The emergence of language: how to simulate it

Domenico Parisi[1], Marco Mirolli[1,2]

[1] Institute of Cognitive Sciences and Technologies, National Research Council,
44 Via San Martino della Battaglia, 00185 Rome, Italy
{domenico.parisi, marco.mirolli}@istc.cnr.it
[2] Department of Philosophy and Social Sciences, University of Siena, 47 Via Roma, 53100
Siena, Italy
Mirolli2@unisi.it

Abstract. The emergence of language in populations of primates that initially lacked language can be simulated with artificial organisms controlled by neural networks and living, evolving, and learning in artificial environment. Some simulations have already been done but most are a task for the future. We discuss language evolution under two topics: language is learned from others on the basis of genetically inherited predispositions, and language has important influences on human cognition. We propose an evolutionary sequence according to which bipedalism and the emergence of the hands represent a selective pressure for developing an ability to predict the consequences of one's actions, this ability is the basis for learning by imitating other individuals, learning by imitating other individuals is applied to learning to imitate their communicative behaviour. The second topic include the consequences of language for various aspects of human cognition, especially when language is used to talk to oneself.

The abstract should summarize the contents of the paper and should contain at least 70 and at most 150 words. It should be set in 9-point font size and should be inset 1.0 cm from the right and left margins. There should be two blank (10-point) lines before and after the abstract. …

1 The Starting Point and the End Point of Language Evolution

Everyone agrees that language is among the most important characteristics that distinguish human beings from other animals. Therefore, it would be an important scientific achievement to clearly understand and explain how language has emerged in human beings' evolutionary history. What we do know is that if we move sufficiently back in time we find ancestors of human beings that lacked language and that all modern humans, except for pathologies, have language. But because of the extreme complexity of human language and because language does not leave fossil traces, we only have speculative theories concerning how and when language has emerged and evolved. In this context, computer simulations can be of help. Simulations are computer programs and to express our hypotheses and theories as computer programs forces us to formulate our hypotheses and theories in a more explicit, detailed, and complete way than if they are just expressed in words because, otherwise, it would be impossible to translate them in a computer program. Furthermore, since a simulation's results are the empirical predictions which are derived from the theory incorporated in the simulation, hypotheses and theories expressed as simulations generate many detailed empirical predictions in a mechanical and, therefore, uncontroversial way. This allows us to make the best possible use of whatever empirical evidence we do have on the initial emergence and further evolution of human language. Of course, simulations simplify with respect to reality but this is true for all theories in science. Scientific theories let us better understand the extreme complexity of empirical phenomena because they simplify and try to capture the essential mechanisms and processes that lie behind the phenomena and explain them. The real problem is that simulations, and theories in general, should make the correct simplifications, that is, they should include the critical entities and factors that explain the phenomena of interest, and leave the rest out. But this can only be judged in each particular case.

If we want to simulate the historical process through which human language has first emerged in some proto-form and has then changed to reach its current form, we should have some idea of the initial state and of the terminal state of this evolutionary process. The initial state is some primate species, probably analogous but not necessarily identical to chimpanzees living today. Chimpanzees have a given body and a given brain, and they have given cognitive and social/communicative abilities. It is possible, however, that a communicative system which was clearly, even if to some limited extent, different from the communication systems of the living chimpanzees, first made its appearance not when our evolutionary line first separated from their evolutionary line, that is, 5 or 6 million years ago, but some time later. During this interval some changes may have occurred in the human evolutionary line that may have played an important role in the first emergence of human language. For reasons that we will discuss later, one of these changes could have been bipedalism and the emergence of the hands with their great manipulative powers. In any case, our simulations should start with populations of "agents" that resemble living chimpanzees or modified forms of living chimpanzees which incorporate whatever changes have occurred in our evolutionary line that may have played a role in the first emergence of language.

With respect to the final state we can be somewhat more detailed since we are dealing with something which is very complex but can be directly observed: current human language. What are the most important aspects of human language that we should be able to simulate? This is a possible list (cf. also Hockett 1960):

1. Human language is culturally learned. A human being is not born with language and language does not emerge as the result of a purely maturational process. Every child acquires a specific form of human language, i.e., an historical language such as English, Italian, or Chinese, which is the specific language spoken in the particular environment in which the child happens to live and grow. Hence, language is learned from others, by imitating others.
2. Human language is learned by the child because the human genotype includes some information that makes the acquisition of language possible. This is sug-

gested by a number of facts. Just to mention a few: nonhuman primates which are exposed to a human language do not acquire language, except in some very limited and special forms, while all normal children do acquire a language without any apparent effort; there seems to be a universal developmental program with which language is acquired by all children, with specific stages and specific timings for those stages; all human languages have shared characteristics. The genetic preparedness for language may include both predispositions that are not specific for language but are necessary for learning a language and may have played a role in the initial emergence of language, and predispositions that are specific for language.

3. Historical languages change constantly. The cultural transmission of language is accompanied by changes that may be the result of many different mechanisms, and this implies that language changes across successive generations. One of the sources of language change is the imperfectness of linguistic cultural transmission. Another is the fact that language changes as the result of the communicative interactions among individuals of the same generation. Furthermore, groups of individuals that interact with each other tend to develop a shared language which is different from the language of other groups because, even if the two groups descend from a single group speaking the same language, the changes tend to diverge.
4. Language is a complex communicative system with specific properties that make it different from animal communicative systems. Among these properties are its compositional and hierarchical structure, with phonemes making up words, words making up phrases, and phrases making up sentences.
5. Unlike animal communicative systems, human language is used not only for communicating with others but also for communicating with oneself (thinking).
6. Unlike animal communication systems, language has a crucial impact on human cognition, changing the way in which humans know and categorize the world, remember the past, and predict and plan the future.

To simulate the evolutionary emergence of language and its subsequent changes is to start with a population of artificial organisms that resemble living apes and do not have language and gradually arrive to a population which has language with the six properties listed above. What we will do in this chapter is to discuss in somewhat more detail some of these properties of human language and how to simulate them. We will organize our discussion under two main headings: (1) language is learned from others on the basis of genetically inherited predispositions, and (2) human language influences human cognition. In some cases we will refer to simulations that have already been done but most of the work remains a task for the future.

2 Language is learned from others on the basis of genetically inherited predispositions

Although there are necessary species-specific genetic predispositions for learning a language, language is learned and is learned from others. The child acquires the specific language which is spoken in its environment by imitating the linguistic behaviour of other individuals. Therefore, one first requirement for simulating the evolutionary emergence of human language is to be able to construct artificial organisms that can learn from others. Simulations in which a communication system emerges in a population of artificial organisms across a succession of generations but the communication system is entirely encoded in the organisms' inherited genotype and there is no individual learning, can be simulations of the evolutionary emergence of animal communication systems but not of human language.

One important consideration is that human beings do not only learn language from others but they learn from others all sorts of behaviours and abilities. In contrast, most other animals do not learn from others at all. Their communicative behaviour tends to be genetically inherited and their other behaviours tend to be either genetically inherited or learned by interacting with the non-social environment. Of course, there are exceptions that it is important to consider, but learning from others as a general adaptive strategy appears to be typical of the human species.

Can we simulate learning from others? One possible simulation model is the following. The brain of our artificial organisms is simulated with a neural network, a simplified model of the nervous system with units corresponding to neurons and connections between units corresponding to synapses between neurons. The basic neural network possessed by all organisms is a sensory-motor network which, in each cycle, maps sensory inputs, encoded as activation patterns in the network's sensory units, into movements, encoded as activation patterns in the network's motor units. The organisms live in a physical environment, which implies that the movements of their motor organs (including their phono-articulatory organs) cause changes in the environment that can be sensed by the sensory units of both their own neural network and the neural network of conspecifics.

One simple way of simulating imitation in neural networks is the following. Both the imitated individual and the imitating individual receive exactly the same input. They both produce an output in response to this input, and the output of the imitated individual (the model) is used by the imitating individual (the learner) as the teaching input of a standard back-propagation procedure, leading to changes in the connection weights of the learner's neural network that cause the learner's output to become progressively more similar to the model's output. In this way, after a number of input/output cycles, the learner will have learned to behave like the model (Denaro and Parisi 1996). This way of implementing imitative learning in neural networks, however, suffers from a fundamental 'ecological' implausibility, being based on the direct comparison between the learner's motor output and the motor output of the imitated individual. Real organisms have no access to the motor commands of other individuals but only to the consequences of their actions on the environment.

We can solve this problem using a more sophisticated model of imitation. Imagine a more complex neural network which, in addition to the basic sensory-motor module, includes an additional set of units which, taken together, constitute a prediction module. On the basis of the current sensory input and the planned movements with which the neural network will respond to the input, the prediction module generates an activation pattern which corresponds to the activation pattern that will appear in the network's sensory units in the next cycle. This activation pattern is a prediction of the next sensory input. Neural networks can learn to predict their next sensory input using the backpropagation procedure, with the actual next sensory input, resulting from the actually executed movement, functioning as teaching input. The network compares its prediction with this teaching input and, on the basis of the discrepancy between the two (error), modifies the connection weights of its predictive module in such a way that in a succession of learning cycles the error goes to (almost) zero. The network has learned to make correct predictions.

After the prediction module has learned to make correct predictions, the sensory-motor module and the prediction module are connected together in such a way that when a sensory input arrives from outside to the sensory-motor module the sensory-motor module can learn to respond by generating a movement that reproduces the sensory input. In other words, the network learns to imitate sensory inputs. This is done by (a) generating a planned movement in response to the sensory input, (b) generating a prediction of the sensory input that will result from this planned movement, (c) comparing the predicted input with the actual sensory input resulting after the planned movement has been executed, and (d) using the discrepancy between the two to change the connection weights of the sensory-motor module, while leaving unchanged the connection weights of the prediction module (which can already make correct predictions). To eliminate the discrepancy, the network will learn to generate movements that reproduce the sensory input which has caused the movement (cf. Jordan and Rumelhart 1992). If the sensory input is the result of the organism's own movements, the network will learn to imitate its own behavior (self-imitation). If the sensory input is caused by the behaviour of another individual, by reproducing with its behaviour this sensory input the network will learn to reproduce the other individual's behaviour (imitation).

This model of learning to predict and learning to imitate can be applied to the pre-linguistic development of the child in its first year of life (Parisi and Floreano 1992). What happens to the child in the first year of its life which is relevant for language can be viewed as a succession of four stages. In Stage 1 (prediction), which covers the very first months of life, the child learns to predict the acoustic events that will result from its phono-articulatory movements. The child generates all sorts of phono-articulatory movements producing all sorts of sounds and, since it belongs to the human species and is a predicting animal, learns to anticipate which sound will result from which phono-articulatory movement. In stage 2 (self-imitation), at around 4-6 months, the child learns to imitate its own sounds. The child produces a sound, hears it, and reproduces it (babbling). In stage 3 (imitation), which covers the second semester of life, the process becomes social. Now the child pays attention to the sounds that are produced by other people and it learns to reproduce those sounds, that is, to produce sounds that resemble the sounds of the particular language which is spoken in the child's environment. Hence, the sounds produced by the child in the second semester of life tend to be different for children living in different linguistic communities. Finally, in Stage 4 (language), which starts at around 1 year of age, true language beings. Other individuals produce specific sounds in response to specific objects and the child learns to produce the same sounds that are produced by other individuals in response to the same objects. In other words, the sounds acquire a meaning for the child because the child notices that specific sounds systematically occur in its experience with specific objects. This is the beginning of language production and language comprehension. The child becomes able to produce one specific sound, not by imitating a heard sound, but in response to the object systematically paired in its experience with the sound (language production), and to respond to a specific sound not by imitating the sound but by executing the non-linguistic action normally evoked by the object systematically paired with that sound (language comprehension) (Mirolli and Parisi 2005a).

What are the implications of this simulation model of the linguistic development which occurs during the child's first year of life? As already noted, human beings do not learn only the sounds of language and the meaning of these sounds by imitating other individuals. They learn all sorts of other behaviours and abilities by imitating other individuals, and this is specifically true of humans, not of other animals. Therefore, one is led to formulate the following hypothesis. After the human evolutionary line has separated from the evolutionary line of living apes 5-6 million year ago, two genetic predispositions have been incorporated in the genotype of our evolutionary line: a predisposition to learn to predict the consequences of one's actions and a predisposition to apply this ability to predict to learning to behave like other individuals by imitating their behaviour.

It is an open question whether the ability to predict has evolved prior to the ability to imitate or the two abilities have evolved together. One important consideration is that the ability to predict can be adaptively useful also at the individual level, that is, independently of its usefulness for learning by imitating others. The ability to predict the effects of one's actions can be useful in hunting, in throwing objects, and in using and constructing artefacts. For example, in using an artefact it may be useful to be able to predict the changes that one's actions mediated by the artefact will cause in the environment. In constructing an artefact, it may be useful to be able to predict the changes that one's actions will cause in the artefact that one is constructing. However, even if some predictive ability may have initially emerged in our evolutionary line with these purely individual functions, it might also be that learning by imitating others has been a selective pressure for developing a more sophisticated ability to predict the results of one's actions. This hypothesis is suggested by our model of imitation which implies that the ability to predict the results of one's actions is a necessary component of learning by imitating others, i.e., to make one's actions similar to the actions of others. (Consider that our model of imitation can be also applied to the imitation of artefacts, i.e., to making copies of existing artefacts.) In any case, by assuming that the incorporation of a tendency to learn to predict the consequences of one's actions in the human genotype has been a critical step in human evolution, we can explain in an economic way many different aspects of the human adaptive pattern: a general ability to generate more effective behaviours, the use, construction, and imitation of artefacts, and the tendency to learn by imitating others.

How and why has the tendency to learn to predict the consequences of one's actions been incorporated in the human genotype? We know that one of the first novelties that has appeared in the human evolutionary line after its separation from the chimpanzees' evolutionary line has been bipedalism and the consequent freeing of the hands for manipulation purposes. This has implied a great enlargement of our ancestors' repertoire of behaviours. With their hands free to manipulate objects, our ancestors of 2-4 million year ago have become able to do many more different things and to cause many more different effects in the environment compared to their quadruped ancestors. The sheer increase in the number of different actions and of different effects of these actions has made the problem of choosing among the different actions more complex. We hypothesize that this has been a selective pressure for developing a tendency to pay attention to the consequences of each of these different actions and to learn to predict their consequences. Using simulations it can been shown that the behaviour of artificial organisms becomes more effective, in a variety of different ways, if they are able to predict the consequences of their actions (Parisi et al. 1990, Nolfi et al. 1994). This can be shown even if the artificial organisms have a very simple behavioural repertoire which includes only one type of action such as approaching food. We assume that the selective advantage of a tendency or ability to learn to predict the consequences of one's actions has become much greater when, endowed with hands that can manipulate the environment, our ancestors' behavioural repertoire has become much more extended.

Once a genetically inherited tendency/ability to learn to predict the consequences of one's actions has been encoded in the genotype of our ancestors, this tendency/ability has been recruited and exploited to develop two other tendencies/abilities: using and then constructing artefacts and imitating the behaviour of others. As we have already said, the influence may have not been only one-way but the adaptive significance of using and constructing artefacts and of imitating the behaviour of others may have represented a selective pressure to further develop the tendency/ability to predict the consequences of one's actions.

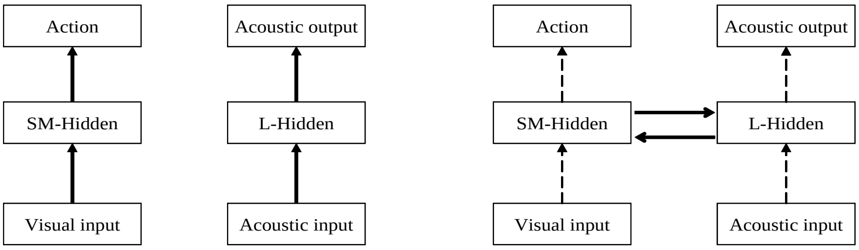
A well developed ability/tendency to learn by imitating others may have had a critical role in the emergence of a communication system such as human language which, unlike most animal communication systems, is culturally, not genetically, transmitted. Notice that learning to imitate the sounds produced by others appears to be easier than learning to imitate their other behaviours, for two reasons. The effects produced by phono-articulatory movements, i.e., the sounds that these movements create in the environment, depend almost uniquely on the phono-articulatory movements themselves, and on no other factor. This is not true for other types of movements, whose effects depend on both the movements themselves and other factors existing in the environment. For example, the effects of hitting a stone with another stone depend on the nature of the hitting movement (its direction and force) but also on the physical properties of the two stones. Hence, it may be easier to learn to predict the effects of one's phono-articulatory movements (sounds) than to predict the effects of other types of movements. The second reason why predicting the sounds resulting from phono-articulatory movements is easier is that sounds are very accessible (e.g., sounds can be perceived from a distance and they go around obstacles) and they are the same sounds for any number of individuals sufficiently close to the source of the sound, i.e., the individual who are produced the phono-articulatory movements. This is less true, for example, for the visual effects of hand movements which tend to be less accessible (they cannot be perceived from a distance and obstacles can make them non-accessible) and they may be somewhat different for different individuals looking at the hands from different spatial locations.

This may be related to the question whether human language has originated in the acoustic/phono-articulatory form of today's language (as commonly assumed), or it has first emerged in a gestural, visuo/motor, form and only some time later it has moved to an acoustic/phono-articulatory form (as proposed, for example, by Corballis 2002 and Arbib 2005). On the one hand, the hypothesis that language have evolved in the acoustic/phonoarticulatory form from the beginning is supported not only by the fact today's natural medium of language is acoustic, but also by the fact that humans seem to inherit a species-specific set of predispositions to process and to produce linguistic sounds. As we have seen, this is shown, among other things, by the regular succession of stages in the phono-articulatory behaviour of the child in its first year of life. On the other hand, there is a growing body of evidence for a visuo-motor origin of human language. One kind of evidence is the easiness with which congenitally deaf children learn the sign languages of the deaf and in the emergence of gestural languages in communities of deaf people. Another is the importance of gestures in the very first phases of linguistic development (Volterra et al. 2005). Still another kind of evidence is the fact that the brain region which is principally devoted to control of speech in humans - Broca's area - appears to be homologous to the brain region which principally controls the production and understanding of hand actions in monkeys - the F5 area, where mirror neurons have been found (Rizzolatti and Arbib 1998). All these could be explained by a visuo-motor origin of language which has then moved to the acoustic/phono-articulatory form of present-day language because of the greater effectiveness of learning and using an acoustic/phono-articulatory language, without erasing the preceding visuo-motor evolutionary stage. This hypothesis is compatible with the important role played by the hands in the evolution of humans in the last 2-4 million years, that we have discussed above. On the other hand, one can hypothesize that language was acoustic/phono-articulatory from its beginning and that gestural languages emerge as a generalization from the genetically inherited tendency to learn an acoustic/phono-articulatory language.

These issues could be decided, or at least illuminated, by doing simulations. For example, can we evolve artificial organisms that learn from others a language which is acoustic/phono-articulatory from its beginning? Will an artificial deaf individual belonging to a population of organisms with a genetically inherited tendency/ability to learn an acoustic/phono-articulatory language, be able to learn a visuo-motor language as easily as an individual without this genetically inherited tendency/ability? Or, for artificial organisms with very able hands, is it easier to first evolve a visuo-motor language and then move to an acoustic/phono-articulatory one because the acoustic/phono-articulatory language is more effective?

Language is culturally transmitted from one generation to the next but it also changes across a succession of generations, and what distinguishes human language from animal communication systems is not only that it is culturally transmitted but that it changes across successive generations with the typical speed of cultural evolution. Cultural evolution can be simulated in populations of artificial organisms if cultural transmission is selective and there are mechanisms that constantly introduce new variability (Hutchins and Hazelnuz1995, Denaro and Parisi 1996, Mirolli and Parisi 2004). One starts with a population of organisms with individually different linguistic behaviours, The different individuals are not all equally imitated by the individuals of the next generation; some potential models have more imitators than others. This is cultural selective reproduction. Furthermore, imitation is never perfect and one single individual can imitate many different "models" so that its linguistic behaviour is not identical to the linguistic behaviour of any of these models but is a novel recombination of parts of their different linguistic behaviours. The selective reproduction of linguistic behaviours and the constant addition of new linguistic behaviours cause language change across generations (Hare and Elman 1995).

Not only language change but also the emergence of different historical languages can be simulated. Although linguistic behaviour is somewhat different in different individuals, in groups of individuals that descend from the same cultural ancestors and that imitate each other linguistic behaviours tend to be similar, constituting an historical language. If the group splits in two subgroups with little interaction between the two subgroups, the original mother-language gives origin to two different, even if historically related, daughter-languages.



(a) **(b)**

3 Human language influences human cognition

Language first emerges in the child at around 1 year of age. It is approximately at this age that the child appears to be able to connect heard or pronounced sounds with specific objects and actions. We have assumed that the basic neural network that controls the behaviour of all organisms is a sensory-motor network that maps sensory inputs into motor outputs. The organism perceives something and it responds with some appropriate movement. We now assume that in the human brain there are two such networks and that from birth to 1 year these two networks are either anatomically or functionally separated (Figure 1a). One sensory-motor network, the non-linguistic network, maps non-linguistic sensory inputs into non-linguistic motor outputs. The connections weights and perhaps even the architecture of connections of this network change during the first year of life as the child learns to coordinate its movements and to respond to sensory inputs with the appropriate movements in reaching, touching, moving its eyes, etc. The other sensory-motor network, the linguistic network, maps linguistic (acoustic) sensory inputs into linguistic (phono-articulatory) motor outputs. As we have already seen, this second network learns during the first year of life to self-imitate and then imitate linguistic sounds.

Fig. 1. (a) The non-linguistic sensory-motor network and the linguistic sensory-motor network are anatomically or functionally separated until 1 year of a child's life. (b) At 1 year the two networks become connected when the child learns the weights of the connections going from the non-linguistic network to the linguistic network (language production) and of the connections going from the linguistic network to the non-linguistic network (language comprehension)

At around 1 year the two networks become connected (Figure 1b). The child begins to learn the appropriate connection weights for the connections going from the non-linguistic network to the linguistic network and for the connections going from the linguistic network to the non-linguistic network. This is the beginning of language proper. Specific sounds produced by other individuals tend to be experienced by the child together with specific objects or specific actions and these systematic correspondences between linguistic sounds and non-linguistic objects and actions are incorporated in the weights of the connections linking the two networks. Given these weights, the child becomes increasingly able to respond with the appropriate motor outputs of its linguistic network (phono-articulatory movements) to sensory inputs to its non-linguistic network (naming a perceived object or a perceived action) and to respond with the appropriate motor outputs of its non-linguistic network (movements of the eyes, face, arms, hands, legs) to sensory inputs to its linguistic network (understanding linguistic sounds). In the model outlined here, language learning is learning the appropriate weights that go from the non-linguistic network to the linguistic network and from the linguistic network to the non-linguistic network. One general consequence of learning these weights is that much of an individual's cognitive activity can consist in going from sounds to meanings and from meaning to sounds. This can make the individual's cognitive activity more effective in a variety of ways.

Categorization. One of the influences of language on cognition concerns categorization. Different objects are put together in the organism's brain if these different objects are to be responded to by the organism with the same action. The different objects that are responded to with the same action constitute a category. In neural network terms, a sensory input is an activation pattern which is transformed in another activation pattern in the network's internal units by the connection weights of the connections linking the sensory input units to the internal units. Let us call the activation pattern in the internal units the internal representation of the sensory input. The ability to categorize consists in the possession of connections weights that tend to make more similar the internal representations of sensory inputs that must be responded to with the same action and to make more different the internal representations of sensory inputs that must be responded to with different actions (Harnad et al. 1995, Di Ferdinando and Parisi 2004). The activation pattern of a set of units can be conceived as one point in a hyperspace with as many dimensions as the number of units and with the point's location in each dimension corresponding to the activation level of the corresponding unit. A category is a "cloud" of points in the hyperspace of the internal units that correspond to the internal representations of sensory inputs which must be responded to with the same action. A neural network which has "good" categories is a neural network whose "clouds" are small (inputs that must be responded to with the same action are internally represented in similar ways) and distant from each other (inputs that must be responded to with different actions are internally represented in different ways).

What is the effect of possessing a language on the organism's categories? As will be recalled, language begins when, at 1 year, the non-linguistic network is functionally linked to the linguistic network, and vice versa. Therefore, the internal representations of the non-linguistic network, which prepare the motor outputs with which the non-linguistic network will respond to its sensory inputs, tend to be influenced not only by the sensory inputs to the non-linguistic network but also by the linguistic network. What changes in the internal representations of the non-linguistic network as a consequence of this linking? The answer is that the non-linguistic network's categories tend to become better categories, that is, the "clouds" of its internal representations become smaller and more distant from each other (Mirolli and Parisi 2005a, in press). Since an organism's categories influence the organism's behaviour by making it easier for the organism to select the appropriate action in response to sensory inputs, an organism endowed with language will have a more effective behaviour.

Learning of categories. The influence of language on categories may go well beyond the mere improvement of existing categorical representations. The categories that form inside an organism's neural network are in fact the result of the organism's experience with the world which allows the organism to learn the appropriate responses to the different sensory inputs. This experience can be long and costly. An individual that acquires a language also acquires the categories marked by the linguistic signals (Steels and Baelpaeme?). In this way the individual can exploit the experience of other individuals, both living and long dead, and acquire more easily more useful categories (Cangelosi and Harnad 2000, Cangelosi et al. 2000). This makes human cognition intrinsically social. Furthermore, it has been shown, both empirically (Waxman and Markov, 1995; Nazzi and Gopnik, 2001) and by neural networks simulations (Schyns 1991, Lupyan 2005), that category learning itself can be improved by the aid of linguistic labels. Linguistic labelled categories can also have a negative side, however, since they may induce a conception of reality as made up of clearly distinct class of entities, internally homogeneous and unchanging.

Selective Attention and Voluntary Control. Language can also be a mechanism for directing attention to specific portions of the input arriving from the environment and for articulating or analyzing complex sensory inputs. All organisms need selective attention mechanisms since all organisms live in environments that send to their sensory organs many different inputs at the same time, and the organism must select which of these inputs to process in order to generate a response, while ignoring all the other inputs. Language can be such a selective attention mechanism. When an individual sees a complex scene, a word originating from another individual which accompanies the perception of the complex scene can help the individual to isolate some particular component of the complex scene and to respond to this component, ignoring the other components. This is a consequence of the co-variation of specific sounds with specific non-linguistic inputs which gives linguistic sounds their meaning. Language can also help the individual to articulate a complex perceived scene into its elements. A sentence is a collection of linguistic sounds (words) each of which co-varies with a different component or aspect of a complex scene so that the sentence makes it easier for the individual that hears the sentence to isolate these different components and aspects and to respond more effectively.

Prediction and Planning. The capacity to predict and to plan can be deeply improved by language for two reasons. First, predictions can become more complex and articulated if they are linguistically labelled. Plans as sequences of actions for reaching some particular goal can also become more complex because the individual can work more effectively with linguistically labelled predictions and with linguistically labelled planned actions in response to these predictions. Finally, evaluations of linguistically labelled predicted effects of planned actions can be used to decide whether to actually execute those actions or not, making planning easier, more effective, and more “reasoned”. Second, linguistically labelled predictions and plans can be shared and discussed with others, making the overall predicting and planning capacity of single individuals and of groups more effective.

Memory. Another example of the importance of language for cognition concerns short-term memory. One way of simulating short term memory with neural networks consists in copying the activation pattern of a network’s internal units in a special set of memory units and then connecting the memory units to the internal units so as to allow the network to retrieve the memory trace (Elman 1990). We can imagine that both the non-linguistic network and the linguistic network (Figure 1) have this type of short-term memory mechanism. However, the quantity of information, as measured by number of units and connections, contained in the non-linguistic network tends to be much greater than the quantity of information contained in the linguistic network. This implies that it is generally easier to remember words than actual sensory-motor experience. This, in turn, has the consequence that an individual possessing language can work more easily with linguistic information and translate this information into the associated non-linguistic information when this becomes necessary. Delegating the memory function to the linguistic system can have the further advantage of leaving the sensory-motor system free to process other information useful for acting in the environment while linguistically remembering previous information. And, indeed, empirical evidence seems to confirm the importance of the linguistic system for human memory (Gruber and Goschke 2004).

Talking to Oneself. A crucial characteristic of human language which distinguishes it from animal communication is that, unlike animal communication systems, language can be used not only for communicating with others but also for communicating with oneself, i.e., for thinking. Inputs to an individual’s linguistic network can come from another individual’s linguistic network but they can also come from the individual’s own linguistic network: the individual talks to itself. If the sounds are actually (physically) produced by the individual's phono-articulatory movements and actually heard by the sensory units of the individual's linguistic network, we call it private speech. If the loop does not include the organism's peripheral motor and sensory organs but is more internal, we call it inner speech (Mirolli and Parisi in press). Inner speech is faster than private speech and as a consequence can be more useful for certain purposes, but in both cases a number of interesting effects on the individual's cognitive activity can be observed.

One could think that using language to talk to oneself is a late development in human evolution and it presupposes an already complex language. However, in some simulations it has been shown that talking to oneself can be a selective pressure for the evolutionary emergence of a very simple communicative system if the linguistic signals are used by the individual to keep in memory some information which has been received from an another individual (Mirolli and Parisi 2005b). In these simulations the signals are genetically inherited and are not learned from others, so they are not linguistic signals, but the simulations can be a demonstration that using language for oneself can have advantages for the individual even if the language is very simple.
Using language to talk to oneself has a number of important consequences for human cognition. As we have seen, one's sensory-memory categories becomes better (smaller and more distant "clouds" of points in the space of sensory-motor internal representations) if sensory-motor categories are linguistically labelled through the bidirectional connections linking the organism's sensory-motor network to its linguistic network. This is true not only when linguistic stimuli arrive from outside but also when they are self-generated by the organism. Hence, an organism with language can work with better sensory-motor categories even when it is all alone and is interacting with the non-social environment. Also the effects of language in directing an individual's attention and in articulating complex scenes can be internalized. Once an individual has experienced the positive effects of being guided by linguistic stimuli produced by other individuals, he or she can learn to linguistically stimulate him- or herself in the same way to produce the same effects. The use of linguistic self-stimulation is indeed very important in child development (Diaz and Berk 1992) and can be considered as the beginning of *voluntary control* (Vygotsky 1962, 1978).

4 Conclusion

The emergence of human language in the descendants of organisms that initially lacked language can be simulated with artificial organisms controlled by neural networks and living, evolving, and learning in artificial environments. The ideal simulation is a simulation in which we start with a population of organisms similar to living chimpanzees and therefore lacking language and these organisms gradually evolve human language as we know it. Aspects of this ideal simulation have already been realized but most are a task for the future.
We have addressed two main topics in the evolutionary emergence of language. Human language is culturally learned from others on the basis of species-specific geneti-

cally inherited predispositions. We propose an evolutionary sequence for the human evolutionary line leading to human language: (1) bipedalism and the evolutionary emerge of hands for manipulating the environment; (2) the emergence of a tendency/ability to predict the effects one's actions on the environment due to the fact that the possession of hands greatly increases the size of one's behavioural repertoire and the number of different effects that can be caused in the environment by using the hands; (3) the use of this ability to predict the effects of one's actions in learning to imitate the behaviour of others by reproducing the effects of their behaviour; (4) the application of the ability to imitate others to imitating their communicative behaviour; (5) the transition from a visuo-motor to an acoustic-phonoarticulatory communicative behaviour.

The other topic of language and of its evolutionary emergence that we have addressed concerns the relation of language to cognition. One critical aspect of human language which makes it different from animal communication systems is that human language has a number of consequences for human cognition and it can be used not only for communicating with others but also for communicating with oneself (thinking). We have examined some of these consequences and how they can be simulated: the improvement of one's sensory-motor categories, the learning of categories without directly experiencing them, the role of language as a mechanism for selective attention and for articulating complex sensory inputs, its role in generating better predictions and better plans by linguistically labelling them and, as a consequence, by becoming able to discussing them with other individuals, and, finally, the role of language in keeping more information in memory in linguistic rather than in sensory-motor form. An important aspect of language learning is language internalization, the transfer of social uses of language to individual uses, i.e., talking to oneself.

Acknowledgements

The research presented in this paper has been supported by the ECAGENTS project founded by the Future and Emerging Technologies program (IST-FET) of the European Community under EU R&D contract IST-2003-1940. The information provided is the sole responsibility of the authors and does not reflect the Community's opinion. The Community is not responsible for any use that may be made of data appearing in this publication.

References

Arbib, M.A.: From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. Behavioral and Brain Sciences, Vol. 28(2) (2005)105-124

Cangelosi, A., Harnad, S.: The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. Evolution of Communication, Vol. 4(1). (2000) 117-142

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

Corballis, M.C.: From Hand to Mouth: The Origins of Language. Princeton University Press, Princeton (2002)

Denaro, D., Parisi, D.: Cultural evolution in a population of neural networks. In Marinaro, M., Tagliaferri, R. (eds.) *Proceedings of the Vietri-96 Conference on Parallel Architectures and Neural Networks*. Springer, New York (1996) 100-111

Di Ferdinando, A., Parisi, D.: Internal representations of sensory input reflect the motor output with which organisms respond to the input. In Carsetti A. (ed.): Seeing, Thinking and Knowing. Kluwer, Dordrecht. (2004) 115-141

Diaz, R., Berk, L. (eds.):: Private speech: From social interaction to self regulation. Erlbaum, New Jersey (1992)

Elman, J.L.: Finding structure in time. Cognitive Science, Vol. 14. (1990) 179-211

Gruber, O., Goschke, T.: Executive control emerging from dynamic interactions between brain systems mediating language, working memory and attentional processes. Acta Psychologica, Vol. 115. (2004) 105-121

Hare, M., Elman, J.L.: Learning and morphological change. Cognition, Vol. 56 (1995) 61-98

Harnad, S., Hanson, S.J., Lubin, J.: Learned categorical perception in neural nets: implications for symbol grounding. In Honavar and Uhr (eds.): Symbol processors and connectionist network models in artificial intelligence and cognitive modelling: steps toward principled integration. Academic Press , New York, NY (1995) 191-206

Hockett, C.F.: Logical considerations in the study of animal communication. In Lanyon, W.E., Tavolga, W.N. (eds.): Animal Sounds and Communication. American Institute of Biological Sciences: Washington, D.C. (1960)

Hutchins, E., Hazlehurst, B.: How to invent a lexicon: the development of shared symbols in interaction. In Gilbert, N. and Conte, R. (eds.) Artificial Societies: the computer simulation of social life. UCL Press , London. (1995) 157-189

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

Lupyan, G.: Carving nature at its joints and carving joints into nature: How labels augment category representations. In Cangelosi, A., Bugmann, G., Borisyuk, R. (eds.) Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop. World Scientific, Singapore (2005) 87-96

Mirolli, M., Parisi, D.: Language, altruism and docility: How cultural learning can favour language evolution. In Pollack, J.B., Bedau, M., Husbands, P., Ikegami, T., Watson, R.A. (eds.): Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems. MIT Press, Cambridge, Mass. (2004) 182-187

Mirolli, M., Parisi, D.: Language as an Aid to Categorization: A Neural Network Model of Early Language Acquisition. In Cangelosi, A., Bugmann, G., Borisyuk, R. (eds.) Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop. World Scientific, Singapore (2005a) 97-106

Mirolli, M., Parisi, D.: How can we explain the emergence of a language which benefits the hearer but not the speaker? Connection Science, Vol. 17(3-4). (2005b) 325-341.

Mirolli, M., Parisi, D.: Talking to Oneself as a Selective Pressure for the Emergence of Language. In Cangelosi, A., Smith, K., Smith, A. (eds.) Proceedings of the 6th Evolution of Language Conference (in press)

Nazzi, T., Gopnik, A.: Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? Cognition, Vol. 80. (2001) 303-312

Nolfi S., Elman J.L., Parisi D.: Learning and evolution in neural networks. Adaptive Behavior, Vol. 3(1). (1994) 5-28

Parisi, D., Cecconi, F., Nolfi, S.: Econets: Neural networks that learn in an environment. Network, Vol. 1. (1990) 149-168

Parisi, D., Floreano, D.: Prediction and Imitation of Linguistic Sounds by Neural Networks. In Paoloni, A. (ed.): Proceedings of the First Workshop on Neural Networks and Speech Processing. Fondazione Bordoni, Roma (1992)

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

Schyns, P.G.: A Modular Neural Network Model of Concept Acquisition. Cognitive Science, Vol. 15(4) (1991) 461-508

Steels, L., Belpaeme ???

Volterra, V., Caselli, M.C., Capirci, O., Pizzuto, E.: Gesture and the emergence and development of language. In Tomasello, M., Slobin, D.: Beyond Nature-Nurture. Essays in Honor of Elizabeth Bates. Lawrence Erlbaum Associates, New Jersey (2005) 3-40

Vygotsky, L.S.: Thought and language. MIT Press, Cambridge, MA (1962)

Vygotsky, L.S.: Mind in society: The development of higher psychological processes. Harvard University Press, Cambridge, MA. (1978)

Waxman, S.R., Markow, D.B.: Words as invitations to form Categories: Evidence from 12- to 13-month-old infants. Cognitive Psychology, Vol. 29. (1995) 257-302