Published as: Mirolli Marco, Parisi Domenico (2005): How can we explain the emergence of a language that benefits the hearer but not the speaker? Connection Science, 17(3-4): 307-324.
doi: 10.1080/09540090500177539

How can we explain the emergence of a language which benefits the hearer but not the speaker?

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

[1] Institute of Cognitive Sciences and Technologies,
National Research Council
15, Viale Marx, 00137 Rome, Italy
domenico.parisi@istc.cnr.it
[2] Philosophy and Social Sciences Department,
University of Siena
47, Via Roma, 53100 Siena, Italy
mirolli2@unisi.it

1. Introduction

1.1 The problem

Language requires the co-evolution of both speakers and hearers. i.e., the co-evolution of both the ability to emit signals that systematically co-vary with specific features of the external or internal environment (language production) and the ability to respond appropriately to these signals (language understanding). If speakers do not produce the appropriate signals in the appropriate circumstances, there are no useful signals for hearers to understand. If hearers do not respond appropriately to the signals produced by the speakers, speakers emitting the appropriate signals are useless. The evolution of both speakers and hearers requires that language must be advantageous to both. Why should speakers evolve an ability to produce signals that appropriately co-vary with specific features of the environment if this does not involve any advantages for them? Why should hearers develop an ability to respond appropriately to the heard signals if this ability does not increase their survival and reproductive chances?

In fact, using computer simulations we will show that, in a situation in which language[1] is advantageous for hearers but not for speakers, a stable and useful language fails to emerge in groups of genetically unrelated individuals, i.e., if the speaker and the hearer do not have the same or similar genes. Of course, there are other uses of language in which both the speaker and the hearer get benefits from, respectively, producing and understanding linguistic signals. And in these circumstances language can emerge even in groups of genetically unrelated individuals. One example is a speaker informing a group of genetically unrelated hearers about the presence of some prey that can only be captured by group hunting. In this context language can emerge because both the speaker and the hearer(s) receive some advantage (capturing the prey) by, respectively, producing the appropriate signals and responding appropriately to the signals.

However, our simulations will show that in at least three conditions language can emerge even if it confers an advantage only to hearers but not to speakers. The first condition is that the speaker and the hearer share the same or similar genes. In these circumstances language emerges because

[1] The computational model presented in this paper is very simple and the 'language' we are dealing with is a very simple communication system whose signals have no internal structure and no syntax. Nevertheless, we will refer to our signalling system indifferently as a 'communication system' and as a 'language'. This is because the problem we are dealing with here (namely the evolution of good speakers when there is no selective advantage for them) is relevant for any kind of communication system used for informing other individuals about the environment, including human language which is particularly sophisticated in this respect. Furthermore, two of the solutions we will find for evolving this kind of communication system (see the 'docility' and the 'talking-to-oneself' simulations, §§ 3.3 and 3.4) are based on two characteristics which are specific of human language: (a) it is a culturally transmitted and culturally evolved system, and (b) it is used to communicate with oneself and not only with others.

the altruistic genes of the speaker are maintained in the population through the advantages conferred by the linguistic behavior of speakers to hearers that possess the same (or similar) genes. In other words, language can emerge in kin-related groups as a result of kin selection (Hamilton 1964).

A second condition in which language emerges even if it is advantageous only to the hearer is a condition in which language is culturally rather biologically transmitted. If what the organisms inherit genetically is not the language itself but only the propensity to learn the language, then this propensity - which, following Simon 1990, we call 'docility' - can be the basis for learning the language from individuals that already know the language, and this irrespective of whether in any particular context of use language confers an advantage to the speaker or to the hearer.

A third condition for the emergence of a language which is useful only for the hearer exploits another peculiar characteristic of human language. Unlike animal communication systems, human language is used not only for social communication, i.e., when the speaker and the hearer are two different individuals, but also for talking to oneself, that is, when the speaker and the hearer are the same individual. This kind of individual use of language constitutes between 20% and 60% of the language produced by three to ten year old children (Berk 1994) and there is evidence that in adult life this private speech is internalized and becomes inner speech (Vygotsky 1962; Diaz and Berk 1992). This use of language for oneself may have been present in the very early evolutionary stages of language and it may have represented an evolutionary pressure for the biological, not cultural, emergence of language. We will show that a useful shared language can biologically evolve in groups of genetically unrelated individuals if language is used as a memory aid, i.e., to allow an individual to keep in memory some useful piece of information which has been communicated by another individual.

1.2 State of the art

The study of the evolution of communication and language using computational models is a very lively and fascinating field. (For a recent and comprehensive review of the field, see Wagner et al. 2003; see also the articles in Cangelosi and Parisi 2002). Indeed, it can be argued that much of the renewed scientific interest in theories of language evolution and origin (Christiansen & Kirby 2003) is due to the introduction of formal and computational methods in the field.

Various models have been described for studying the emergence of both simple, non-syntactic communication systems (Hutchins and Hazlehurst 1995, Di Paolo 1998, Hurford 1999, Billard and Dautenhann 1999, Steels and Kaplan 2002) and more complex, compositional ones (Batali 1998, Hazlehurst and Hutchins 1998, Steels 1998, Kirby 2000, Cangelosi 2001, Cangelosi and Parisi 2001). Most of these models focus on the *mechanisms* that can sustain the emergence of a good communication system and adopt means for assuring such an emergence. Evolutionary models tend to directly include communication success in the fitness function of the individual, while learning models tend to introduce learning mechanisms (such as reinforcement) that force the system towards the emergence of a shared communication system.

Only a few computational and formal models have directly tackled the issue which is of interest here, that is, the adaptive factors that can lead to the emergence of a communication system which is advantageous only for the hearer, but not for the speaker. Ackley and Littman 1994 and Oliphant 1996 have shown that this kind of communication system does not evolve under 'normal' conditions because of its altruistic character. Only if the population is spatialized, then a communication system can evolve even if there is no advantage for the speaker. Both Ackley and Littman and Oliphant argue that in spatialized populations communicating organisms will tend to be kin-related so that altruistic genes can emerge because altruistic agents will tend to benefit other altruistic agents. Di Paolo 1999 has criticized this interpretation by formally analysing kin relatedness in a model similar to that used by Ackley and Littman 1994 and Oliphant 1996, and he

has shown that kin selection cannot fully explain the effects that spatialization has on the emergence of a communication system; instead, spatial organization by itself, together with other factors such as discreteness and stochasticity, can play a major role in the emergence of cooperative communication.

The role of spatial factors in the emergence of simple, altruistic signalling systems is also discussed in Noble et al. 2002, together with other adaptive factors like the handicap principle (Zahavi 1975) and other ecological factors such as the presence of noise.

In a model very similar to the one presented here, Cangelosi and Parisi 1998 were able to evolve an altruistic communication system in a population of ecological neural networks. These authors interpret the emergence of a good communication system in terms of the cognitive pressure towards communication: the evolution of good signallers is explained as a by-product of the independently evolving ability to categorize the environment.

Finally, Marocco et al. 2003 tried to evolve a communication system in a population of neural networks controlling a robot arm whose task was to categorize objects by physically interacting with them. The evolution of such a communication system proved to be very difficult because of its altruistic character. In fact, a good communication system evolved only in 7 out of 10 replications of the simulation, even if (a) the networks were cognitively pre-adapted to resolve the task before the introduction of communication, and (b) the speaker was always the parent of the hearer, so that there was a strong pressure against the emergence of cheating due to kin-selection.

2. Method

In our simulations we use a simplified version of the 'mushroom world' (Parisi 1997, Cangelosi and Parisi 1998, Cangelosi and Harnad 2000). To survive and reproduce individuals must be able to distinguish between edible and poisonous mushrooms and they must eat the edible mushrooms and avoid the poisonous ones. Furthermore, organisms can send signals to each other for communicating the quality of the encountered mushrooms. The only difference with respect to the original mushroom world is that the new world is one-dimensional rather than bi-dimensional. This simplification has been introduced in order to make the study of the population dynamics easier and to discriminate better the roles played by the different adaptive factors.

2.1 The environment and the task

The environment is a corridor made up of 11 cells (figure 1). In each trial an organism is placed in the start cell at the beginning of the corridor and a mushroom is placed in the last cell of the corridor (the 'mushroom cell'). If the organism reaches the mushroom cell it eats the mushroom contained in the cell. Mushrooms can be either edible or poisonous. If an organism eats an edible mushroom its reproductive chances (fitness) are increased by some amount, while eating a poisonous mushroom decreases the organism's fitness (see below). The sensory system of the organisms is very limited: an organism can see the mushroom only if the organism finds itself in the cell immediately preceding the mushroom cell. Hence, without language the only strategy which is available to the organisms is to run through the entire corridor in each trial in order to reach the last but one cell and see whether the mushroom is edible or poisonous. However, if another organism is near to the mushroom so that it can see it, and this organism sends to the first organism a linguistic signal that tells the first organism whether the mushroom is edible or poisonous, the first organism can avoid spending its energy in order to run through the corridor in the trials with poisonous mushrooms. This confers an advantage to the first organism (the hearer) but not to the other organism (the speaker).

[Insert figure 1 about here]

2.2 The neural network

The neural network that controls the organisms' behavior is made up of 17 nodes and 32 connection weights, biases included (figure 2). There are 10 visual input nodes whose activation encodes the perceptual properties (when visible) of mushrooms, and 2 linguistic input nodes with continuous activation between –1 and +1 which encode heard signals.

[Insert figure 2 about here]

Each individual mushroom has perceptual properties that are different from the properties of all other mushrooms. The perceptual properties of mushrooms are encoded as vectors of 10 bipolar values[2] in the input nodes of the neural network controlling the organism's behavior. There are 210 edible mushrooms and 210 poisonous ones. The perceptual properties of edible mushrooms are encoded as all and only the patterns that differ by 4 bits from the prototype represented by ten – 1s, whereas the perceptual properties of poisonous mushrooms are all and only the patterns that differ by 4 bits from the prototype represented by ten +1s.

When organisms act as hearers the activation of the 2 linguistic input nodes corresponds to the activation of the 2 linguistic output nodes of the other organism which is near to the mushroom, while when an organism is playing the role of the speaker the activations of the linguistic input nodes are set to zero.

There are only 2 hidden nodes (with bias) fully connected both with the input nodes and with the output nodes. The locomotion of the organisms is controlled by a single output node with binary activation. If the activation is one, the organism moves to the next cell in the corridor. Otherwise, it stays still. Movement has a cost, so that individual fitness is decreased by one energy unit each time the individual makes one step forward.

Finally, there are two linguistic output nodes with continuous activation between –1 and +1. The activation vector of these two nodes constitutes the linguistic signal that can be transmitted either to some other organism (in the simulations with language) and/or to oneself (in the 'talking-to-oneself' simulation, see § 3.4).

2.3 Individual Life

The life of each individual lasts 420 trials, one for each possible mushroom. In each trial the organism is placed in the start cell of the corridor and one mushroom is placed in the last cell. Another organism is chosen randomly from among the remaining members of the population, it is placed near the mushroom so that it can perceive the mushroom's perceptual properties, and it emits a linguistic signal (linguistic output) which is received by the tested organism as a linguistic input. If the tested organism reaches the mushroom cell, it eats the mushroom and goes on to the next trial. In any case, a trial is terminated after 11 input-output cycles of the neural network.

---

[2] "Bipolar" means that each value can be either +1 or –1. We have used bipolar values instead of binary values (1/0) because bipolar encoding permits to deal better with situations in which the organism does not perceive the mushroom. In such situations we give as perceptual input of the organism's neural network the vector composed by all zeros, so that the influence of the perceptual input on the hidden vector will be null. This strategy seems to be quite straightforward but has a drawback when used with a binary encoding. In fact, in such an encoding zeros have a "meaning" and even though this "meaning" cannot be "represented" in the connection weights between the input nodes and the hidden nodes, it would be represented in hidden nodes' biases.

2.4 The fitness formula

At the end of the 420 trials, the fitness of an individual is calculated according to the following rule: (1) 30 energy units are added for each edible mushroom eaten; (2) 5 energy units are subtracted for each poisonous mushroom eaten; (3) 1 energy unit is subtracted for each movement the organism has made during its life[3]. Hence, for the organisms in the simulations with language, the maximum possible fitness is:

Fmax = 30 × 210 (edible mushrooms) – 5 × 0 (poisonous mushrooms) – 210 × 10 (steps necessary for eating a mushroom) = 6300 – 2100 = 4200

In the graphs the fitness is always normalized (that is, it is divided by 4200) so that the maximum possible fitness is 1.

2.5 The genetic algorithm[4]

In each generation the population is constituted by 100 organisms. The genome of the organisms contains all the connection weights and biases of their neural network. In the first generation the weights and biases are initialized with a random value in the range [– 4; + 4]. After all the organisms have lived their lives and their fitness has been calculated, individuals are selected for reproduction by the roulette with windowing method, they reproduce (sexually), and some mutations are added to the genomes inherited by the offspring. Ten replications (with different initial conditions) are run for each different simulation and each replication lasts 1000 generations.

In the standard roulette method (Mitchell 1996), one creates a 'roulette wheel' such that for each individual the size of its slice of the roulette wheel is proportional to the individual's fitness and, for each wheel spin, the individual selected for reproduction is the one under the wheel marker. In this way, the probability of reproduction for each individual corresponds to the fitness of the individual divided by the average fitness of the population. Adding the 'windowing' method to the roulette method, one subtracts the fitness value of the worst individual of the population for all the individuals, so that each individual will have a fitness value between 0 and the fitness value of the best individual minus the fitness value of the worst individual. This is done in order to maintain a reasonable selective pressure even when the average fitness reaches a high value and the differences in fitness between individuals are very low.

Reproduction is sexual and the population is divided by sex: in each generation there are 50 'males' and 50 'females'. Each individual has two parents, one male and one female. One male individual and one female individual are selected for reproduction using the method described above and they reproduce. Each couple makes two offspring with a double point cross-over. For each offspring, we select two randomly chosen integer numbers between 0 and 31 (the genome is made up of 32 connection weights) which constitute the two points for the cross-over; then, we take one segment between these two points from the 'father' and the other from the 'mother' and generate in this way the genome of the offspring. This double-point cross-over treats the genome as a circle, in that there is no difference in the probability of cross-over between the 'centre' and the 'periphery' of the genome.

---

[3] The values used in the fitness formula are arbitrary. The reason of the asymmetry between the contributions of edible and poisonous mushrooms to fitness (30 vs. 5) is that there must be some advantage in moving and eating randomly; otherwise, at the beginning of the simulation organisms that stay still would be selected and no evolution could start at all (consider that just reaching for a mushroom has its own cost of 10).

[4] This genetic algorithm has been used for all the simulations except the simulation on docility. The algorithm used in this last simulation is described in § 3.3.

After all the new individuals are created, we apply random mutations to the genome in the following way: each single weight of each individual has 0.1% probability to have its current value replaced by a new value randomly chosen in the range [– 4; + 4].

3. Simulations

3.1 Language does not evolve if it is advantageous to hearers but not to speakers

In the baseline simulation we try to biologically evolve language when in each trial the speaker is randomly selected from the population and, therefore, it tends not to share the same genes of the hearer.

We measure the quality of the emergent language in the following way. A particular linguistic signal produced by a particular organism is constituted by a vector of two continuous numbers in the range [– 1; + 1], i.e., the activation vector of the two linguistic output nodes of the organism. Hence, a signal can be considered as a point in a bi-dimensional space. Let's call *E* and *P*, respectively, the set of points ('cloud') which represent the signals produced by all the organisms of one generation in presence of all edible and poisonous mushrooms. A good language is one in which the two clouds *E* and *P* are (1) as small as possible in size (all the mushrooms belonging to the same category are named in similar ways), and (2) as distant as possible from one another (the mushrooms belonging to different categories are named in different ways)[5]. We calculate (1) by normalizing (in the range [0; 1]) the mean distance of the points of a cloud from its geometrical centre and taking its inverse (i.e. 1– x). Then, we measure (2) as the distance between the geometrical centres of the two clouds (also normalized in the range [0; 1]). Finally, the quality of the language produced in a generation is calculated as the mean of these two measures.

To give the reader an idea of what communication systems look like and of our language quality measure, in figure 3 we show three communication systems of different quality: the black circles represent the signals produced by all the individuals of one population in response to all edible mushrooms (i.e., cloud E) and the grey crosses represents the signals produced in response to poisonous mushrooms (i.e. cloud P). Figure 3a represents the communication system at the beginning of the simulation, when organisms have random connection weights: as can be seen, signals are completely random and one cannot tell the category of the mushroom which is signalled (language quality = 0.03). Figure 3b represents a communication system of low quality (0.33), in which the two clouds exist but they overlap significantly. Finally, figure 3c represents a very good communication system (language quality = 0.83) in which the two clouds are quite small and distant from each other.

[Insert figure 3 about here]

Figure 4a shows average fitness and language quality in a single replication of the simulation. Both the language quality and the average fitness are very unstable, fluctuating continually together between 0 and 1. How can this be explained? If we analyse the curves of the base-line simulation, we actually find a temporal pattern that can be described in the following way:

(1) fitness tends to stabilize at approximately 0.55;

(2) language quality tends to grow;

[5] These two properties guarantee both that (a) each individual produces similar signals for mushrooms that are members of the same category and different signals for members of different categories, and (b) language is shared across individuals in the population. On the contrary, if the two conditions are not satisfied, then we don't know whether this is because single individuals do not produce good languages or because there is no agreement between the (good) private signaling systems of different individuals. As we are interested in the language quality of the entire population, we can ignore this distinction and use the size of the two clouds and the distance between the two clouds as good indicators of overall language quality.

(3) after language quality has reached a certain level, fitness increases very rapidly until language quality suddenly drops down, and then fitness also drops down;

(4) finally, fitness tends to increase again and the cycle starts again.

[Insert figure 4 about here]

This pattern is present in every replication of the simulation, and this implies that the selective pressures that produce such results are reliable and strong. As it turns out, the fitness value of 0.55 is exactly the maximum value that can be reached without the help of language:

FmaxNoLang = (Fmax – 210 × 9 (steps necessary for checking the quality of poisonous mushrooms)) / 4200 = (4200 – 1890) / 4200 = 0.55

A possible interpretation of the pattern described above is the following. First, the organisms learn to discriminate between edible and poisonous mushrooms without listening to the signals received, where not listening to the signals means simply having very low weight values in the connections linking the linguistic input units with the hidden units. Then, for reasons that we will discuss in a moment, language quality increases so that the organisms start to rely on linguistic signals for not moving when a poisonous mushroom is present, and this leads to an increase in fitness. However, quite suddenly, language quality drops down and this causes the organisms' fitness to decrease because organisms are misled by the signals they receive. After they have learned not to listen to the signals, fitness re-starts to rise towards 0.5, and the cycle begins again.

The increase in language quality during the phase in which signals are ignored can be explained by both cognitive and genetic factors. Surviving and reproducing in the mushroom world implies that organisms must develop some ability to categorize the mushrooms appropriately. This means that all edible mushrooms tend to elicit similar activation patterns in the network's internal units which are different from the activation patterns elicited by poisonous mushrooms. Since the internal units are connected to the linguistic output units, these similar internal activation patterns tend to result in similar linguistic signals for all edible mushrooms and very different linguistic signals for all poisonous mushrooms. This cognitive explanation for the emergence of a good language (Cangelosi and Parisi 1998) can be supplemented by an explanation based on genetic convergence. In these simulations genetic variability tends to decrease because of selective reproduction which causes the long term survival of only a few lineages. Since a good language is a language which is the same in the entire population, increasing genetic similarity can result in better language quality.

The sudden drop in language quality in the next phase can be explained as a direct consequence of the altruistic character of the kind of communication we are dealing with here. Since selection favours only good understanding but not good speaking, bad speakers will be favoured against good ones because they can play the role of cheaters: bad speakers take advantage from the signals emitted by good ones while they emit signals that mislead other organisms. As a consequence, once a good communication system is in place, bad speakers will start to be selected against good ones and in a few generation this will result in a sudden decrease in overall language quality and in an accompanying drop in the average fitness of the population.

The outcome of these conflicting evolutionary pressures is that a good communication system never stabilizes and this does not permit the full exploitation of the adaptive value of communication. If we look at the average results of 10 replications of the base-line simulation (figure 4b), we see that language quality does not reach very high values and, as a consequence, the average fitness of the population fluctuates around a value (0.65) which is just a little higher than the maximum value that can be reached without communicating at all (0.55).

3.2 Kin selection allows language to evolve

If language quality decreases because of the altruistic character of speaking, then language quality should not decrease if we add to the base-line simulation some adaptive factor that makes it possible for altruistic behaviours to evolve. So, we tested whether adding kin selection to our simulation would result in the emergence of a good and stable communication system.

Kin selection has already been claimed to play a role in other language evolution models (Ackley and Littman 1994, Oliphant 1996), but this hypothesis has never been tested directly. In fact, previous work assumed kin selection was at work due to the spatialization of the organisms' interactions and reproduction (for a criticism of such a use of kin selection theory, see Di Paolo 1999). So, we decided to implement kin selection directly, by varying the probability that our organisms interact with close kin. The rationale for doing so is the following. Close kin tend to share the same genes; so, the higher the probability to speak to a close kin, the higher the chances that the altruistic behavior of producing a good language is preserved because it tends to confer an advantage to organisms with the same genes and, therefore, with the same altruistic behavior.

In our simulations 50 out of the 100 individuals in each generation are males and 50 are females. For each breeding cycle, one male and one females are selected for reproduction with the roulette selection algorithm and the pair generates two offspring, one male and one female, with different cross-over points and different mutations on connection weights. This breeding method makes sure that each organism of the population has at least one full sibling. So, we implemented kin selection by varying the probability p with which in any trial the speaker is one of the full siblings of the hearer (with probability $1 - p$ the speaker is not one of the hearer's full siblings, but can be a half sibling).

We run different simulations with different values of p. Figure 5 compares the results of the base-line simulation with those of the kin-simulations with $p = 0.01$, 0.05, 0.1 and 1, in terms of language quality (a) and average fitness (b). The results of the simulations are very clear: the higher the probability p of speaking to a close kin, the higher language quality and average fitness. Vice-versa, the lower the p, the lower the language quality and the average fitness of the population. With $p = 0.01$, the results are practically the same as those of the baseline simulation.

[Insert figure 5 about here]

The fact that the results of the simulation with $p = 0.1$ are not very different from those with $p = 1$ should cause no surprise. After all, p determines only the probability that the speaker is a full sibling of the hearer, but even when this is not the case, the speaker can share the same genes of the hearer either because it is a half sibling or simply because of the genetic convergence of the population discussed above. In fact, augmenting the 'kin selection factor' has the effect of adding an additional selective pressure for the stabilization of the language-production genes: in other words, kin selection simply strengthens a genetic pressure towards language emergence already present also in the base-line simulation.

3.3 Cultural evolution of language through docility

The preceding simulations have shown that the kind of communication system we are dealing with here is altruistic and, as such, it fails to emerge through genetic evolution unless under kin selection conditions. But one of the specific properties of human language with respect to animal communication systems is that human language is culturally, not genetically, transmitted. Could this property have in influence on the emergence of the kind of (altruistic) use of communication we are dealing here? In 1990 Herbert Simon proposed an original hypothesis on the evolution of altruistic behaviour in humans which was based on the notion of 'docility' (Simon 1990). Docility

refers to the propensity, characteristic of the human species, to learn socially (by imitation or by explicit teaching) how to behave. This propensity is biologically inherited in our species and it appears to be so strong that we can make the hypothesis that there has been a strong adaptive pressure during hominid evolution for the emergence of such a trait. (Consider the advantages that could derive from an innate predisposition to learn to imitate the behaviour of others in a context of highly social primates that are evolving the capacity of tool making and tool use, for instance). But for an individual to determine the contribution to its fitness of each behaviour which the individual is learning from others would be extremely hard, if not impossible. As a consequence, a docile organism will tend to learn whichever behaviour it will be taught, be it egoistic or altruistic, provided that the overall advantages of learning from others are higher than the disadvantages. According to Simon's docility theory, in a population of docile individuals an altruistic behaviour can evolve if the following three conditions are satisfied:

1) there is some advantage *d* in being disposed to learn from others, i.e., in being docile;
2) organisms are not able to evaluate the contribution of each particular behavior that they learn from others to their own fitness;
3) the advantage *d* of being docile is greater than the cost *c* of the altruistic behaviour.

As the use of language that we are dealing with in this paper is altruistic and as human language is culturally transmitted (although on a probable genetic species-specific basis), Simon's explanation of altruism seems to be applicable to language evolution.

So, we have run a new simulation, the 'docility simulation', in which the connection weights of all individuals are always random at birth and are not inherited from parents. Instead, the genome of these organisms is constituted by one only gene, encoded as an integer number, which specifies an individual's 'docility', i.e., the number of language learning trials for that particular individual. In the first generation each individual is assigned a random value in the interval [0; 200] for this gene and this value is genetically transmitted with a 2% probability of being changed by adding or subtracting a random number in the range [– 100; 100]. In any case, docility is forced to stay in the interval [0; 500].

The life of organisms in this second simulation is divided into two periods: infancy and adulthood. During infancy, the organism is supposed to follow its parent and learn from its parent how to behave in different situations. Its inherited docility gene determines the number of back-propagation learning cycles to which the infant exposes itself. Learning is imitative in that the teaching input of the back-propagation algorithm is the output of the infant's parent (Hutchins and Hazlehurst 1995, Denaro and Parisi 1997). Since there are three kinds of situations to which organisms are exposed during their life, there are three different learning conditions: (1) comprehension learning, (2) decision learning, and (3) naming learning. Comprehension learning takes place when the organism is distant from the mushroom and has to decide whether to move or not to move according only to the signal it receives from another organism. Decision learning takes place when the organism is near the mushroom and its decision whether to move or not to move into the mushroom cell and eat the mushroom depends on both the visual input from the mushroom and the linguistic input. Finally, naming learning takes place when the organism acts as a speaker: the organism receives the perceptual properties of a mushroom as input and it has to produce a linguistic signal.

In short, the number of learning cycles for each organism is determined by the organism's inherited docility and for each learning cycle this is what happens:

1) one of the three learning situations is randomly chosen together with one of the 420 mushrooms;
2) the appropriate input is given both to the learner and to its parent;
3) both the organism's output and its parent's output are calculated;

4) the output of the parent is given to the child as teaching input with some added random noise (a random value chosen in the interval +0.25/-0.25 is added to the teaching input[6]);
5) finally, the child's connection weights are changed according to the back-propagation algorithm (with a learning rate of 0.3 and a momentum of 0.8).

At the end of infancy an individual starts its adult life, which is identical to that of the genetic simulation.

[Insert figure 6 about here]

Figure 6 shows language quality, average fitness, and average value of the docility gene in this simulation. Since in this simulation organisms are born with random connection weights, they are bound to behave randomly unless they learn from their parents. The fact that the teachers of the organisms of any given generation are the selected organisms of the previous generation guarantees that what is learnt is a good behavior and the fact that some noise is always present in the cultural transmission of behavior guarantees that there is some added variability which is necessary for cultural evolution to take place. The results show that the average value of the docility gene rises very quickly until it reaches almost its maximum value (figure 6, thin line). Since organisms with a good foraging behavior tend to produce also a good language because of the cognitive pressure towards language emergence, young organisms will learn not only to discriminate edible from poisonous mushrooms, but also to produce a good language. And because of the fact that the behavior of these organisms is culturally learnt, the invasion of the population by cheaters is prevented: in fact, only individuals that are not very docile will produce a bad language, but those individuals cannot learn how to behave efficiently and therefore they will probably have a lower fitness than more docile (and altruistic) individual. As a result, to the increase in the average value of the docility gene corresponds a parallel increase in the quality of the language produced by the organisms, which reaches after about 200 generations the very high value of about 0.6 (figure 6, medium line). As it turns out, the correlation between docility and language quality is very high: 0,967. Finally, since the organisms of this simulation can exploit all the advantages provided by a good communication system, their average fitness reaches almost the maximum possible value (figure 6, thick line).

3.4 Language evolves if it is used to talk to oneself

In all the simulations described so far we have assumed that language is used for social communication purposes, that is, in situations in which the emitter and the receiver of signals are two different individuals. But language can have useful functions even in situations in which the emitter and the receiver of a signal are the same individual, i.e., when an individual talks to itself. The particular function that we will consider here is language as an aid for memory.

The initial situation is identical to that of the preceding simulations. An individual is placed in the start cell of the corridor, another randomly chosen individual is placed near the mushroom and it generates a signal which is heard by the first individual. On the basis of this signal the first individual can decide whether to approach the mushroom or refrain from doing so. However, since it takes 10 cycles to reach the mushroom, the signal must be available to the first individual in all these cycles in order for the individual to know what to do in each successive cycle. In the preceding simulations this problem was solved by assuming that the second individual, the speaker, continued to emit the signal until the end of the trial. In the present simulation the situation is

[6] The results are quite robust with respect to the quantity of noise added, provided that this quantity is adequate for cultural evolution: if there is no noise, there is no room for improvement in behavioural capacity; on the other hand, if there is too much noise, good behaviours cannot be preserved. All noise values between 0.1 and 0.4 produce the same qualitative results.

different. The emitter of the signal emits the signal only in the first cycle and then it goes away. All the first individual can do in this situation is to try to remember the signal by repeating the signal to itself until it reaches the mushroom. In the first cycle, when the signal arrives from the conspecific, the individual responds to the signal not only by either moving one cell forward toward the mushroom or avoiding doing so but also by producing a signal using its own linguistic output units. In the next cycles, the individual hears this self-produced signal and responds to it. Will a language evolve in these conditions?

Notice that in the baseline simulation language was very unstable because there were both cognitive and genetic pressure for its emergence and a strong selective pressure against language emergence due to its altruistic character. In the present simulation, in which individuals talk to themselves, they have an interest in producing good signals because sometime they are the receivers of the signals produced by themselves. Therefore emitters that produce good signals will tend to have more chances to reproduce than emitters of bad signals because the emitters of good signals can remember correctly the information received about the quality of the mushroom present at the end of the corridor. This effect is independent from kin-relatedness. (It might be interpreted as kin-relatedness in a single individual.) Hence, talking to oneself may constitute a selective pressure for the emergence of a good language even in populations in which language is exchanged between pairs of non-kin-related individuals.

This prediction is confirmed by the results of the talking-to-oneself simulation. Figure 7 compares average fitness and language quality in the baseline and the talking-to-oneself simulations. Language quality in the talking-to-oneself simulation is more stable and significantly higher than that of the baseline simulation: the range of fluctuation of language quality is [0.43; 0.57] and [0.25; 0.45], respectively. As a result, the average fitness of the population is higher and more stable in the talking-to-oneself condition than in the baseline condition: fluctuation range is [0.78; 0.95] and [0.5; 0.8], respectively.

[Insert figure 7 about here]

4. Discussion

For language to emerge biologically it must involve some selective advantage for both the emitters and the receivers of signals. Language requires both good speakers, i.e., individuals that emit the appropriate signals in the appropriate circumstances, and good hearers, i.e., individuals that respond appropriately to the signals that are produced by the speakers. But good speakers emerge only if it is advantageous to them to produce the appropriate signals in the appropriate circumstances, and good hearers emerge only if it is advantageous to them to respond appropriately to these signals. If linguistic signals provide the hearer with useful information, then the production of linguistic signals by the speaker is an altruistic behaviour provided that the behaviour with which the hearer responds to the linguistic signals have no useful consequences for the speaker and do not increase the speaker's reproductive chances. In fact, our simulations show that if it benefits the hearer but not the speaker language fails to emerge. Individuals that happen to be good speakers may increase the reproductive chances of hearers which benefit from the linguistic signals produced by the speakers but which, when it is their turn to function as speakers, may turn out not be good speakers. In this manner, although it would be generally useful, a stable shared language may never evolve.

Indeed, the results of our base-line simulation support this analysis and replicate the results of other simulations of the emergence of altruistic communication systems (Ackley and Littman 1994; Oliphant 1996; Marocco et al. 2003). However, Cangelosi and Parisi 1998 have described a model very similar to the one we adopt here in which a good communication system did evolve even without any benefit to the speaker. The explanation that those authors gave for their striking result

was that the emergence of a communication system was a by-product of the organisms' ability to categorize the environment. Since the organisms' fitness depends on their ability to tell poisonous from edible mushrooms, there is a strong selective pressure for the emergence of good representation abilities, that is, for having very similar internal representations (vectors of activation values in the internal layer neurons) for mushrooms belonging to the same category and very different representations for mushrooms belonging to the other category. This will result in the spontaneous emergence of a good language just because the activation pattern of the linguistic output units depends on the activation pattern of the internal units. However, this might be only part of the explanation. The relationship between language and categorization can only explain why the quality of the language produced by a single individual tends to be high, but it cannot explain why language is shared in the population. In order to explain why different individuals belonging to the same population share the same language something more is to be added: that is, the genetic similarity between individuals produced by genetic convergence of the population. Only the fact that all organisms share similar connection weights linking the perceptual input to the hidden layer and the hidden layer to the linguistic output can explain why the cognitive pressures towards the emergence of a good individual communication system (with similar signals for mushrooms belonging to the same category and different signals for mushrooms belonging to the other category) results in the spontaneous emergence of a shared language in the whole population. If this analysis is correct, than the difference between the results of our simulation and that of Cangelosi and Parisi 1998 might be explained by a difference in the genetic convergence of the population due to differences in the simulation set-up such as the different difficulty of the task or the different genetic algorithm used. Further analyses are needed on this topic but some preliminary results of such analyses seem to support this interpretation.

Apart from the role played by the structural coupling between categorization abilities and linguistic behaviour, three distinct mechanisms may have made it possible for language to emerge even if the production of linguistic signals is altruistic behavior which benefits the hearer but not the speaker.

One mechanism is kin selection. Our simulations show that, in accordance with kin selection theory (Hamilton 1964), the probability that the speaker and the hearer share the same genes directly affects the stability of a communication system which benefits the hearer but not the speaker. Animal signals mostly communicate information about the sender of the signal - its location, species, identity, emotional state, intentions and attitudes -, not information about the external environment (Hauser 1996). Surely, there are some exceptions like honey bees dance and vervet monkeys alarm calls, but human language clearly is the most sophisticated communication system for communicating information about the external environment. While kin selection was likely one of the major factors in the evolution of social insects' communication systems, we don't know what could have been the role of kin selection in the evolution of primate (and human) communication. Hominid evolution took place in small, kin-related groups, and long juvenile period of humans may have contributed to a substantial increase in the probability that communicative interactions took place between related individuals. As our kin selection simulations demonstrate, if this is the case, then the increase in genetic relatedness between the speaker and the hearer might have been one factor, among many others, which favoured the evolution of (the altruistic use of) language.

However, if language is restricted to kin-related groups its usefulness appears to be limited. Language becomes much more useful if it represents a vehicle of communication and interaction among larger groups of genetically unrelated individuals. Language that benefits the hearer but not the speaker may emerge in larger groups of unrelated individuals if it is culturally rather than biologically transmitted. Cultural transmission is learning from others. Human beings appear to have a genetically inherited tendency to learn from others whatever behavior others may care to teach them. This 'docility' evolves because it confers an advantage to the individual that possesses

it: a docile individual can learn from others behaviours which would be more difficult to learn by directly interacting with the non-social world. These behaviours generally tend to increase the reproductive chances of the learning individual but in some circumstances and in some of their uses they may benefit others. But for docile organisms, learning from others is 'blind', in the sense that the learning individual does not (easily) distinguish between what is beneficial for himself or herself and what is beneficial for others, especially because this kind of learning takes place mostly when the individual is young. Thus, altruistic behaviours, including the altruistic use of language we are dealing with here, can emerge as a cultural by-product of the genetically evolving docility of a population.

Studying the influence of cultural learning on language emergence is one of the major themes in the literature on computational models of language evolution (see, for example, Hutchins and Hazlehurst 1995; Batali 1998; Hurford 1999; Kirby 2000; Steels and Kaplan 2002). But none of the previous computational models of language evolution tested the hypothesis that the evolution of cultural transmission could have favoured the emergence of an altruistic use of language. Our docility simulation shows exactly this. It shows that linguistic docility, that is, a tendency to learn language from others, can emerge biologically and be biologically transmitted, and, if this is the case, the homogenization of linguistic behaviour induced by cultural transmission can favour the (cultural) emergence of an altruistic use of language. (For a general discussion of our docility simulation in the context of the biological hypotheses on the evolution of altruism, see Mirolli and Parisi 2004).

Finally, a third mechanism may explain the emergence of a language that benefits the hearer and not the speaker in groups of genetically unrelated individuals even if the language is biologically and not culturally transmitted and evolved. This third mechanism is using the language not (only) to communicate with others but (also) to communicate with oneself. Generally, there is a tendency to think that language was used by humans to communicate with oneself only when language was already well developed and was sophisticated and complex; hence, quite recently compared with the first appearance of a proto-language. However, this is not necessarily the case. Even a very simple proto-language, for example, a language made up of single words (or holophrases), may be used to talk to oneself, for example as an aid for memory, with advantages for the individual that uses the language in this way.

Based on this hypothesis, our talking-to-oneself simulation shows that linguistic signals that benefit the hearer but not the speaker can emerge biologically among genetically unrelated individuals if the hearer has to repeat the signals to himself or herself in order to keep them in memory. When one speaks to another individual and the signals benefit the hearer but not the speaker, the speaker acts altruistically and may benefit a hearer who is a poor speaker. This, as we have seen, is an obstacle for the biological emergence of language. But if the hearer has to repeat the signals to himself or herself in order to keep them in memory, then there is a positive selective pressure towards good speaking abilities because in talking-to-oneself the same individual is both speaker and hearer, and hence the advantage of understanding linguistic signals can only be exploited if it is accompanied with good productive capacities.

We think that linguistic memory can be advantageous for at least two reasons: (a) delegating the memory function to the linguistic system can leave the sensory-motor system free to process other information useful for acting in the environment while linguistically remembering previous information, and (b) linguistic signals may occupy less space in memory than the raw information they refer to[7]. It might also be that another peculiar characteristics of human language, namely the

[7] Recent neuro-psychological evidence seems to suggests that humans have indeed (at least) two different working memory systems: the first one is non-linguistic and it is shared with other primates; the second one seems to involve linguistic areas. Furthermore, these two systems seem to have not only different neuro-anatomical organization and different evolutionary origins, but also quite different functional properties (Gruber and Goschke 2004)

fact that it uses displaced signals (Hockett 1960), has become possible only after hominids had improved their memory by talking to themselves. (For the possible role of linguistic memory in the evolutionary origin of language, see Aboitiz and Garcìa 1997.) In any case, we think it would be worthwhile to investigate those issues using a more ecological and embodied simulative framework.

Furthermore, the use of the linguistic system as an aid to memory is only an example of the many possible individual uses of the social communication system (Vygotsky 1962, Jackendoff 1996, Clark 1998). For example, in other simulations we showed how learning a simple language can improve the ability to categorize the world (Mirolli and Parisi in press). In general, we think that the use of language as an aid to cognition can have played an important role in language evolution. Our talking-to-oneself simulation is just a very simple demonstration that this could be the case. Much more work (both empirical and computational) is to be done in order to better understand the intimate relationship between language and cognition and the phylogenetic and ontogenetic co-evolution between the two.

5. Conclusion

Human language is an incredibly complicated phenomenon which depends on the individual brain, on ontogenetic development, and on genetic and glossogenetic (cultural) evolution. All those systems are complex systems, acting at different time scales, which in turn interact with one another in complex ways. So, it is very unlikely that a single simple explanation can be found for the emergence of human language. Indeed, it is very unlikely that a single discipline or a single methodology can suffice for understanding all the amazing features of human language or the mechanisms and factors that generated them.

In the simulations described in the present paper we singled out one particular feature of human language, namely, the fact that it is a particularly sophisticated system for informing other individuals about the external environment, and we tried to analyse some of the factors that could have played a role in the evolutionary emergence of this feature: its altruistic character, the social scope of its use which could result in kin selection, the fact that language is culturally transmitted, and the fact that human language is not only a communication system, but is also used for improving individual, cognitive functions such as memory. A lot of more work needs to be done in order to increase our understanding of the emergence of all the important features of human language.

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

Aboitiz, F. and Garcìa, R.V., 1997, The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. **Brain Research Review**, **25**, 381-396.

Ackley, D. H. and Littman, M. L., 1994, Altruism in the evolution of communication. In *Artificial Life IV: Proceedings of the International Workshop on the Synthesis and Simulation of Living Systems*, edited by R.A. Brooks and P. Maes (Cambridge, Mass.: MIT Press), 40-48.

Batali, J., 1998, Computational simulations of the emergence of grammar. In *Approaches to the evolution of language*, edited by J.R. Hurford, M. Studdert-Kennedy and C. Knight (Cambridge: Cambridge University Press), 405-426.

Berk, L., 1994, Why children talk to themselves. **Scientific American**, November, 78-83.

Bickerton, D., 2002, Foraging versus social intelligence in the evolution of protolanguage. In *The transition to language*, edited by A. Wray (Oxford: Oxford University Press), 207-226.

Billard, A. and Dautenhahn, K., 1999, Experiments in learning by imitation: Grounding and use of communication in robotic agents. **Adaptive Behavior**, **7(3/4)**, 415-438.

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

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

Cangelosi, A. and Parisi, D., 1998, The emergence of a 'language' in an evolving population of neural networks. **Connection Science**, **10(2)**, 83-97.

Cangelosi, A. and Parisi, D. (eds.), 2002, *Simulating the evolution of language* (London: Springer-Verlag).

Christiansen, M.H. and Kirby, S. (eds.), 2003, *Language evolution: The states of the art* (Oxford: Oxford University Press).

Clark, A., 1998, Magic words: How language augments human computation. In *Language and thought: Interdisciplinary themes*, edited by P. Carruthers and J. Boucher (Cambridge: Cambridge University Press), 162-183

Denaro D. and Parisi D., 1997, Cultural evolution in a population of neural networks. In *Neural Nets. Wirn-96*, edited by M. Marinaro and R. Tagliaferri (New York: Springer Verlag), 100-111.

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

Di Paolo, E.A. 1999, A little more than kind and less than kin: the unwarranted use of kin selection in spatial models of communication. In *Advances in Artificial life. Proceedings of the 5th European Conference*, edited by D. Floreano, J.D. Nicoud, and F. Mondada (Lausanne: Springer-Verlag), 504-513.

Elman, L.J., 1990, Finding structure in time. **Cognitive Science**, **14,** 179-211

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

Hamilton, W. D., 1964, Genetic evolution of social behavior. **Journal of Theoretical Biology**, **7(1)**, 1-52

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

Hauser, M.D., 1996, *The Evolution of Communication* (Cambridge, Mass.: MIT Press).

Hazlehurst, B. and Hutchins, E., 1998, The emergence of propositions from the co-ordination of talk and action in a shared world. **Langauge and Cognitive Processes**, **13(2/3)**, 373-424.

Hocket, C.F., 1960, The origin of speech. **Scientific American**, **203(3)**, 89-96

Hurford, J.R. 1999, Expression-Induction models of language evolution: Dimensions and issues. In *Linguistic evolution through language acquisition*, edited by T. Briscoe (Cambridge: Cambridge University Press), 301-344

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

Jackendoff, R., 1996, How language help us think. **Pragmatics and Cognition**, **4**, 1-34.

Kirby, S., 2000, Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight et al. (eds.) 2000, 303-323

Knight, C., Studdert-Kennedy, M. and Hurford, J. (eds.), 2000, *The evolutionary emergence of language: social function and the origins of linguistic form* (Cambridge: Cambridge University Press).

Marocco, D., Cangelosi, A. and Nolfi, S., 2003, The emergence of communication in evolutionary robots. **Philosophical Transactions: Mathematical, Physical and Engineering Sciences**, **361(1811)**, 2397-2421.

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

Mirolli, M. and Parisi, D., in press, Langauge as an aid to categorization: A neural network model of early language acquisition. In *Modeling language, cognition and action: Proceedings of the 9th Neural Computation and Psychology Workshop*, edited by A. Cangelosi, G. Bugmann and R. Borisyuk (Singapore: World Scientific).

Mitchell, M., 1996, *Introduction to Genetic Algorithms* (Cambridge, Mass.: MIT Press).

Noble, J., Di Paolo, E.A. and Bullock, S., 2001, Adaptive factors in the evolution of signalling Systems. In *Simulating the evolution of language*, edited by A. Cangelosi and D. Parisi (London: Springer-Verlag), 53-78.

Oliphant, M., 1996, The dilemma of Saussurean communication. **Biosystems**, **37(1-2)**, 31-38.

Parisi, D., 1997, An Artificial Life approach to language. **Brain and Language**, **59**, 121-146.

Pedone, R. and Parisi, D., 1997, In what kinds of social groups can altruistic behavior evolve?. In *Simulating social phenomena*, edited by R. Conte, R. Hegselmann and P. Terno (Berlin: Springer-Verlang), 195-201.

Simon, H.A., 1990, A mechanism for social selection and successful altruism. **Science**, **250**, 1665-1668.

Steels, L., 1998, The origins of syntax in visually grounded robotic agents. **Artificial Intelligence**, **103(1-2)**, 133-156.

Steels, L. and Kaplan, F., 2002, Bootstrapping grounded word semantics. In *Linguistic evolution through language acquisition: Formal and computational models*, edited by T. Briscoe (Cambridge: Cambridge University Press), 53-73.

Vygotsky, L.S., 1962, *Thought and language* (Cambridge, Mass.: MIT Press) (Original work published in 1934).

Wagner, K., Reggia, J.A., Uriagereka, J. and Wilkinson, G.S., 2003, Progress in the simulation of emergent communication and language. **Adaptive Behavior**, **11(1)**, 37-69.

Zahavi, A., 1975, Mate selection. A selection for a handicap. **Journal of Theoretical Biology**, **53,** 205-214.

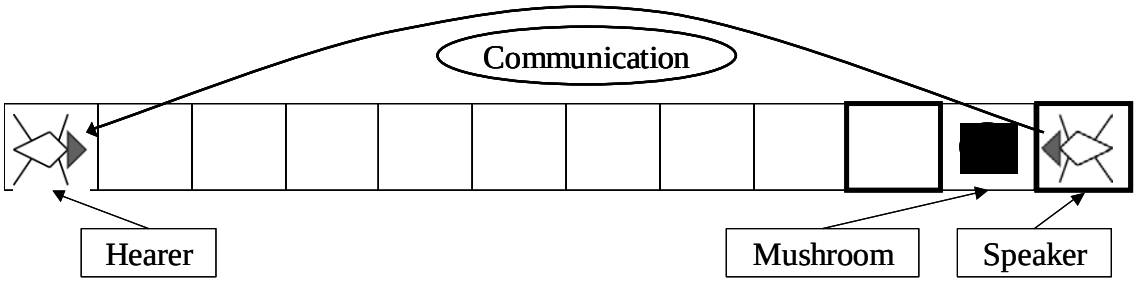

Figure 1. The environment at the beginning of each trial. The tested organism (hearer) is placed at the beginning of the corridor and a mushroom is placed at the end. Another organism (speaker) is placed near to the mushroom and it sends a signal (linguistic output) to the hearer. The perceptual properties of the mushroom are perceived only when the organism is in the cell represented with thick lines. If an organism enters the mushroom cell, it eats the mushroom and the trial ends.

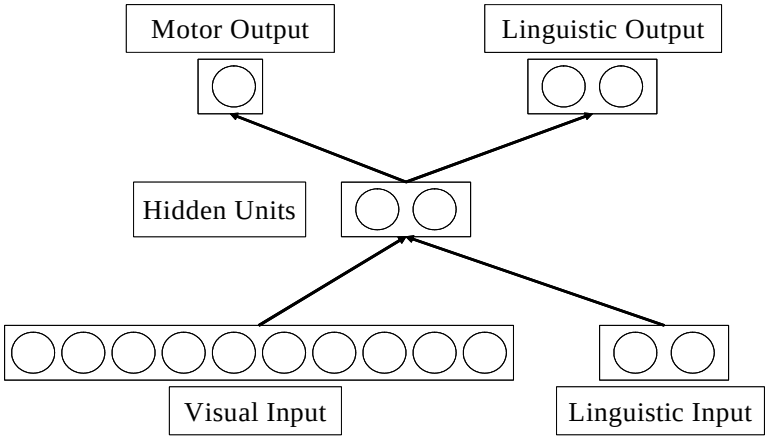

Figure 2: The neural network.

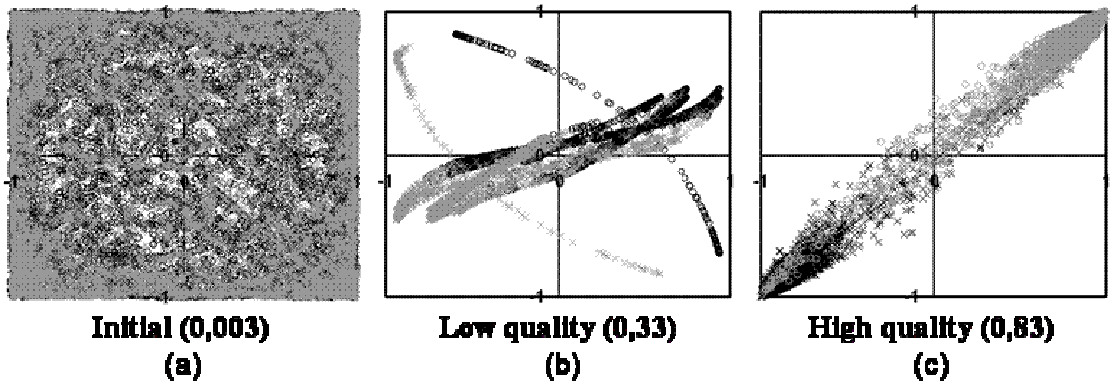

Figure 3: Communication systems of very low (a), medium (b) and high (c) quality. Black circles represent signals produced for edible mushrooms (i.e. cloud *E*) and grey crosses represent signals produced for poisonous mushrooms (i.e. cloud *P*). See text for details.

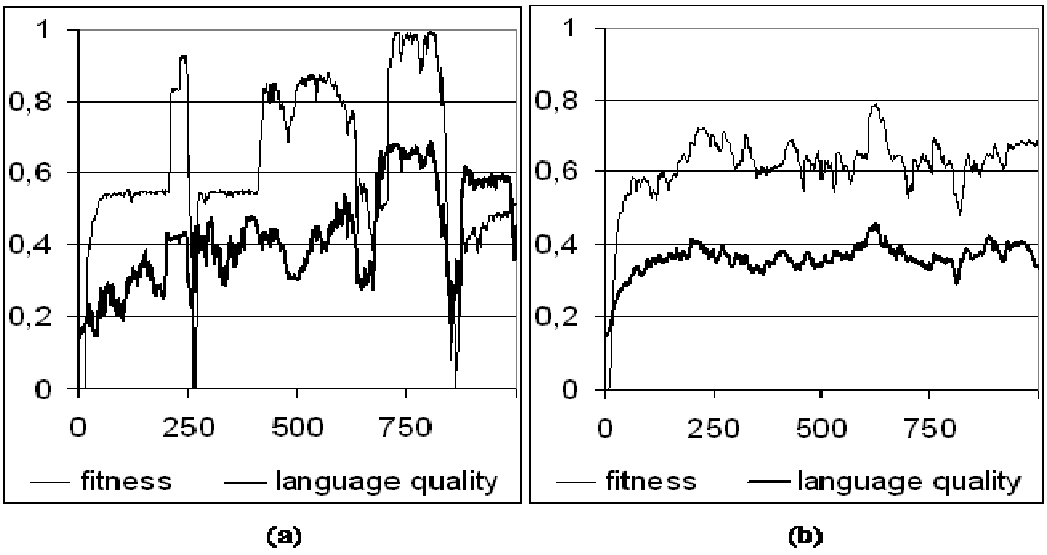

Figure 4: Average fitness and language quality of a single replication (a) and of 10 replications (b) of the base-line simulation.

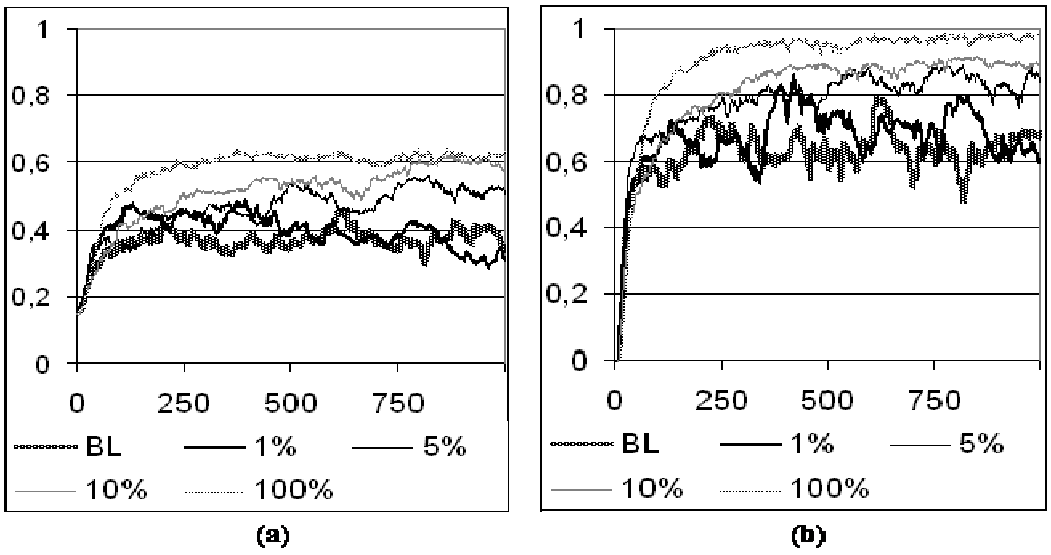

Figure 5: Language quality (a) and average fitness (b) of the base-line and 1%, 5%, 10% and 100% kin simulations. Average results of 10 replications.

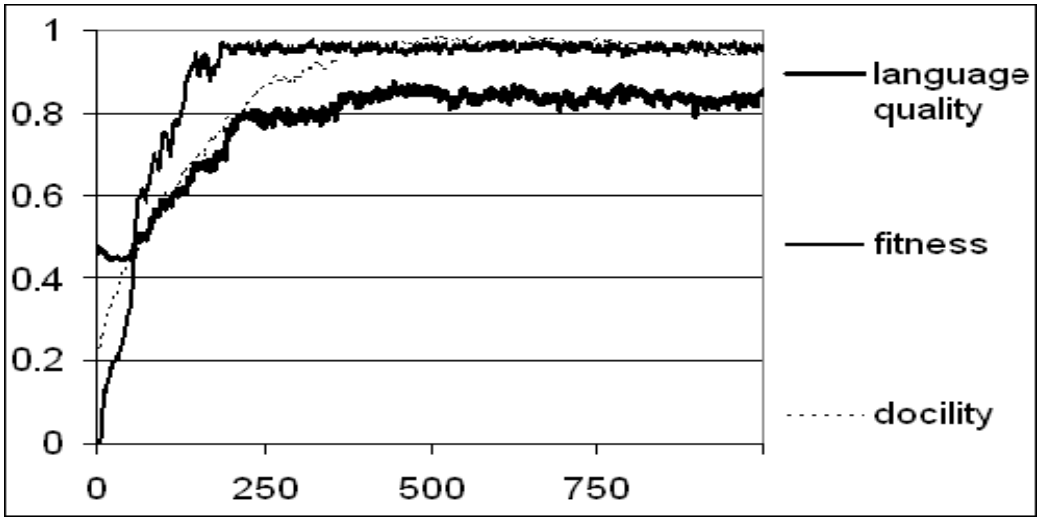

Figure 6: Language quality, average fitness and average value of the docility gene of the docility simulation. Average results of 10 replications of the simulation.

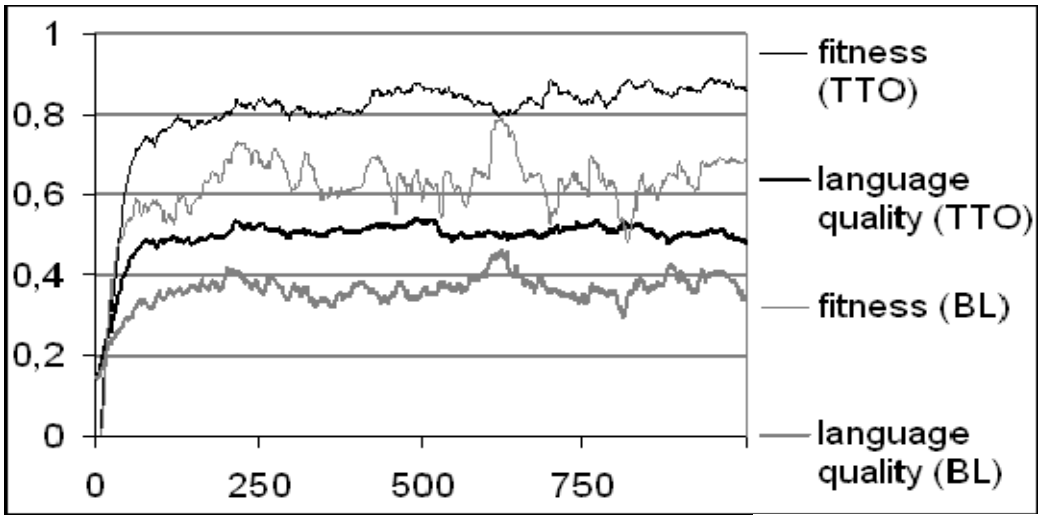

Figure 7: Average fitness and language quality of the talking-to-oneself (TTO) simulations and of the base-line (BL). Average results of 10 replications of both simulations.