A Contribution to the Special Issue on "Integrative Models of Broca's Area and the Ventral Premotor Cortex"
of *Cortex: A Journal Devoted to the Study of the Nervous System & Behavior* (Revised September, 2004)

Arbib, M.A., 2005a, A sentence is to speech as what is to action? *Cortex,* in press.

A sentence is to speech as what is to action?[1]

Michael A. Arbib
Computer Science, Neuroscience and USC Brain Project
University of Southern California
Los Angeles, CA 90089-2520
arbib@pollux.usc.edu

This article offers a conceptual framework for integrated analysis of subprocesses in action and language, based on goal-directed action. Anatomical substrates are discussed in the companion paper (Arbib & Bota, 2003) which approaches "Integrative Models of Broca's Area and the Ventral Premotor Cortex" within the context of explaining why the evolution of the human brain yielded mechanisms which support language in a multi-modal vocal-manual-facial system rather than privileging the vocal mode. Arbib & Bota examine homologies between different cortical areas in macaque and human to revisit the Mirror System Hypothesis (MSH) of Rizzolatti and Arbib (1998) – the notion that the mirror system for grasping (which has its frontal outpost in premotor area F5 of the macaque) provides the substrate for the evolution of the language-ready brain which supports parity of communication. They also offer a critique and extension based on the work of Aboitiz and García (1997; Aboitiz et al., 2004). Arbib & Bota also discussed the utility of neuroinformatics in relating information across diverse cortical atlases and evaluating degrees of homology for brain regions of interest in different species (for discussion, see Deacon, 2004, and Arbib & Bota, 2004).

1. From Communicative Goal to Sentential Form

Consider a conditional, hierarchical motor plan for opening a child-proof aspirin bottle:

> While holding the bottle with the non-dominant hand, grasp the cap, push down and turn the cap with the dominant hand; then repeat (release cap, pull up cap and turn) until the cap comes loose; then remove the cap.

This hierarchical structure unpacks to different sequences of action on different occasions, with subsequences conditioned on the achievement of goals and subgoals. To ground the search for similarities between action and language, I suggest we view an action such as this one as a "sentence" made up of "words" which are basic actions. A "paragraph" or a "discourse" might then correspond to, e.g., an assembly task which involves a number of such "sentences".

[1] This article is based in part on comments presented at the September 2003 Workshop in Leipzig and an ensuing email debate on the issues with David Caplan. The debate has refined my original argument but not, alas, sufficiently to convince David. I also thank the anonymous referees whose thoughtful critiques of the previous draft enabled me to develop my ideas in a way which, hopefully, will encourage others to probe them more deeply.

Now consider a sentence like "Serve the handsome old man on the left.", spoken by a restaurant manager to a waiter. From a "conventional" linguistic viewpoint, we would apply syntactic rules to parse this specific string of words. But let us look at the sentence not as structure to be parsed but rather as the result of the manager's attempt to achieve a *communicative goal*: to get the waiter to serve the intended customer. His *sentence planning strategy* repeats the "loop"

<add adjective or prepositional phrase>

until (he thinks) ambiguity is resolved:

(1) Serve the old man.

Still ambiguous?

(2) Serve the old man on the left.

Still ambiguous?

(3) Serve the handsome old man on the left.

Still ambiguous? Apparently not. So the manager "executes the plan" and says "Serve the handsome old man on the left." to the waiter.

Here, a noun phrase NP may be expanded by adding a prepositional phrase PP after it [as in expanding (1) to (2)] or an adjective Adj before it [as in expanding (2) to (3)]. The suggestion is that syntactic rules of English which I approximate by NP → NP PP and NP → Adj NP are abstracted from procedures which serve to reduce ambiguity in reaching a communicative goal. This example concentrates on a noun phrase – and thus exemplifies ways in which reaching a pragmatic goal (identifying the right person, or more generally, object) may yield an unfolding of word structures in a way that may clarify the history of syntactic structures.

While syntactic constructions can be usefully analyzed and categorized from an abstract viewpoint, the pragmatics of what one is trying to say and to whom one is trying to say it will drive the goal-directed process of producing a sentence; while the hearer has the inferential task of unfolding multiple meanings from the word stream (with selective attention) and deciding (perhaps unconsciously) which ones to meld into his cognitive state/narrative memory.

2. The Saussurean Sign

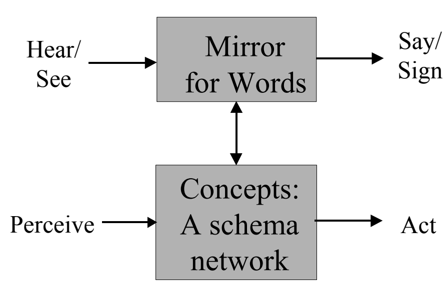

Figure 1: The bidirectional sign relation links (neural codes for) words and concepts (Arbib 2004b, inspired by Hurford 2004).

Hurford (2004) makes the crucial point that we must (in the spirit of Saussure) distinguish the "sign" from the "signified". In Figure 1, we distinguish the "neural representation of the sign" (top row) from the "neural representation of the signified" (bottom row). The top row of the figure makes explicit the result of the progression (Arbib 2002, 2005) of mirror systems for:[2]

1) grasping and manual praxic actions.
2) imitation of grasping and manual praxic actions.
3) pantomime of grasping and manual praxic actions.
4) pantomime of actions outside the panto-mimic's own behavioral repertoire (e.g., flapping the arms to mime a flying bird).
5) conventional gestures used to formalize and disambiguate pantomime (e.g., to distinguish "bird" from "flying")
6) conventionalized manual, facial and vocal communicative gestures ("words") separate from pantomime.

Arbib (1981) distinguished *perceptual schemas* which determine whether a given "domain of interaction" is present in the environment and provide parameters concerning the current relationship of the organism with that domain, and *motor schemas* which provide the control systems which can be coordinated to effect a wide variety of actions. Recognizing an object (an apple, say) may be linked to many different courses of action (to place the apple in one's shopping basket; to eat the apple; to discard a rotten apple, etc.). In this list, some items are apple-specific whereas other invoke generic schemas for reaching and grasping. Only for some actions (run, walk, hit, …) or expressions of emotion will perceptual and motor schemas be integrated into a "mirror schema". Viewing Figure 1 as an anatomical diagram, patterns of activity encoding "words" in the top row mirror system become associatively linked with various activity patterns in the various parts of the brain serving action and perception. I do not see a "concept" as corresponding to one word, but rather to a graded set of activations of the schema network.

In developing the Mirror System Hypothesis for the evolution of brain mechanisms underlying language, I make the point (Arbib, 2005) that a key transition in "brain power" was the ability – *complex action analysis* – to recognize that a novel action is composed of (approximations to) known actions. This is crucial not only for "complex imitation" and the child's ability to acquire language and social skills, but also for adult use of language – we recognize a novel utterance as in fact composed of known actions (the speaking or signing of words) and this stock of words is open-ended.

3. What corresponds to "Reach and Grasp"?

Is a "reach and grasp" is more like a word or a phoneme. My answer is, paradoxically, both – because signed language takes a very different approach from speech.

Signing exploits the fact that the signer has a rich repertoire of arm, hand and face movements, and thus builds up vocabulary by variations on this multi-dimensional theme (move a handshape [or two] along a trajectory to a

[2] Barrett et al. (2005) usefully review evidence for differential localization of neuronal systems controlling limb praxis, speech and language, and emotional communication, and assert that such data justify rejection of an evolutionary relationship between the underlying mechanisms. However, studies of mammalian brain evolution suggest that increasing complexity of behavior is paralleled by increases in the number of functional subdivisions of neocortex and that reduplication of circuitry may form the basis for differential evolution of copies of a given system, with differing connectivities, etc., to serve a variety of functions.

particular position while making appropriate facial gestures). Being omnivores, humans also have a great deal of "oral dexterity", but (a) manual control is under visual feedback, providing a wide range of hand movements available for the inspection of others even in the absence of intended communication, whereas (b) the ingestive use of "oral articulators" has few auditory correlates, and so their use for vocalization requires a dramatically different mode of control.[3] Thus, speech employs a system of articulators for which there is no rich behavioral repertoire of sound producing movements to build upon. Instead evolution "went particulate", so that the word is built (to a first approximation) from a language-specific stock of phonemes, actions defined by the coordinated movement of one or more articulators, but with only the goal of "sounding different from other phonemes" rather than conveying meaning in themselves.[4]

An "action word" (i.e., the answer to "A word is to speech as what is to action?") is a basic action in the sense of a movement that achieves some goal. A "language word" is defined by language internal criteria – basically as the smallest element of freestanding meaning. It is then a contingent fact that a single "vocal action" does not contain the flexibility required to support a large vocabulary, and so the phonemes and syllables become meaningless counters. But note too that in sign even apparently pantomimic gestures get their meaning within a language specific system of distinctions rather than necessarily inheriting the praxic significance of the motion.[5] Some may find it disturbing that the basic "action word" in speech corresponds to a unit smaller than the word. However, the "level" of the word is highly fluid even internal to language. As we translate from language to language, we preserve a general sense of "word", yet a word with case markings in one language may translate into a prepositional phrase in another, and a verb in some Amerindian languages may unpack into a whole sentence. I thus think it quite reasonable that "action words" in the vocal and manual action may map to different levels in language analysis.

4. From Phonology to Grammar

Research in motor control tends to be at the level of *phonology* – how do effectors produce a basic action, what "co-articulation" may modify one action on the basis of another? – than at the level of *syntax and semantics* which analyzes the structure of sentences, and related issues. MSH claims that mechanisms supporting language evolved atop the mirror system for grasping, not that circuits for praxis and communication are identical. At least two different changes in the brain were required to support the emergence of language: (i) the availability of complex action analysis and complex imitation at "both ends of the arrow" of Figure 1, abilities in which the evolution of the mirror system play a role; and (ii) the complementary ability for situational analysis which links sentence production and perception to assemblages of conceptual schemas.

---

[3] Recent neurophysiology helps us expand our view of the MSH to build on the orofacial as well as the manual mirror system of macaques: see Kohler et al. (2002) for manual mirror neurons responsive to sounds, Ferrari et al. (2003) for orofacial motor neurons responsive to observation of oro-facial communicative gestures, and Fogassi & Ferrari (2004), Arbib (2004a) and MacNeilage & Davis (2004a) for discussion.

[4] Studdert-Kennedy (2002) relates mirror neurons and vocal imitation to the evolution of particulate speech.

[5] Corina et al. (1992) described patient WL with damage to left hemisphere perisylvian regions. WL exhibited poor sign language comprehension and production. Nonetheless, he could produce stretches of pantomime and tended to substitute pantomimes for signs, even when the pantomime required more complex movement.

The hearer's processes for understanding (more or less) what the speaker intends, and the speaker's processes for conveying the intended message with (more or less) reduced ambiguity must, to succeed, be *approximately* inverse to each other. I distinguish *production grammar* – getting from a communicative goal to words that express it – from *perception grammar* – getting from a sequence of words to the goal behind it. Syntax in the normal sense is a compact answer to the question: "In this community, what regularities seem to be shared by the sentences that are produced and understood?" In this way, the linguist may define a *single* grammar to represent regularities common to both perception and production of utterances – but I deny that there is a single grammar represented in the brain that is consulted by separate processes of perception and production.

The neuroscience of action must go far beyond the neurophysiology of controlling muscles. Successful planning or interpretation in general requires the interweaving of many motor activities to achieve an overall goal through a hierarchy of subgoals. "Action" at this relevant level integrates perceptual, conceptual and motor activities. Miller, Galanter, and Pribram (1960) related motor sequences to plans defined as hierarchical structures of elementary "TOTE (Test-Operate-Test-Exit) units" to make explicit the complex contingencies within a plan. Arbib (1981) offered the assemblage of available schemas to form new schemas as *coordinated control programs* which could control the interweaving and overlapping of perceptual and motor acts.

Linguists view the sentence as a basic locus of well-formedness and focus syntax on the characterization of sentence structure and the constituents that yield it. However, sentences only approximate the way words are gathered in real speech (Iacoboni, 2004) as different thoughts vie for expression, or partial expressions seem to do the job of a whole sentence. One may employ run-on phrases which cumulatively convey some message, yet do not cohere into a well-formed sentence. Thus in some sense a sentence is a theoretical construct sanctioned by a specific choice of grammar for a language. Written language reflects the lessons of grammar more than does spoken language. I say this not to disparage the utility of the notions of sentence and grammar but rather to note they describe approximations to oral language use rather than its full complexity.

It is thus somewhat arbitrary how one inserts the punctuation to turn a speech stream into a sequence of sentences. It is similarly arbitrary to break the "action stream" into a sequence of actions. In our bottle cap example, one must recognize that the correct analysis would not include, say, "turn the cap three times" but "turn the cap repeatedly till it comes loose." The ability to perform and recognize skills rests on the ability to extract "constituents" by linking action patterns to subgoals.

I would argue that the full syntax of any modern language is the result of "bricolage" – a historical accumulation of piecemeal strategies for achieving a wide range of communicative goals complemented by a process of generalization whereby some strategies become unified into a single general strategy. In correspondence related to Arbib (2005), David Kemmerer notes that the formal criteria used by linguists to identify grammatical categories in one language may be absent or employed very differently in another: verbs are often marked for tense, aspect, mood, and transitivity but some languages, like Vietnamese, lack all such inflection, while Makah applies aspect and mood markers not only to words that are translated into English as verbs, but also to words that are translated into English as nouns or adjectives. Croft (2001) uses "construction grammar", seeking to identify the grammatical categories of individual languages according to the constructions employed in those languages. Of course, one may still relate these to semantic and pragmatic prototypes, but in a language-specific way.

Just as we see the compositionality of language at work in the elaboration of the sentence “He picked a berry” to “He pulled the branch to pick the berry” so can we see the ability to take the “action constructions” <If a berry is within reach, pick it to eat it> and <If something not within reach is attached to something within reach, then grasp the latter and pull it to bring the former within reach> and compose them to plan the action which combines pulling a branch with one hand to permit picking a berry with the other. Informed by this example, we can suggest how a “grammar for action” may relate to a “grammar for language”, without blurring the distinction between the two.

(a) A “grammar for action” would start from multiple “action constructions” and seek to replace ad hoc verbal descriptions by a more formal typology whereby groups of individual constructions could be revealed as instances of a single more abstract construction, and then attempt a principled account of how those structures may be combined.

(b) Praxic action requires temporal and spatial coordination of simultaneously occurring actions whereas language encode “conceptual overlap” in a linear ordering of words and morphemes. Words are required to make explicit both more and less than what is simply “there” in vision and action – we have to say “put the apple on the plate”, but our act simply puts the apple in one place in a whole continuum of resting places (and this language can only handle by “number plug-ins” such as “place the apple 3.15 cm to the left of the daisy in the plate pattern”).

(c) Schema assemblage is not only “somewhat syntactic” as in (a) but also involves “tuning”. In learning to drive, one may rapidly acquire the “syntax” that relates turning direction to motion of the steering wheel, etc., but this novel coordinated control program will be of little use in heavy traffic unless tuned by extensive practice.

(d) The genius of language is that it supports metaphorical extension whereby constructions emerge that can be linked step by step back to the concrete but in the end gain their independence as a domain of expression. Just consider how much experience is required to go from a 5-year old mind to one which can see the previous sentence as meaningful.

5. Forward and Inverse Models in Relation to Grammars for Production and Perception

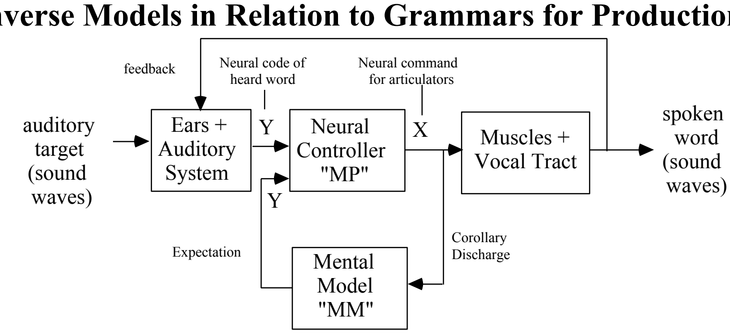

Figure 2. A recasting of the Jordan-Rumelhart model of imitation: train a *direct (forward) model* MM (mental model) of the neural correlates of the transformation of motor commands to auditory output, and then train an *inverse model* MP (motor program) which can generate motor commands which match auditory inputs.

Perception and production grammars may be related to inverse and forward models of motor control. The present section introduces these ideas and shows how they have been related to the mirror system for grasping. I leave for other papers the challenge of extending these ideas to coordinated control programs rather than single actions, and the exploration of the implications of this for an action-oriented theory of grammar.

Jordan & Rumelhart (1992) considered the problem of a human infant learning to pronounce words. The Neural Controller for the Motor Program (MP, Figure 2) does not have direct access to the sound of the word or the articulations that produce a sound. Instead, it must go from the neural code for a heard word to the neural code for its articulation.

Let X be the set of neural motor commands and Y the set of sensory patterns produced by auditory inputs.[6] Then the physical system

Muscles + Vocal Tract + Sound Waves + Ear + Auditory System

defines a function $f: X \to Y$ such that f(x) is the neural code for what is heard when neural command x is sent to the articulators.

The challenge is to train MP to yield successful imitation of an auditory target – the catch being that neither signal set X nor Y is accessible to the teacher. The answer offered by Jordan & Rumelhart (1992) is to "backpropagate through the world":[7]

- First learn the "Mental Model" MM, a network which provides a *direct* model of the effect of commands – MP models the "physics" $f: X \to Y$, predicting the sensory effects in Y of a wide set of commands in X.
- Once MM is built, it provides a stable reference for training MP by adjusting the total network [MP → MM] so that input sounds are reproduced fairly accurately.

MP thus comes to implement a function

$$g: Y \to X \text{ with the property } f(g(y)) \approx y$$

for each desired sensory situation y, generating the motor program for a desired response. I.e., MP becomes an *inverse model* of the effect of commands: Given the neural code for an auditory target y, MP will produce the neural code for a motor command which will generate a spoken sound that approximates y.

[6] For simplicity, this model ignores timing. Each y in Y is an *instantaneous* neural pattern that encodes *all of the last word heard*; and each x in X is an *instantaneous* neural pattern that encodes *all of the next word to be uttered.*

[7] "Backpropagation" is a way of training a network across many trials so that it comes to match input patterns to desired output patterns. However, there is nothing in this discussion that rests on the training method used, and the results described here have also been attained with reinforcement learning.

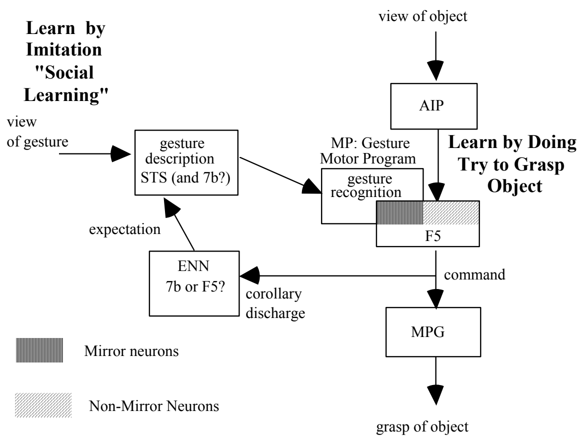

Figure 3. A conceptual framework analyzing the role of F5 in grasping (Arbib and Rizzolatti, 1997). The right hand path provides mechanisms for grasping a seen object. The loop on the left provides mechanisms for imitating observed gestures in such a way as to create expectations which enable the visual feedback loop to serve both for (delayed) error correction during one's own actions and for "social learning" through imitation of the actions of others.

Figure 3 presents a conceptual framework for analysis of the role of F5 in grasping (Arbib and Rizzolatti, 1997, building on Jordan & Rumelhart, 1992). Here

• The input is the visual sign of a gesture.

• The direct model of Command → Response is called the Expectation Neural Network (ENN), not MM: When F5 generates a command, ENN generates the *expected neural code* for the visual signal generated by the resulting gesture. They explicitly label the input to ENN as the corollary discharge of the motor command.

• The Motor Program MP is the inverse model of Command → Response, going from a desired response to a command which can generate it.

Reconfiguring Figure 2, Arbib and Rizzolatti (1997) obtained Figure 3:

i) The right hand, vertical, path is the *execution system* from "view of object" via parietal area AIP (visual recognition of affordances – possible ways to grasp an object) and F5 (motor schemas) and motor cortex to grasp an object. This pathway (and the way in which prefrontal cortex may modulate it) has been analyzed in the FARS model (Fagg and Arbib, 1998).

ii) The loop on the left provides both the *observation matching system* from "view of gesture" via gesture description (posited to be in superior temporal sulcus, STS) and gesture recognition (mirror neurons in F5 or area 7b) to a representation of the "command" for such a gesture, and the *expectation system* from an F5 command via ENN to MP, the motor program for generating a given gesture. The latter path may provide visual feedback comparing "expected gesture" and "observed gesture" for monkey's self-generated movements, and also create expectations which enable the visual feedback loop to serve for learning an action through imitation of the actions of others.

Arbib and Rizzolatti (1997) conflate three separate roles of the mirror system in (ii):

(a) the use of the mirror system to provide visual feedback appropriate to dexterous action;

(b) the ability to recognize the actions of others; and

(c) the ability to use that recognition as the basis for imitation.

I currently would argue (Arbib, 2002, 2005) that (a) and a simple form of (b) were already present in the common ancestor of human, chimpanzee and monkey; that (a) and simple forms of (b) and (c) were already present in the common ancestor of human and chimpanzee; and that the extended capacity for (b) and (c) – complex action analysis and complex imitation – evolved along the hominid line. But since we are dealing here with the relation between action and language in the human brain, we may proceed with the diagram as labeled, but with F5 now interpreted loosely as being in or near Broca’s area with somewhat different localization at each end of the Saussurean sign.

The integrated model of Figure 3 thus relates the "grasp an object" system to the "view a gesture" system. The expectation network is driven by F5 irrespective of whether the motor command is "object-driven" (via AIP) or "gesture-driven". It thus creates expectations both for what a hand movement will look like when "object-driven" (an instrumental action directed towards a goal) or "gesture-driven" (a "social action" aimed at making a self-generated movement approximate an observed movement). The right hand path of Figure 3 exemplifies "learning by doing", refining a crude "innate grasp" - possibly by a process of reinforcement learning, in which the success/failure of the grasp acts as positive/negative reinforcement. The left hand path of Figure 3 exemplifies another mode of learning (the two may be sequential or contemporary) which creates expectations about gestures as well as exemplifying "social learning" based on imitation of gestures made by others.

Arbib and Rizzolatti (1997) discussed two main possibilities. The first is that MP is located along the path leading from STS to F5 via 7b. The reciprocal path from F5 to superior temporal sulcus would provide the direct model, ENN. An alternative would be that both ENN and MP are located in F5. Oztop & Arbib (2002) developed the MNS model of MP on the former basis. They showed how signals from STS and 7b encoding the relation between hand and object over time could become correlated with the F5 codes for a variety of grasps. Iacoboni et al. (2001; Iacoboni , 2004) analyze human brain imaging studies to support the view that STS is responsible for the visual representation of observed actions with (i) the connections from STS to PF and onwards to the mirror cells in F5 forming an inverse model, converting this visual representation into a motor plan, and (ii) connections from mirror cells in F5 to PF, and back to STS forming a forward model converting the motor plan back into a predicted visual representation (a sensory outcome of action). Miall (2003) has offered a brief essay on possible new modeling studies inspired by this viewpoint.

6. Discussion

I argue that language and action both build on the evolutionary breakthrough that gives us a brain capable of complex action analysis and complex imitation.. However, the challenges of “linearizing thought” posed by language are sufficiently different from those of spatial interaction with the world that I accept that there is some specialization of neural circuitry to support language. But specialization in what sense? We may distinguish (i) a “language-ready brain” able to master language as the child matures within a language-using community from (ii) a brain which “has” language (in the sense, e.g., of an *innate* principles and parameters Universal Grammar). I argue

(Arbib, 2005) that biological evolution endowed us with the former not the latter, and that the transition from protolanguage to full languages is a matter of human history rather than biological selection.

The question "A sentence is to speech as what is to action?" forced us to confront the prior question "What is a sentence to speech?" and see that the answer will be far more equivocal than "What is a sentence to well punctuated writing?" But this legitimizes the view that we can characterize basic actions and then define a grammar of action whereby coordinated control programs can be combined to define "sentence level actions". As is well known, a compound action may become so well practiced that they then become available to act as new "action words". In the same way, complex concepts which must be referred to by complex phrases may acquire single words if referred to frequently enough. Thus the line between word, phrase and sentence can shift over time. On the other hand, just as sentences may be concatenated to yield the verbal equivalent of a paragraph, so too may "action sentences" may be combined in less and less loosely structured ways.

Admitting this, I suggest that an *action sentence* is the interwoven structure of basic actions (recall that in general execution of actions is not strictly sequential) that results from combining a relatively small number of schemas to form a coordinated control program and then executing that program according to the contingencies of the current situation. I'm not insistent that action privileges the "action sentence" in the same way that written language privileges the sentence. But just as we ask what constitutes a constituent in language, and then ask how constituents aggregate to yield sentences that are relatively free standing, so we can do the same for action, and in this understanding of the relation of a complex behavior to the underlying coordinated control program that describes the interweaving of its components. It may be that construction grammar provides the best fit for the pragmatic skills of language use – with specific subskills corresponding to specific constructions.

Whereas many accounts of language are so structured as to hide the commonalities between action in general and language use in particular, I stress that much of what one does with language involves machinery that is available more generally. This does not deny that special machinery is needed to augment the general machinery to use special tools (words) to conduct communication in a relatively sequential manner. It is the task of neurolinguistics to understand how mechanisms supporting syntax and pragmatics compete and cooperate to make the perception and production of language a successful tool for communication and social coordination.

References

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

Aboitiz, F., García, R., Brunetti, E., and Bosman, C. (2004) The origin of Broca's area and its connections from an ancestral working memory network, in *Broca's region* (Yosef Grodzinsky and Katrin Amunts, Editors), Oxford University Press, in press.

Arbib, M.A., 1981, Perceptual Structures and Distributed Motor Control, in *Handbook of Physiology, Section 2: The Nervous System, Vol. II, Motor Control, Part 1* (V. B. Brooks, Ed.), American Physiological Society, pp. 1449-1480.

Arbib, M.A. (2002) The Mirror System, Imitation, and the Evolution of Language, in Imitation in Animals and Artifacts, (Chrystopher Nehaniv and Kerstin Dautenhahn, Editors), The MIT Press, pp. 229 - 280.

Arbib, M.A. (2004a) Interweaving Protosign and Protospeech: Further Developments Beyond the Mirror, *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* (to appear).

Arbib, M.A. (2004b) How Far is Language Beyond Our Grasp? A Response to Hurford, in *Evolution of Communication Systems: A Comparative Approach*, D.K. Oller and U. Griebel, E.). The MIT Press, in press.

Arbib, M.A. (2005) From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, *Behavioral and Brain Sciences*. (In press.)

Arbib, M.A., and Bota, M. (2003) Language Evolution: Neural Homologies and Neuroinformatics, *Neural Networks* 16:1237–1260.

Arbib, M.A., and Bota, M. (2004) Response to Deacon: Evolving mirror systems: homologies and the nature of neuroinformatics, *Trends in Cognitive Sciences*, 8:290-291.

Arbib, M., and Rizzolatti, G., 1997, Neural expectations: a possible evolutionary path from manual skills to language, *Communication and Cognition*, 29, 393-424.

Barrett A.M., Foundas, A.L., Heilman, K.M., (2005): Speech and gesture are mediated by independent systems, *Behavioral and Brain Sciences,* (in press).

Corina DP, Poizner H, Bellugi U, Feinberg T, Dowd D, O'Grady-Batch L. (1992) Dissociation between linguistic and nonlinguistic gestural systems: a case for compositionality. *Brain Lang*. 43(3):414-47.

Croft, W. (2001) Radical construction grammar: syntactic theory in typological perspective, Oxford: Oxford University Press, 2001.

Deacon, T. (2004) Monkey homologues of language areas: computing the ambiguities. *Trends in Cognitive Sciences* 8:288-290.

Fagg, A.H., and Arbib, M.A., 1998, Modeling Parietal-Premotor Interactions in Primate Control of Grasping, *Neural Networks*, 11:1277-1303.

Ferrari PF, Gallese V, Rizzolatti G, Fogassi L. (2003) Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex.. Eur J Neurosci. 17(8):1703-14.

Fogassi, L., and Ferrari, P.F., 2004, Mirror neurons, gestures and language evolution, *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* (to appear)

Hurford (2004) Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution, in *Evolution of Communication Systems: A Comparative Approach*, (D. Kimbrough Oller and Ulrike Griebel, Eds.), Cambridge, MA: The MIT Press.

Iacoboni M, Koski LM, Brass M, Bekkering H, Woods RP, Dubeau MC, Mazziotta JC, Rizzolatti G (2001) Reafferent copies of imitated actions in the right superior temporal cortex. *Proc Natl Acad Sci USA* 98:13995-13999.

Iacoboni M. (2004) Understanding others: Imitation, language, empathy. In: *Perspectives on Imitation: From Cognitive Neuroscience to Social Science, Volume 1: Mechanisms of Imitation and Imitation in Animals*. (S. Hurley and N. Chater, Eds.). Cambridge, MA: MIT Press.

Jordan, M.I., and Rumelhart, D.E., 1992, Forward Models: Supervised Learning with a Distal Teacher, *Cognitive Science*, 16:307-354.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. Science, 297(5582), 846-848.

MacNeilage, P.F. & Davis, B.L. (2004) The frame/content theory of evolution of speech: Comparison with a gestural origins theory. *Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems.* (to appear)

Miall R.C., 2003, Connecting mirror neurons and forward models, *NeuroReport*, 14:2135-7.

Miller, G.A., Galanter, E., and Pribram, K.H., 1960, *Plans and the Structure of Behavior*, Henry Holt & Co.

Oztop, E., and Arbib, M.A., 2002, Schema Design and Implementation of the Grasp-Related Mirror Neuron System, *Biological Cybernetics,* 87:116-140.

Rizzolatti, G., and Arbib, M.A. (1998) Language Within Our Grasp, *Trends in Neurosciences*, 21(5):188-194.

Studdert-Kennedy, M. 2002. Mirror neurons, vocal imitation and the evolution of particulate speech, in *Mirror Neurons and the Evolution of Brain and Language,* (M. Stamenov and V. Gallese, Eds.), Amsterdam/Philadelphia: John Benjamins, pp.207-227.