velocity of lip aperture increased when the movement was directed to the large object. In a further experiment, participants were asked to pronounce a syllable (e.g., GA). It was found that lip aperture was larger when the participants grasped the larger object. Furthermore, the maximal power of the voice spectrum recorded during syllable emission was also higher when the larger object was grasped.

These experiments show that mouth movements and the oro-laryngeal synergies necessary for syllable emission are indeed linked to specific manual gestures. Some sound are linked to (and describes) large objects, while others are linked to small objects.

Let us now come back to Paget theory. When we eat, we move our mouth and tongue in a specific manner. This combined series of motor acts constitutes a gesture whose meaning for the observer is transparent. If, while making these movements, we blow air through the oro-laryngeal cavities, we produce a sound like "*mnyam-mnyam*", a word whose meaning is almost universally recognized (Paget 1930). Thus, the meaning of an action, "naturally" understood, is transferred to a sound. Accepting this view, it appears logical to postulate that the next step toward speech acquisition consisted in the generation of sounds of actions (e.g.

"*mnyam-mnyam*") *without* actually performing those actions. In analogy to the mirror neurons for gestures, this evolutionary step must have been accompanied by the appearance of neurons that controlled oro-pharyngeal gestures for sound emission *and* resonated in response to the same sounds.

This prediction was recently confirmed. Fadiga et al. (2002) recorded motor evoked potentials from the tongue muscles in volunteers instructed to listen to acoustically presented words and non-verbal stimuli. In the middle of words either a double "f" or a double "r" (requiring tongue movements) were embedded. During stimulus presentation, the participants' left motor cortex was magnetically stimulated. The results showed that the mere listening to words containing the double "r" determined a significant increase of the amplitude of the potentials recorded from tongue muscles with respect to listening to bitonal sounds and words containing the double "f".

What is the functional role of this mechanism? It may represent the neural basis of imitation of verbal sounds; mediate their perception (Liberman & Mattingly 1985); or both. It may, however, also have a role in semantics.

As discussed above, the primate communication system was originally based on gestures. The appearance of neurons active during production of verbal material *and* during perception of the same verbal material changed the scenario. Because these new neurons (e.g., those activated by "*mnyam-mnyam*" sound), were connected (given their origin) with the mirror neurons coding the corresponding actions (e.g. to eat), when they discharged the perceiving individual recognized not only the sound of the word but also its meaning.

In other words, at a certain stage of evolution the activity of these "echo mirror neurons" became sufficient to represent the semantic content of mirror neurons linked to them by itself. Their activation *represented*, albeit indirectly, an action (*second order action representation*). This second order representation had clear advantage on the first order representation. For example, it could create new associations between words based on the probability of word occurrence rather than on the occurrence of actions and gave, in this way, greater communicative possibilities to individual possessing it.

This sketch of speech evolution does not take into consideration the problem of grammar evolution. Yet, a better understanding of the origin of semantics may, in the future, throw light also on this fundamental problem.

References/ Bibliografie

Arbib M.A. (2002). Beyond the mirror system: imitation and evolution of language. In Nehaniv C. and Dautenhan K. (Eds) *Imitation in animals and artifacts*. Cambridge MA: The MIT Press, pp. 229-280.

Armstrong A.C., Stokoe W.C., Wilcox S.E. (1995). *Gesture and the nature of language.* Cambridge, UK: Cambridge University Press.

Binkofski F., Buccino G., Posse S., Seitz R.J., Rizzolatti G., Freund H. (1999). A fronto-parietal circuit for object manipulation in man: evidence from an fMRI-study. *Eur. J. Neurosci.,* 11: 3276-86.

Buccino G., Binkofski F., Fink G.R., Fadiga L., Fogassi L. et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.,* 13: 400-4.

Cochin S., Barthelemy C., Lejeune B., Roux S., Martineau J. (1998). Perception of motion and qEEG activity in human adults. *Electroencephalogr. Clin. Neurophysiol.,* 107: 287-95.

Cochin S., Barthelemy C., Roux S., Martineau J. (1999). Observation and execution of movement: similarities demonstrated by quantified electroencephalograpy. *Eur. J. Neurosci.,* 11: 1839-42.

Corballis M.C. (2002). *From hand to mouth. The origins of language*. Princeton: Princeton University Press.

Decety J., Grezes J., Costes N., Perani D., Jeannerod M., Procyk E., Grassi F., Fazio F. (1997) Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain,* 120: 1763-1777.

Fadiga L., Craighero L., Buccino G., Rizzolatti G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.,* 15: 399-402.

Fadiga L., Fogassi L., Pavesi G., Rizzolatti G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *J. Neurophysiol.,* 73: 2608-11.

Gallese V., Fadiga L., Fogassi L., Rizzolatti G. (1996). Action recognition in the premotor cortex. *Brain,* 119: 593-609.

Gangitano M., Mottaghy F.M., Pascual-Leone A. (2001). Phase specific modulation of cortical motor output during movement observation. *NeuroReport* 12: 1489-92.

Gastaut H.J., Bert J. (1954). EEG changes during cinematographic presentation. *Electroencephalogr. Clin. Neurophysiol.,* 6: 433-44.

Gentilucci M., Benuzzi F., Gangitano M., Grimaldi S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *J. Neurophysiol.,* 86: 1685-99.

Grafton S.T., Arbib M.A., Fadiga L., Rizzolatti G. (1996). Localization of grasp representations in humans by PET: 2. Observation compared with imagination. *Exp. Brain Res.,* 112: 103-11.

Grèzes J., Armony J.L., Rowe J., Passingham R.E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage,* 18: 928-37.

Hari R., Forss N., Avikainen S., Kirveskari S., Salenius S., Rizzolatti G. (1998). Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proc. Natl. Acad. Sci. USA,* 95: 15061-5.

Iacoboni M., Koski L.M., Brass M., Bekkering H., Woods R.P. et al. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proc. Natl. Acad. Sci. USA* 98: 13995-9.

Iacoboni M., Woods R.P., Brass M., Bekkering H., Mazziotta J.C., Rizzolatti G. (1999). Cortical mechanisms of human imitation. *Science,* 286: 2526-8.

Jürgens U. (2002). Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.,* 26: 235-258.

Kohler E., Keysers C., Umiltà M.A., Fogassi L., Gallese V., Rizzolatti G. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science,* 297: 846-48.

Koski L., Iacoboni M., Dubeau M.C., Woods R.P., Mazziotta J.C. (2003). Modulation of cortical activity during different imitative behaviors. *J. Neurophysiol.,* 89: 460-71.

Koski L., Wohlschlager A., Bekkering H., Woods R.P., Dubeau M.C. (2002). Modulation of motor and premotor activity during imitation of target-directed actions. *Cereb. Cortex,* 12: 847-55.

Liberman A.M., Mattingly I.G. (1985). The motor theory of speech perception revised. *Cognition,* 21: 1-36.

Maeda F., Kleiner-Fisman G., Pascual-Leone A. (2002). Motor facilitation while observing hand actions: specificity of the effect and role of observer's orientation. *J. Neurophysiol.*, 87: 1329-5.

Manthey S., Schubotz R.I., Von Cramon D.Y. (2003). Premotor cortex in observing erroneous action: an fMRI study. *Brain Res. Cogn. Brain Res.,* 15: 296-307.

Nishitani N., Hari R. (2000). Temporal dynamics of cortical representation for action. *Proc. Natl. Acad. Sci. USA,* 97: 913-8.

Nishitani N., Hari R. (2002). Viewing lip forms: cortical dynamics. *Neuron,* 36: 1211-20.

Paget R. (1930). *Human Speech*. London: Kegan Paul, Trench and Co.

Petrides M., Pandya D.N. (1997). Comparative architectonic analysis of the human and the macaque frontal cortex. In Boller F., Grafman J. (Eds) *Handbook of Neuropsychology* Vol. IX. New York: Elsevier, pp. 17-58.

Rizzolatti G., Arbib M.A. (1998). Language within our grasp. *Trends Neurosci.*, 21: 188-94.

Rizzolatti G., Craighero L. (2004). The mirror neuron system. *Ann. Rev. Neurosci.* (in press).

Rizzolatti G., Fadiga L., Matelli M., Bettinardi V., Paulesu E. et al. (1996). Localization of grasp representation in humans by PET: 1. Observation versus execution. *Exp. Brain Res.,* 111: 246-52.

Rizzolatti G., Fogassi L., Gallese V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.,* 2: 661-70.

Strafella A.P., Paus T. (2000). Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *NeuroReport,* 11: 2289-92.

Vygotsky L.S. (1934). *Thought and language*. Cambridge, MA: MIT Press.