Modeling the evolution of communication: From stimulus associations to grounded symbolic associations

Angelo Cangelosi

Centre for Neural and Adaptive Systems
University of Plymouth
Plymouth PL4 8AA (UK)
angelo@soc.plym.ac.uk

Abstract. This paper describes a model for the evolution of communication systems using simple syntactic rules, such as word combinations. It also focuses on the distinction between simple word-object associations and symbolic relationships. The simulation method combines the use of neural networks and genetic algorithms. The behavioral task is influenced by Savage-Rumbaugh & Rumbaugh's (1978) ape language experiments. The results show that languages that use combination of words (e.g. "verb-object" rule) can emerge by auto-organization and cultural transmission. Neural networks are tested to see if evolved languages are based on symbol acquisition. The implications of this model for Deacon's (1997) hypothesis on the role of symbolic acquisition for the origin of language are discussed.

1. Symbol acquisition in the evolution of communication

The synthetic approach of Artificial Life has recently been applied to studying the evolution of communication and language (Steels, 1997a). Some models have been used for the simulation of the emergence of simple lexicons in populations of simulated organisms (e.g. Cangelosi & Parisi, 1998; Steels, 1997b) or in small communities of robots (Steel & Vogt, 1997). In these studies organisms evolve shared lexicons for describing entities and relations of the environment. Other models have focused on the evolution of syntax (e.g. Batali, 1994; Kirby, in press). Simulated organisms evolve different syntactic languages starting from a given set of syntactic structures and constraints, and devices for syntax acquisition.

The first type of models, that focus on lexicon emergence, do not make any explicit reference to the role of syntax in language origin. Their aim is to model the early stages of the evolution of (animal) communication. Indeed, in animal communication systems, no syntactic structures have been observed. For example, no animal communication systems have been found that share one of the main properties of human languages, i.e. the combination of words to express different and new meanings. These models of lexicon evolution study communication systems based on simple signal-object associations. Organisms learn and evolve simple stimulus associations between objects in the environment and signals.

In the second type of models the evolution of syntax is simulated. These models, that for example show how syntax can emerge without natural selection (Kirkby, in press), do not explain the possible role of syntactic languages for organisms' adaptation and survival. Moreover, the associations that organisms learn are self-referential symbol-symbol relationships. These models are subject to the symbol grounding problem (Harnad, 1990) since they lack an intrinsic link between their symbols and the entities and relations existing in the organisms' environment. Internal symbols need some form of sensorimotor grounding. Due to the symbol grounding problem, the role of these models for understanding the evolution of cognition is reduced. The Artificial Life methodology used in this model, instead, allows to overcome the symbol grounding problem. Simulated organisms will use symbols whose semantic referents are constituted by categorical representations in the neural network's hidden layer. These semantic representations are activated by the actual presence of their referents in the organism's world.

Recently, Terrence Deacon (1997) proposed an explanation for the fact that animal communication and human language differ. A variety of animal communication systems have been studied (Hauser, 1996), however, there is no apparent continuity between animal communication systems and complex human languages. That is, no "simple languages", using some elementary forms of word combinations or syntax, have been found in the animal kingdom. The existence of simple languages could explain the gap between animal and human communication. Deacon (1997; 1996) believes that this is due to the symbol acquisition problem. In fact the main difference between animals and humans relies on symbolic references. There is a significant difference between the referencing system of simple object-word associations and that of symbolic associations. In animals, simple associations between world entities and words can be explained by mere mechanisms of rote learning and conditional learning. An animal acquires genetically, or learns, that a word's sound is always associated with a specific object. Instead, symbolic associations have double references, one between the word (symbol) and the object, and the second between the symbol itself and other symbols. A language-speaking human knows that a word refers to an object and also that the same word has logical (syntactic) relation with other words. Due to the possible combinatorial interrelationships between words, there can be an exponential growth of reference with each new added word.

The difference between these types of associations, and their relation to the models of language origin, is graphically represented in Figure 1. Figure 1a represents a communication system based on simple associations between objects and words. It refers to the models of the acquisition of lexicons. Figure 1b represents the models of the origin of syntax. It only shows word-word associations, but this system is not based on real symbolic associations as a link is missing between words and objects. Words are self-referencing and they lack a grounding in the external world. Figure 1c shows a system based on grounded symbolic associations. The arrows represent references between words, and references between words and objects. This third type of association system can be simulated through Artificial Life methodology, upon which the present model is based.

It should be noted that the relationship between words and objects, that constitute the grounding of symbols to entities of the real world, is not a direct link between mental symbols and real objects. Instead, it is a link between mental entities (the symbols or words) and other mental entities (such as concepts) that constitute the semantic reference. These categorical representations, that Deacon (1996) calls "indexical", are useful to "sort out" the extensive perceptual variability of objects in the real worlds. The ability of humans and animals to create categories, e.g. through categorical perception, constitutes the "groundwork" of cognition (Harnad, 1987). From this it is possible to build more complex cognitive skills, such as language.

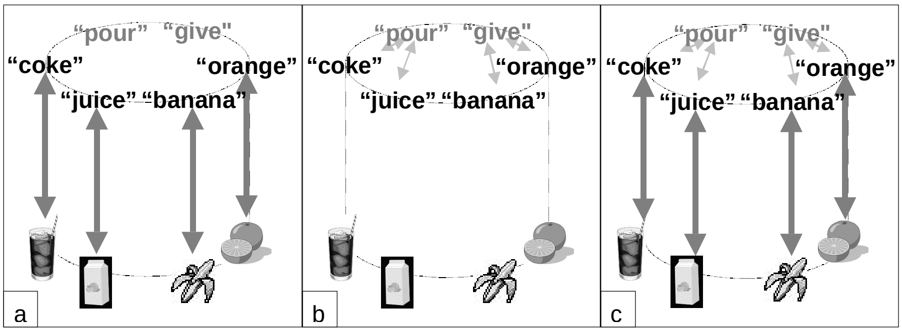

Fig. 1. Associations between objects (pictures) and symbols (words) in language origin models. (a) Simple stimulus associations between objects and words in the models of the origin of lexicons. (b) Self-referential associations between words that lack sensory-motor grounding in the models of the origin of syntax. (c) Grounded symbolic associations. Words have links with objects and logical relationships between themselves. Objects and words were chosen from Savage-Rumbaugh & Rumbaugh's (1978) experiments on ape language.

Deacon's hypothesis on the role of symbolic learning in the evolution of human language is supported by ape language studies and by neuropsychological and neurophysiolgical evidence (Deacon, 1997). For example, experiments on language acquisition in chimpanzees have been used to support the idea that animals tend to learn language using simple word-object associations. However, apes can be taught real symbolic associations under special experimental conditions (Savage-Rumbaugh & Rumbaugh, 1978). Moreover, in these language-speaking animals the spontaneous use of the grammatical rule "verb-object" has also been observed (Greenfield & Savage-Rumbaugh, 1990).

This paper aims to test a model of the origin of communication and language that deals with the evolution of different types of associations. The model should be able to study how different word/object relationships can evolve and also to define the mechanisms that explain the passage from communication based on simple stimulus association to languages based on grounded symbolic references. For this reason languages based on two-word combinations will be evolved. The behavioral task is influenced by ape language studies (Greenfield & Savage-Rumbaugh, 1990).

2. Method

The simulation method combines the use of artificial neural networks and genetic algorithms. It uses the methodology and theoretical framework of Ecological Neural

Networks (ECONET: Parisi, Cecconi & Nolfi, 1990). Populations of organisms are evolved according to their behavioral performance in foraging tasks. Organisms' behavior is controlled by neural networks.

In the present simulation, the environment setting for the foraging task consists of a 2D grid of 100x100 cells. About 1200 cells are occupied by randomly placed foods (mushrooms). The foods are grouped into two main functional categories: edible mushrooms (E), i.e. foods that need to be collected to increase organisms' fitness, and toadstools (T), i.e. mushrooms that must be avoided. The first category of edible mushrooms is then split into three functional subcategories: white (e1), yellow (e2), and gray (e3). These are called functional categories because they require organisms to perform a different task when approached (e.g. white mushrooms e1 should be picked and cut, whilst other colored mushrooms require different actions). The fitness formula adds one point for each e1/e2/e3 mushroom that an organism approaches <u>and</u> properly treats according to its color. When a toadstool is collected, the fitness is decreased by one point. The toadstool category does not have any functional subcategory. Even though toadstools are perceptually classifiable into three categories (t1, t2, t3), these are not functional categories because the fitness formula removes one point for each toadstool that the organism reaches, regardless of their appearance.

The organization of the foraging task stimuli into a hierarchy of functional categories was derived from the experimental setting of ape language studies. For example in Savage-Rumbaugh & Rumbaugh (1978) chimpanzees had to learn to use different lexigrams (graphic symbols in a keypad) to name solid foods (e.g. banana, orange) and drinks (coke, milk). Since they receive food from a vending-machine, they also need to learn a lexigram for the verb associated to solid foods ("give") and that for the liquid drinks ("pour"). These stimuli constitute a hierarchy of two high-level functional categories (verbs) followed by four low-level categories (two foods and two drinks). In our model organisms will have to learn a name for each of the three edible subcategories (e.g. "white" for e1, "yellow" for e2, and "gray" for e3), plus a common verb for the whole edible category, e.g. "approach". All toadstools will require the use of a common verb, such as "avoid". The three toadstool subcategories do not require a specific name to be identified, but organisms will be allowed to name them.

The neural networks controlling the organisms' behavior have a 3-layer feedforward architecture (see figure 2). The input layer has 29 units, organized into three groups of sensory units. In the first group there are 3 units, one for each of the 40-degree neighboring visual field. The unit corresponding to the visual field in which the closest mushroom is perceived will be activated. Its activation value id the distance of this mushroom (range 0-1). The second group of input units has 18 nodes that encode some (visual) features of mushrooms. In fact, each mushroom has a set of 18 binary features. The mushrooms of each subcategory share a common binary pattern. For example e1 mushrooms share the prototype 111*************** and e2 have ***111************. An asterix (*) represents random bits. The third group of input nodes are localist language units. Each unit is activated whenever the corresponding word is used. The hidden layer has 5 units. The output layer has two groups of unit. The first 3 units encode the actions. Two binary bits control the movement (move one cell forward, turn left, turn right, stand still) and one unit for discriminating between e1, e2, and e3 (activation $a<.2$, $.2\geq a>.8$, $a\geq.8$ respectively). The second group has 8 linguist units. These units are organized into two clusters of winner-takes-all units. Only two words at a time will be active (one per cluster). One cluster has 2 units, the second 6 units. The hidden and output units use the sigmoid activation function.

Evolution is organized into two sequential stages. The first stage takes 300 generations and organisms do not communicate at all. They only use the mushroom position and feature information to evolve the proper foraging action. The population consists of 80 individuals. Organisms live in the same environment for 1000 actions (20 epochs of 50 moves each). The 20 organisms with the highest fitness level are selected and each reproduce making 5 offspring. The new organisms' genotypes, i.e. the set of connection weights encoded as real numbers, are then mutated by slightly modifying 10% of the weights.

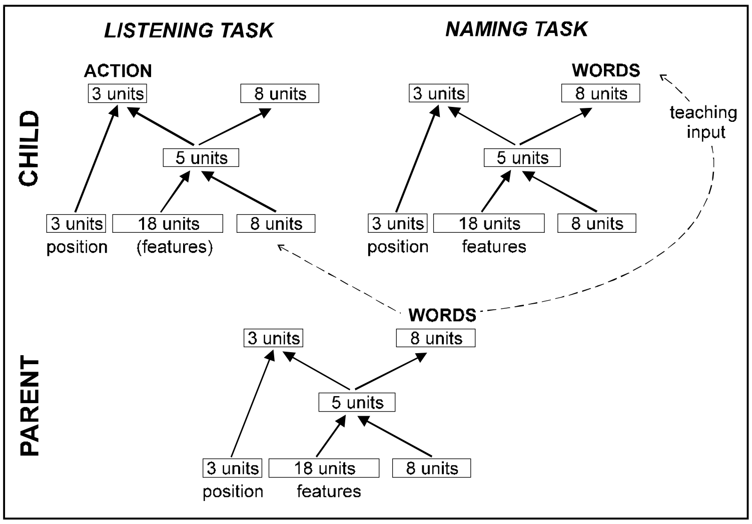

Fig. 2. Neural network architecture and the interaction between the child organism and its parent. The parent uses words to describe the closest mushroom. In the Listening Task the child uses these words to decide which action to take. In the following Naming Task the child uses the parent's words as teaching input for error backpropagation.

In the second stage of evolution, starting at generation 301, communication between organisms is allowed. The 20 parent organisms are kept together with the 80 new siblings. The parent organisms work only as speakers and language teachers. They cannot eat mushrooms and do not replicate. The need for two-stage simulation, in which only the foraging behavior is evolved at first, and subsequently communication is introduced, was suggested by a previous study (Parisi, Denaro & Cangelosi, 1998) in which the simultaneous evolution of foraging and language proved difficult. During each time interval, child organisms perform two actions (figure 2). The first is a Listening Task. To discriminate the type of mushroom children use the words suggested by their parents as input. In fact most of the time children rely solely on the parents' linguistic input because they only perceive the mushroom’s features 10% of the time. After the network activation cycle of the Listening Task, children perform a Naming Task. They use the mushroom’s 18-bit features to name the food type. An error backpropagation algorithm is then applied. The error is computed using the parent’s words as teaching input, so that children learn the same linguistic description given by their parents. Some noise is added to the error between the child’s linguistic output and the parents’ teaching input. This is to allow variability in the process of cultural transmission of language (Parisi *et al.*, 1998). The same backpropagation algorithm is used for an Imitation task, where the organism’s neural network learns the auto-association of the input-output linguistic stimulus. Organisms live for 2000 time steps, a longer lifetime than in the first stage as backpropagation learning requires more stimulus presentations. When the select organisms reproduce, their new offspring inherit the parent's connection weights before backpropagation learning occurred. No Lamarkian inheritance of learned weights is allowed. Darwinian selection will continue evolving the ability to approach/avoid mushrooms. The language learning between parents and offspring permits cultural transmission between consecutive generations.

During the second stage of evolution the interaction between parents and children can result in the emergence and auto-organization of a shared language. As they are only allowed to perceive the mushroom’s features 10% of the time, it should facilitate the evolution of a good language that discriminates at least the functional categories `T`, `e1`, `e2`, `e3`. Fitness depends on the correct identification of these four categories.

3. Results

The first stage of evolution, which does not permit communication, was repeated 10 times using different random populations. Nine of these replications resulted in an optimal classification behavior. Organisms evolved the ability to approach edible mushrooms `E` and avoid all toadstools `T`. Moreover, according to the type of edible mushrooms `e1`, `e2`, `e3`, they produced the correct activation in the third node of the output action units. In the sole population where the evolved behavior was poor, organisms were unable to discriminate between `e2` and `e3`. The average fitness for the 9 successful populations is shown in Figure 3 (first 300 generations). At generation 300 the best individual of each population on average collects 90 edible mushrooms (i.e. 4.5 mushrooms per each of the 20 epochs), and avoid all toadstools.

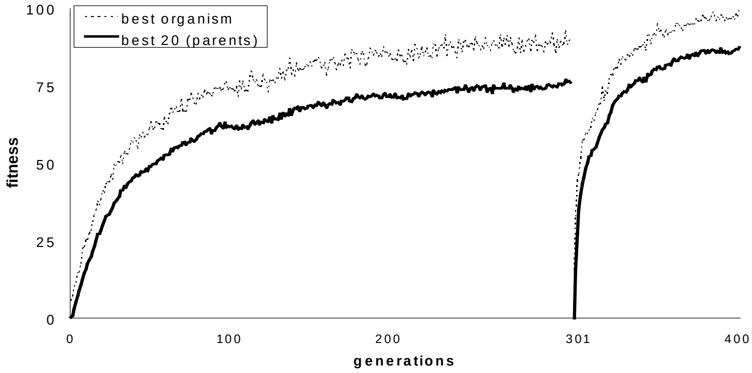

Fig. 3. Fitness of the best individuals and of the groups of 20 selected parents. Simulation without communication (generations 1-300) and during the evolution of communication (generations 301-400). The values of generations 1-300 are averaged over 9 successful simulations. The fitness of generations 301-400 is averaged over 11 populations.

For the second stage of evolution only the nine successful populations were used. The only simulation with unsuccessful fitness growth was not used because language evolution requires a preliminary ability to discriminate behavioral categories (Parisi *et al*, 1998). This stage took 100 generations. For each population, two random starting conditions were executed. In total, 18 replications were performed.

The results of the distribution of evolved languages are shown in Table 1. In 11 of the 18 runs, populations evolved good languages, i.e. the use of at least four words/word-combinations to distinguish the four behavioral categories `T`, `e1`, `e2`, `e3`. These languages emerged through a process of auto-organization of the lexicon, due to the interaction between organisms and the process of cultural transmission. The average fitness for the 11 successful populations is shown in figure 3 (generations 301-400). In the remaining 7 populations the emerged language was poor. That is, some mushroom types were incorrectly labeled due to the lack of a specific symbol, or symbol combination. Therefore the fitness is very low since these mushrooms were incorrectly described and collected.

	Single word	Word combination	Verb-object	TOTAL
Good languages	1 (9%)	3 (27%)	7 (64%)	11
Imperfect languages	1 (14%)	2 (29%)	4 (57%)	7

Table 1. Distribution of language types in the simulations for the evolution of communication.

In the previous section we explained that the linguistic output units are organized into two winner-takes-all clusters. The first cluster is made up of 6 linguistic units (words), and the second has 2 units. The cluster-based structure does not imply that a combination of two words is always necessary to describe a mushroom. In fact the optimal behavior requires the production of only four actions, and therefore four words from the first cluster are enough to name these categories. This is what happened for one of the good language populations. Here organisms used four words of the first cluster to name the four categories T, e1, e2, e3. The two words of the second clusters were not systematically associated to any mushroom.

When both clusters are used, there are several possibilities of combining words. However, we are interested in identifying word-combination rules that resemble known syntactical structures. In particular we want to establish if a verb-object rule has emerged. Considering the populations where good communication evolved, ten (91%) evolved languages that use combinations of symbols. Among these, three populations (27%) use various combinations of two words, and seven (64%) use verb-object rules. The way we can identify the verb-object rule is because in the two-word cluster each linguistic unit is systematically associated only to one of the high-order categories T and E. One "verb" symbol is always used for all toadstools ("avoid") and the other for all edible mushrooms ("approach"). The units in the 6-word cluster are used for distinguishing single "objects" (mushroom types) with which the two verbs systematically couple.

4. Discussion

The aim of this research was to develop a model of the evolution of communication systems based on simple syntactic rules, such as word combination. Moreover we were interested in establishing whether the evolved word-object relationships were based on symbolic learning or mere object-word associations. The resultant description shows that languages that use combination of words (e.g. verb-object rule) can emerge by auto-organization and cultural transmission. During the first stage organisms forage using the 18-bit feature information to discriminate mushrooms. Throughout the second stage, the foraging strongly depends on the evolution of a useful language, as features are rarely available. This condition has resulted in the rapid emergence of a shared language. In fact, within 30 generations the organisms' average fitness is the same as in generation 300, when the mushroom feature information was available all of the time. Moreover, the final fitness at generation 400 is higher (99 for the best organism) than that at generation 300 (91). This could be due to the fact that for neural networks it is easier to process discrete information, such as localist linguistic input, rather then processing the 18-bit feature information.

The percentage of evolved good language is 61%, as 11 out of 18 populations evolved useful languages. The remaining 7 populations (39%) evolved imperfect languages. However the discriminative quality of these languages was relatively good. In the majority of them only one of the four functional categories is incorrectly labeled. Two of the edible mushrooms categories are named by the same word/word-combination. Note that these imperfect languages also tend to use word combinations, and in particular most of them evolve the use of a verb-object rule.

After having shown that it is possible to evolve by auto-organization communication systems based on the combinatory rule "verb-object", we want to analyze the kind of referencing systems that organisms use when they associate words with objects. We are interested in establishing if the evolved languages are based on the use of grounded symbols, i.e. words that have a direct association with objects and that have logical relationships between them. We used a symbol acquisition test consisting of the training of organisms with a perfect combinatory language using the verb-object rule. The test was structured into three stages. In the first stage, organisms learn to name each of the four categories `e1`, `e2`, `t1`, `t2`. The teaching input is not provided by the parent organisms, but directly from the researcher. At this stage verbs are not used, and no names are taught for the two categories `e3` and `t3`. In the second stage, organisms learn to associate the two verbs "approach" and "avoid" with the categories `e1`/`e2` and `t1`/`t2` respectively. It is now expected that organisms learn the logical relationship between the names of the two edible mushrooms `e1` and `e2` and the verb "approach". The same symbolic association between the verb "avoid" and the names of toadstools `t1` and `t2` should be learned. In the final stage the learning of the names of categories `e3` and `t3` is finally introduced. The association the two verbs with these new names is not taught. In fact it is expected that after the training the organisms that learned real symbolic relationships between verbs and names will be able to generalize the verb-object relationship to the new mushroom names. If the verb "approach" is not associated with `e3` it means that organisms did not learn any symbolic association between the names of `e1` and `e2` and the verb "approach". They simply learned two independent object-word associations, one between `e1` and its name, and another between the same `e1` and the verb "approach".

This type of symbol acquisition test has been used in ape language studies. In the experiment where chimpanzees learned to associate "pour" with the name of the solid foods banana and orange (Savage-Rumbaugh & Rumbaugh, 1978), animals were then tested with new names of foods. Only those animals that made the correct generalization were considered to have learned symbolic associations.

The symbol acquisition test was repeated with 10 different populations. After the three learning stages, seven populations produced the correct associations `e3`-"approach" and `t3`-"avoid". The success criterion was the production of the correct verb for more than 75% of `e3` and `t3` mushroom types (N=8). In three populations the learning of the names for `e3` and `t3` did not produce the activation of the proper verb. It means that these organisms did not learn any symbolic association. In the seven successful populations, instead, the language is based on logical relationships between the mushrooms' names and the two verbs. The relationships between words and real objects, and between verbs and objects' name, allow neural networks to generalize the association of new names with the correct verb category.

These results show that neural networks can learn simple languages that use symbolic associations. These symbols are grounded in the environment because of the ecological simulation framework that allows a direct link between words and the objects with which organisms interact. However, the network's simple feedforward architecture allows the use of other non-symbolic strategies for language learning. In fact, during some simulations organisms appear to learn languages that do not use symbolic relationships. More complex and biologically-inspired neural network architectures could steer the learning towards symbolic acquisition, rather than simple stimulus associations. As neurophysiological data suggests (Deacon, 1997) the cortico-cortical connections in the human brain could help explain why humans can easily learn symbolic associations, while animals tend to learn conditional stimulus associations (except in controlled experiments as shown in ape language studies). Deacon's analysis of neuropsychological experiments on patients with prefrontal cortex lesions suggests that this area, and its connection with other cortical regions, could play a major role on symbol acquisition. The absence or underdevelopment of the prefrontal cortex in animals would subsequently explain the lack of symbol-based languages in animal communication systems. Our simple neural networks are not meant to represent any real neural systems. However, it is possible to design more articulate neural architectures that are inspired to specific connection patterns observed in the brain. Research is ongoing into understanding which particular neural network architectures allow symbol acquisition in language learning.

The model described in this paper allowed the simulation of the evolution of languages using simple syntactic rules and symbol acquisition. To create a selective pressure for the auto-organization of useful languages the experimental condition made the foraging behavior highly dependent on the parents' linguistic input (as mushroom features are only available 10% of the time). Moreover, the communication between parents and children, and the cultural transmission of language in the Naming Task, allowed the auto-organization of a population-shared language. The potential of this model for the study of the evolution of animal communication and human language is high. In future studies both the experimental setting, defining the availability of input mushroom features and of the linguistic input, and the model parameters controlling the pattern of communication and language learning between organisms, can be systematically changed to test specific hypothesis of the origin of language. For example, this model allowed us to focus on the important distinction between communication systems based on simple object-signal associations and languages based on symbolic relationships. The integration of the model with ongoing studies of the role of neural network architectures for the learning of symbolic representations will help to test Deacon's (1997) hypothesis. It will evaluate the role of symbol reference in language origin and on the co-evolution of brain structures, i.e. the prefrontal cortex, and language and symbolic acquisition.

References

Batali J. (1994). Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In R. Brooks & P. Maes (eds), *Artificial Life IV*, Cambridge, MA: MIT Press, 160-171.

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

Deacon T.W. (1996). Prefrontal cortex and symbol learning: Why a brain capable of language evolved only once. In B.M. Velichkovsky & D.M. Rumbaugh (eds*), Communicating meaning: The evolution and development of language,* Mahwah NJ: LEA Publishers

Deacon T.W. (1997). *The symbolic species: The coevolution of language and human brain,* London: Penguin.

Greenfield P.M. & Savage-Rumbaugh S. (1990). Grammatical combination in *Pan paniscus*: Process of learning and invention in the evolution and development of language. In S.T. Parker & K.R. Gibson (eds), *Language and intelligence in monkeys and apes,* Cambridge University Press, 540-579.

Harnad S. (Ed.) (1987). *Categorical Perception: The groundwork of cognition*. New York: Cambridge University Press.

Harnad S. (1990). The Symbol Grounding Problem. *Physica D* 42: 335-346

Hauser M.D. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Kirby S. (in press). Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners. In Chris Knight (ed.) *Approaches to the Evolution of Language*.

Kirby S. & Hurford J. (1997). Learning, culture and evolution in the origin of linguistic constraints. In P. Husband & I. Harvey (eds). *Proceedings of the Fourth European Conference on Artificial Life*, London: MIT Press

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

Parisi D., Denaro D., & Cangelosi A. (1998). The evolution of language as an aid to categorization. *Technical Report,* CNR-GRAL, Institute of Psychology. Poster presented at the *2nd International Conference on the Evolution of Language*, London.

Savage-Rumbaugh S. & Rumbaugh D.M. (1978). Symbolization, language, and Chimpanzees: A theoretical reevaluation on Initial language acquisition processes in four Young *Pan troglodytes*. *Brain and Language*, **6**: 265-300.

Steels, L. (1997a) The synthetic modeling of language origins. *Evolution of Communication*, **1**(1).

Steels L. (1997b). Self-organising vocabularies. In C.G. Langton & K. Shimohara (eds), *Proceedings of the ALIFE V conference*, Cambridge MA: MIT Press, 179-184.

Steels L. & Vogt P. (1997). Grounding adaptive language games in robotic agents. In P. Husband & I. Harvey (eds). *Proceedings of the Fourth European Conference on Artificial Life*, London: MIT Press

Acknowledgement. This research was partially supported by a Grant of the Nuffield Foundation (UK) for Newly Appointed Lecturers (SCI-180-97-116).