MIRROR NEURONS AND THE ACTION THEORY OF LANGUAGE ORIGINS

Luc Steels
Vrije Universiteit Brussel (AI Lab) and
Sony Computer Science Laboratory Paris 6 Rue Amyot 75005 Paris
tel: 33-1-44 08 05 05
fax: 33-1-45-87-87-50
email: steels@arti.vub.ac.be

a. Introduction and Goal

The research reported here attempts to understand how language may have originated from sensori-motor competences. Recently the observation of mirror neurons [1] has lead to the suggestion that there is not only a rich representation of motor action but also that this representation is used for multiple purposes: action execution, action planning, action imaging, and action recognition. Of particular importance is the observation that one agent can recognise an action plan of another one and that the same neurons are involved.

The relevance of this for the origins of language has been pointed out by Rizzolatti and Arbib [2]. Here we go a step further, arguing that the meaning of a language utterance in general is a series of physical or mental actions that the speaker wants the hearer to perform, rather than a declarative statement to be stored whose only relevance are its truth conditions. For example, when a speaker says "Can you give me the black box on the table?", he wants the hearer to hand over an object (which means to grasp it and move it in the direction of the speaker). To know which object is involved, the speaker wants the hearer to direct his or her attention to a table in the shared context, to identify the objects which can be compared to the prototype of a box, and then focus on the one box which has a black colour. These mental actions are as situated and grounded as motor actions like grasping.

From this action-oriented view of language semantics, language understanding amounts to the recognition of the plan intended by the hearer and the utterance is seen as giving hints about which plan is intended. The production of an utterance can also be seen as involving the contruction of an action plan and thus parsing amounts to the recognition of which production plans have been used by the speaker. So the production of an utterance, both the conceptualisation of what to say and the decision on how to say it, can be viewed as the planning of a series of actions, and the interpretation of an utterance can be seen

as the recognition of these action patterns and their subsequent execution.

Taking this point of view has two important implications: (1) It helps to understand how language might have originated. If the mechanisms required for language are essentially the same as those required for motor planning, execution, and recognition, then it is less a mystery how homo sapiens could have started to evolve language. We no longer need a scenario based on genetic mutations (as in [5]) but can assume a pre-adaptation scenario, in which existing brain structures and processes became used for language communication. (2) It leads to a greater overall economy of the human cognitive system because fewer special-purpose components (like a dedicated language organ) are needed.

To demonstrate the theoretical viability of this thesis we have to show that the same representational framework is adequate for sensori-motor behavior and both for the conceptualisation and interpretation of utterances and for the verbal behavior itself (the production and recognition of utterances). We also have to show that the same learning mechanisms are involved. This is obviously a very non-trivial exercise given the complexity involved.

b. Materials and methods

So far, we have been developing formal models and have conducted computer simulations and experiments with physical robots to test them. The robots have a sensori-motor layer for executing autonomous behaviors, and a fully integrated cognitive layer for planning, memory, and communication. The robot bodies in our experiments range in complexity from steerable cameras [3] to small mobile robots, animal-like robots (specifically the dog-shaped SONY AIBO), and humanoid torsos.

The robots play language games, either among themselves or with a human player. Each language game is a situated interaction between at least two agents about something in their shared environment. It involves perception, conceptualisation, communication, interpretation, and action.

An example game that we have used extensively is the guessing game, in which the speaker draws the attention of the hearer to an object in the shared reality by verbal means [3]. In one large-scale experiment, a growing population of close to 3000 (virtual) agents was employed which

used the (real) robot bodies to engage in guessing games about scenes consisting of geometrical figures on a white board in front of them. Another example game that we have used extensively is the "Where-Is-It?" game, in which agents locate objects based on a spatial map acquired by exploring and remembering the environment and verbal suggestions of a path to follow.

The first step in our research has been to operationalise a representational framework of actions and action plans in the form of schemata. Each schema has a number of slots, constraints on each slot, and an action plan in the form of augmented finite state machines. The automaton schedules and de-schedules sensori-motor behaviors and moves from one state to another based on success or failure in behavior execution. The constraints are maintained by propagating information as fast as it becomes available using data-flow computation. A schema may itself be a specialisation of a more abstract schema and may call upon other schemas. This representational framework was demonstrated to be adequate for the actual high level control of grounded robotic behaviors. We have also developed a learning system capable to acquire new motor schemata by the exploration of a search space of possible concatentations of the primitive actions and by a chunking of successful paths.

Our second step has been to use the same framework to plan the meaning of natural language utterances as needed for language games. The primitive actions in this case are operations over cognitive spaces, such as filtering a set into a subset, shifting the focus of attention from one object to another, or ordering the members of a set into a sequence and retrieving the first member. These conceptual schemata are tightly coupled with the sensori-motor layer in the sense that the information items and facts used by them have all been deposited in memory by sensori-motor behaviors and are continuously upgraded by them.

The third step has been to use the same framework for the execution and the recognition of the utterances themselves. The primitive actions of verbal schemata center on the production (or recognition) of parts of utterances in a specific order and on the realisation of suprasegmental modulations such as prosody and stress patterns. The planning of verbal behavior is itself a highly complex process and known to be distinct from the actual execution of the plan. We are interested in natural dialogues which are highly situated in the specific interaction context of speaker and hearer, with many false starts, hesitations, irrelevant words, etc. This makes verbal behavior much closer to sensori-motor

behavior than is usually assumed, particularly by linguistic theories that exclusively lok at "clean" written language.

Finally we have developed a two-way associative memory that is mapping conceptual schemata to verbal schemata. While parsing an utterance the hearer must recognise which verbal schemata wre involved and map them to the conceptual schemata that could have been intended by the speaker. While producing an utterance the speaker must conceptualise what he wants to say in terms of conceptual schemata and map them onto verbal schemata that constitute a plan for how to express the meaning. We have been experimenting with memory-based learning techniques to gradually build up the repertoire of form-meaning mappings [4].

c. Results

At this point we have been able to demonstrate the complete architecture on autonomous robotic agents. For example, in the large-scale experiment alluded to earlier [3], we have observed that a stable communication system based on a vocabulary of a few thousand words indeed emerges and is maintained in the population even if new members continuously enter or leave the system. A self-organising semiotic dynamics has been observed damping synonymy and polysemy due to a positive feedback loop between use and success [5].

These grammatical forms express the conceptual plans made by the speaker and recognised by the hearer. Even though a vast amount of work is still required to enrich the schema repertoires by the addition of more primitive actions and by integrating more complex learning mechanisms, we can say that based on the results so far the original thesis has gained in plausibility. The planning and plan execution mechanisms required for sensori-motor behavior can form the basis of language.

d. Conclusions

The Action Theory of the origins of language argues that there is a very tight analogy between the ability to plan and recognise a motor action and the ability to plan and recognise an utterance, both its content (what to say) and its form (how to say it). Our research is developing in full detail this analogy by operationalising it on physical robots. We believe that such experimentation is complementary to neurobiological observation and a potentially rich source for detailed models of human verbal behavior.

e. References

[1] Gallese, V., L. Fadiga, L. Fogassi, G. Rizzolatti (1996) Action recognition in the premotor cortex. Brain 119:593-609.
[2] Rizzolatti, G. and M. Arbib (1998) Language within our grasp. Trends Neuroscience. 21:188-194.
[3] Steels, L. (1998) The origins of syntax in visually grounded robotic agents. Artificial Intelligence 103 (1-2), 133-156.
[4] Steels, L. (2000) The Emergence of Grammar in Communicating Robots. In: Proceedings of the European Conference on AI, Berlin. IOS Press, Amsterdam.
[5] Pinker, S. and P. Bloom (1990) Natural Language and Natural Selection. Behavioral and Brain Sciences, 13, 707-784.