First International Workshop on Emergence and Evolution of Linguistic Communication, May 31- Jun 1 2004, Kanazawa Japan

Language Emergence and Grounding in Sensorimotor Agents and Robots

Angelo Cangelosi[1], Thomas Riga[1], Barbara Giolito[1,2], Davide Marocco[1,3]

(1) Adaptive Behaviour & Cognition Research Group, University of Plymouth (UK)
(2) Department of Human Studies, Universita' del Piemonte Orientale - Vercelli (Italy)
(3) Artificial Life and Robotics Lab, Institute of Cognitive Sciences and Technology - CNR (Italy)

email: angelo.cangelosi@plymouth.ac.uk
http://www.tech.plym.ac.uk/soc/research/abc

1. Modeling Language with Grounded Agents and Robots

The grounding of linguistic symbols in the organism' cognitive system, and indirectly in the physical and social environment in which individuals live, is one of the most important issues in recent experimental and computational approaches to language. This is normally referred as the Symbol Grounding Problem (Harnad 1990). In cognitive science, psychological experiments have focused on the relationship between language and perception (Barsalou 1999; Coventry & Garrod 2004) and language and action (Glenberg & Kaschak 2002). These empirical investigations show a strong inter-dependence between language development and perceptual and embodiment factors. In robotics and artificial intelligence, various models have been proposed to ground language in neural networks (Joyce et al. 2003; Cangelosi et al. 2000; Dyer 1994) and in interactive robots (Steels 2002; Vogt 2002; Roy & Pentland 2002). Moreover, computational approaches to the evolution of language (Cangelosi & Parisi 2002) commonly deal with the issue of symbol grounding.

In this paper we present a computational modeling approach to language based on an integrative view of the agent's cognitive system. This work has mainly been developed at the "Adaptive Behaviour & Cognition" research group[1] of the University of Plymouth. All models reviewed here are characterized by the fact that the emergence of linguistic abilities (both evolutionarily and developmentally) is strictly dependent on, and grounded in, other behaviors and abilities. These include sensorimotor skills (e.g. action categories), cognitive abilities (e.g. categorical perception), neural mechanisms, and social and evolutionary factors. Agents are able to build an intrinsic link between the linguistic symbols (words) they use to communicate and their own sensorimotor and cognitive representations (meanings) of the perceptual and sensorimotor interaction with the external world (referents). This approach is called "grounded adaptive agent modeling" for the emergence of language (Cangelosi in press). Linguistic simulations with such agents imply the use of groups of autonomous agents that interact via language games to exchange information about the environment. It also entails the fact that their coordinated communication system is not externally imposed by the researcher, but emerges from the interaction between agents. The paper will present a series of past and current studies on grounded adaptive agent models. Some will be based on simulated sensorimotor agents, others on evolutionary and epigenetic robots.

2. Emergence of Nouns and Verbs in Adaptive Agents

Models based on adaptive agents simulate a multi-agent scenario in which individuals use a lexicon to communicate about their environment. Various adaptive agent models have been used to model the evolutionary emergence of language (e.g. Cangelosi 2001; Hashimoto & Ikegami 1996). Here we focus on grounded agent models, where the individual's neural network is used to control all sensorimotor, cognitive and linguistic abilities. In previous studies (Cangelosi & Parisi 2001) we showed that agents were able to evolve the ability to use a syntactic lexicon consisting of proto-verbs (names of actions) and proto-nouns (names of objects). The lexicon was externally imposed by the researcher. In these models, the analysis of the agents' neural representations can highlight the neural mechanisms responsible for the integration of language, cognition and action. For example, categorical perception analyses (Cangelosi et al. 2000; Cangelosi & Parisi 2001; Cangelosi in press)

[1] http://www.tech.plym.ac.uk/soc/research/ABC

have shown that when agents use verbs and nouns, the similarity space of the representations of verbs is enhanced and optimized with respect to that of nouns. In addition, synthetic brain imaging techniques (Arbib et al. 2000; Cangelosi & Parisi 2004; Cangelosi in press) have shown that the neural representations of syntactic word classes are sensitive to the level of integration of linguistic information and sensorimotor knowledge. In particular, verbs produce more neural activity in the regions of the network that specialize for the integration of sensorimotor information. Nouns are more active in the modules dedicated to the processing of sensory (e.g. visual) information only.

Recent simulations (Giolito & Cangelosi in preparation) have extended this model to include (i) the cultural acquisition of language through communication with other agents, (ii) the linguistic production of nouns and verbs. The main aim of such a simulation consists in testing some hypothesis on the emergence of nouns and verbs. We hypothesize that there is a semantic contribution to the grammatical distinction between nouns and verbs. In particular, we suggest that some semantic aspects of the grammatical category of nouns are related to the objects to which nouns refer, and that some semantic aspects of the grammatical category of verbs are related to the actions that subjects are performing. In other words, our hypothesis is that the distinction among nouns depends - at least in part - on the variations of the objects with which organisms are interacting, while the distinction among verbs depends - at least in part - on the variations in the actions that organisms are performing on these objects. A term is a noun because its semantic aspect depends on the objects it refers to, whereas another term is a verb because its semantic aspect depends on the actions it refers to. Our intention is not to deny the existence of an additional 'grammatical' basis for the distinction between nouns and verbs, but to suggest that a part of this distinction is based on a sort of semantic differentiation. Moreover, we suppose that such a differentiation originates from the evolutionary interaction among organisms and between the agents and their environment.

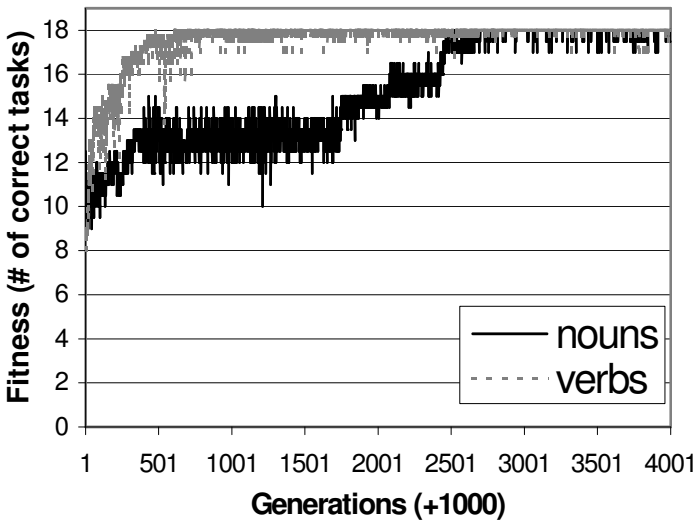

Figure 1: Emergence of a verb-noun lexicon in a population of adaptive agents.

To test such an hypothesis, we used agents whose neural controller consists of: 4 proprioceptive units, 25 visual units and 4 language units in the input level; 2 layers of hidden units; and 8 output units (4 motor and 4 linguistic units). Agents are subject to an evolutionary algorithm for 5000 generations. Each generation consists of a no-language object manipulation task (push/pull 2 different objects) and 10 other linguistic epochs (various combinations of visual images and noun/verb linguistic inputs). During the first 1000 generations, agents are evolved solely for their ability to perform the object manipulation task. The linguistic tasks are introduced at generation 1001 and last until generation 5000.

Preliminary results on the emergence of nouns and verbs indicate a tendency to evolve a lexicon based on both verbs and nouns. Although only 2 populations out of 15 evolve a full verb-noun lexicon (see Figure 1 for a sample population), in all other populations a great deal of correct verbs or nouns also emerges. These results show that agents evolve an ability to differentiate between the two word classes. This confirms that at least a partial grounding on the noun-verb distinction can be originated

from the evolutionary use of sensorimotor information and interaction with the environment. New simulations are currently underway to improve the robustness of the results and produce more verb-noun languages. These will mainly focus on the modification of the neural network architecture, as suggested in a related model on verb-noun control in modular neural networks (Cangelosi, in press).

3. Emergence of Communication in Evolutionary Robots

Evolutionary robotics has been successfully applied to the synthesis of robots able to exploit sensorimotor coordination (Nolfi & Floreano, 2000) and signaling capabilities (Quinn, 2001). Most of the properties of evolutionary robotics (e.g. sensorimotor coordination, social interaction, evolutionary dynamics, use of neural controllers) also are beneficial for modeling the emergence of grounded communication. For example, in recent simulations, robotic agents first evolve an ability to create action categories by interacting with physical objects.

Recent linguistic experiments (Marocco, Cangelosi & Nolfi, 2003) have shown that the ability to form categories from direct interaction with the environment constitutes the ground for subsequent evolution of names of objects (e.g. nouns). In this model, agents are implemented with simulated robots provided with a 3-segments arm with 6 degrees of freedom (DOF) and extremely coarse touch sensors. Agents use proprioceptive information behavior to actively explore the environment and build categories, on the basis of tactile information, and the communication, about the type of objects that are in it. The environment consists of an open three-dimensional space in which one of two different objects is present in each epoch. The two objects used in this simulation are a sphere and a cube.

The sensory system consists of a simple contact sensor placed on the body that detects when this body collides with another and proprioceptive sensors that provide the current position of each joint of the arm. The controller of each individual consists of an artificial neural network in which, in addition to proprioceptive sensors, two sensory neurons receive their input from the other agents. The output layer has motor neurons, which control the actuators of the corresponding joints, and two additional output neurons, which encode the signal to be communicated to the other agents. The two linguistic units work as a small winner-takes-all cluster, where the neuron with the highest activation is set to 1 and the other to 0. This means that, in addition to the proprioceptive information, agents also receive in input a 2-bit signal produced by some other agent in the population, such as the parent or any agent from the population (linguistic comprehension task). The protocol of interaction and communication between agents was systematically changed. Before they act as speaker, agents undergo a linguistic production task. That is, each agent is put in the environment and asked to interact with the object. The value of the two output neurons in the last cycle of the epoch is saved and used as the signal produced to "name" the object. A genetic algorithm is used to evolve the behavior of agents.

The evolutionary robotics model was used to run a series of experiments on the role of various social and evolutionary variables in the emergence of shared communication. The first independent variable of the experimental design is the selection of speakers: each agent receives communication signals solely from its own parent or from any individual of the population. This looks at the role of different social groups of speakers in facilitating shared communication. The second independent variable is the time period in which communication is allowed: agents can communicate right from the initial random generation or only after the pre-evolution of the ability to touch/avoid the two objects. Through this variable it will be possible to investigate the initial behavioral and cognitive abilities necessary to evolve communication. The simulation results show that populations evolve stable shared communication (i.e. using two different signals are produced for the two different objects) mostly when the parents act as speakers and when signaling is introduced in the second stage (Marocco et al., 2003). Additional analyses of results support the following findings: (a) the emergence of signaling brings direct benefits to the agents and the population, in terms of increased behavioral skill and comprehension ability (but the agents' fitness does not assess the ability to communicate well); (b) there is a benefit in direct communication between parents and children, not only because of kinship mechanisms, but also because parents produce more stable and reliable input signals; (c) the pre-evolution of good sensorimotor and cognitive abilities permits the establishment of a link between production and comprehension abilities, especially in the early generations when signaling is introduced.

A second model has been tested, in order to simulate the emergence of different types of syntactic categories. In the previous model, the signal associated to each object can be simultaneously

interpreted as the name of the object ("noun") or as the name of the action ("verb"). For example, when a parent produces the signal X at the end of its interaction with the sphere, the child can interpret it as "sphere" (which indicates the object) or as "touch" (which indicates the action to be performed). In the new model, the touch/avoid actions are not rigidly linked to any specific object. The agents can touch and avoid both the sphere and the cube, depending on the task context. This is defined by the parent's language and the fitness formula. When a parent produces a signal X after having touched the sphere, the child must interpret it as the "touch" verb and touch any object regardless of its shape. Thus this signal can be clearly considered as the name of the action, which is a typical case of verbs. Our aims was the identification of significant differences between the verb lexicon of the second simulation and the signaling system in the first simulation. Such comparisons provided useful insights on the evolutionary transition from signaling to syntactic languages reinforcing the fact that the pattern of results in the first simulation correspond to that of the evolution of verbs, as observed in the adaptive agent model of verbs and nouns (Cangelosi & Parisi, 2001).

4. Imitation and Language in Epigenetic Robots

More recently, new experiments have focused on the developmental emergence of communication in epigenetic robots (Riga & Cangelosi, in preparation). In this study simulated robots observe and execute actions via imitation learning, while using an artificial language to communicate about the names of actions and objects. Analyses of the model, through synthetic brain imaging techniques, highlight the specific role of sensorimotor knowledge in the processing of the words for actions.

We developed a computer simulation of two robots embedded in a virtual reality world (Figure 2), that accurately models physical constraints present in real-world situations, using the physics engine Open Dynamics Engine. Each robot has 12 degrees of freedom and consists of two 3-segment arms attached to a torso and a base with 4 wheels. The teacher robot has preprogrammed behavior to manipulate objects; the imitator agent learns to perform the same actions by observing the teacher executing them. The imitator approximates the teacher's movements using an on-line mimicking algorithm, resulting in movement dynamics that are processed by a neural network, which memorizes action patterns and enables for their autonomous execution. Visual pre-processing is simulated, thus the agent receives joint angles directly in input instead of analyzing visual input to compute joint angles. This integrated system models the mechanism responsible for matching perceived actions with equivalent self-generated actions and the way in which actions are memorized and successively reproduced autonomously. It addresses the question of how agents learn to perform actions on objects using simple imitative mechanisms like mimicking.

In a first simulation the imitator agent learned to perform actions when receiving a linguistic description of them. Furthermore it learned to give a linguistic description of actions performed by the teacher agent. The neural network controlled both motor and linguistic behavior. Descriptions consisted of a verb, indicating the action, and a noun, referring to the object involved. Brain imaging studies (Pulvermueller, 2003; Cappa & Perani, 2003) reveal that language comprehension and production activates different brain areas for verbs and nouns. In particular they indicate that the left temporal neocortex exhibits strong activation in tasks involving lexical-semantic processing, while additional regions of the left dorsolateral prefrontal cortex are recruited during the processing of words related to actions. We applied synthetic brain imaging techniques (Arbib et al., 2000) to analyze the internal neural network structure that emerged during training.

The results showed that noun processing involved the area responsible for object recognition while verb processing recruited neurons in both the object recognition area and the area responsible for motor program execution. The neurons in the object recognition area differentiated by specializing in either noun processing or motor program execution. The fact that the agents, in absence of a linguistic input, performed a different default action for every object caused the object recognition area to have a double function: categorizing the objects and bootstrapping a default action in absence of linguistic input. This simulation supports the view that language builds on existing neural structures responsible for action execution and observation (Glenberg & Kaschak, 2002; Rizzolatti & Arbib, 1998). Linguistic representations are embodied: they build on sensorimotor representations. In fact, embodiment effects can be detected when the imitator agent observes its teacher: it performs very small movements in synchrony with the observed action.

During the second simulation agents learned to perform basic actions by mimicking them, while simultaneously learning words corresponding to these actions. Furthermore they learned higher-level composite behaviours by receiving linguistic descriptions containing these previously acquired words. The agents merged basic actions into a composite action by transferring the neural grounding of the words referring to basic actions to the word indicating the higher-level behaviour. This process of grounding transfer (Riga et al., 2003) grounds words, known exclusively from linguistic descriptions, on the neural level by adapting neural activations of the words contained in the description.

The imitator robot, during training, learned the basic actions of opening and closing their left and right arms (upper arms & elbows), lifting them (shoulders), and moving forward and backward (wheels), together with the corresponding words. At the 50th epoch it received 1st level linguistic descriptions of combined actions, consisting in a new word and two known words referring to basic actions. A combined action consisted for example in grabbing the object in front of them and was described like: "close_left + close_right = grab". Grounding was transferred from "close_left" and "close_right" to "grab". Consequently, when the agent was given the command "grab" it successfully executed the combined action of pushing its arms towards the object and grabbing it. At the 100th epoch it received second level descriptions, in which a defining word was itself learned exclusively from a linguistic description. Following the example of before, we combined grabbing and moving forward into carrying: "move_forward + grab = carry". Also at this level grounding was successfully transferred to the new word, enabling the agent to correctly perform the action of carrying on hearing the word "carry": it pushed both arms against the object and moved forward, effectively exhibiting the behavior of carrying the object. The system learned several of these combined actions simultaneously, and also four-word definitions and grounding transfers of up to three levels have been realized.

The second simulation sheds light on language as a cognitive enhancer, as a means through which new behaviors can be acquired quickly and effortlessly, building on experience accumulated by previous generations of agents. The importance of cultural transmission in cognitive development is highlighted. Our long-term goal is to develop a framework for training robots by demonstration, using both imitation and a natural language interface, enabling for a neuro-robotic approach to investigating imitation as a precursor of communication.

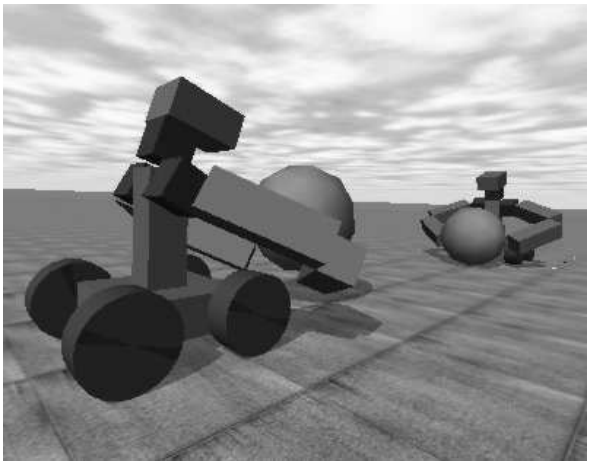

Figure 2: Simulation setup for the model of imitation and communication in epigenetic robots.

5. Conclusion

Adaptive agent models and evolutionary and epigenetic robots can significantly contribute to a better understanding of the strict interdependence between language and perceptual, motor and cognitive capabilities. Such models of language emergence have important scientific and technological implications for research in language and communication. In robotics and artificial intelligence, they provide new approaches and algorithms for the development of autonomous interactive systems. In cognitive science, these models permit a deeper understanding of the psychological and cognitive bases of language and its grounding in perceptual and sensorimotor abilities. Finally, in linguistics and other disciplines interested in language origins, agent and robotics models allow the simulation of the evolutionary emergence of language and the test of language origin hypothesis.

6. References

Arbib M.A., Billard A., Iacoboni M., Oztop E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Networks*, 13: 975-997.

Barsalou L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22: 577-609.

Cangelosi A. (in press). The sensorimotor bases of linguistic structure: Experiments with grounded adaptive agents. *The Eighth International Conference on the Simulation of Adaptive Behaviour – SAB04*, Los Angeles, July 2004

Cangelosi A. (2001). Evolution of communication and language using signals, symbols, and words. *IEEE Transactions on Evolutionary Computation*. 5(2): 93-101

Cangelosi A., Greco A., Harnad S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, 12(2): 143-162

Cangelosi A., Parisi D. (2001). How nouns and verbs differentially affect the behavior of artificial organisms. In J.D. Moore, K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, London: LEA, 170-175.

Cangelosi A., Parisi D. (2002). *Simulating the Evolution of Language*. London: Springer.

Cangelosi A., Parisi D. (2004). The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging. *Brain and Language*, 89(2): 401-408

Cappa S.F., Perani, D. (2003). The neural correlates of noun and verb processing. *Journal of Neurolinguistics*, 16 (2-3): 183-189.

Coventry K.R., Garrod S.C. (2004). *Saying, Seeing and Acting: The Psychological Semantics of Spatial Prepositions*. Psychology Press. Hove and New York

Dyer M.G. (1994). Grounding language in perception. In V. Honavar, L. Uhr (Eds.), *Artificial Intelligence and neural networks: Steps toward principled integration*. Boston: Academic Press.

Glenberg A., Kaschak M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.

Harnad S. (1990). The symbol grounding problem. *Physica D*, 42: 335-346

Hashimoto T., Ikegami T. (1996). Emergence of net-grammar in communicating agents, *BioSystems*, 38: 1-14

Joyce D., Richards L., Cangelosi A., Coventry K.R. (2003). On the foundations of perceptual symbol systems: Specifying embodied representations via connectionism. *Proceedings of the 5th Intl. Conference on Cognitive Modeling (ICCM 2003)*. Bamberg

Marocco D., Cangelosi A., Nolfi S. (2003), The emergence of communication is evolutionary robots. *Philosophical Transactions of the Royal Society London – A*, 361: 2397-2421

Nolfi S., Floreano D. 2000 *Evolutionary Robotics*. Cambridge, MA: MIT Press.

Pulvermueller F. (2003). *The Neuroscience of Language. On Brain Circuits of Words and Serial Order*. Cambridge University Press.

Quinn M. (2001). Evolving communication without dedicated communication channels. In J. Kelemen & P. Sosík (Eds.), *Proceedings of ECAL01*, pp. 357-366, London: Springer.

Riga T., Cangelosi A., Greco A. (in press). Symbol grounding transfer with hybrid self-organizing/supervised neural networks. *IJCNN04 International Joint Conference on Neural Networks*. Budapest, July 2004

Rizzolatti G., & Arbib M. (1998). Language within our grasp. *Trends in Neuroscience*, 21: 188-194.

Roy D., Pentland A. (2002), Learning words from sights and sounds: A computational model, *Cognitive Science,* 26: 113-146.

Steels L. (2002). Grounding symbols through evolutionary language games. In A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*. London: Springer-Verlag.