TALKING TO ONESELF AS A SELECTIVE PRESSURE FOR THE EMERGENCE OF LANGUAGE

MARCO MIROLLI[1,2]   DOMENICO PARISI[1]

[1] *Institute of Cognitive Sciences and Technologies, CNR, 44 Via San Martino della Battaglia, Roma 00185, Italy*

[2] *Philosophy and Social Sciences Department, University of Siena, 47 Via Roma, 53100, Siena, Italy*

Selective pressures for the evolutionary emergence of human language tend to be interpreted as social in nature, i.e., for better social communication and coordination. Using a simple neural network model of language acquisition we demonstrate that even using language for oneself, i.e., as private or inner speech, improves an individual's categorization of the world and, therefore, makes the individual's behavior more adaptive. We conclude that language may have first emerged due to the advantages it confers on individual cognition, and not only for its social advantages.

1. Introduction

1.1. *Talking-to-oneself and language evolution*

Human language has evolved. If we go back sufficiently in time we find ancestors of present-day humans who did not have language and then, some time later, ancestors who did have language. What is less clear is how and why language has evolved. If any capacity has evolved one can ask what adaptive advantages the possession of that capacity conferred to the individuals that possessed the capacity over other individuals that lacked it, even if this should not necessarily imply a pan-adaptivist view of evolution.

In the case of language it seems clear that language may have evolved because it conferred social advantages by dramatically improving the communicative capabilities of human beings which in turn improved coordination among individuals. The idea that language is just a very complex and powerful communication system can indeed be considered as the 'standard view' in the debate on language evolution (see, for example, Knight et al., 2000; Hurford, 2002; Pinker & Jackendoff, 2005).

Another less often explored possibility is that human language may have evolved because it made the cognitive functioning of single individuals more effective. Unlike animal communication systems human language can be used not only socially but also individually. It can be used to communicate with other individuals, by asking them information or providing them with information or

by asking them to do one thing or another, but it can also be used to talk with oneself, to comment on what one sees, to put one's predictions, explanatory hypotheses, and plans into words. These individual uses of language can result in more effective behaviours on the part of the individual even when the individual is acting alone, and this may have represented a reproductive advantage and a selective pressure for the emergence of language.

The idea that the main function of language is communication has been criticized by Bickerton (1990) and Chomsky (2002), who argue, from different standpoints, that language is fundamentally a system of mental representation and thought. Nevertheless, in the literature on language evolution the individual functions of language tend either to be ignored or to be thought to have appeared much more recently than it would be required if these uses acted as a selective pressure for the initial emergence of human language. It is assumed that human beings have learned to talk to themselves when their language was already completely developed and indistinguishable from the language spoken by present-day humans. The initial selective pressures for the emergence of language were social. When language was already evolved and fully modern, humans found that it could be usefully used to talk to oneself and not only to talk to others. But it is not clear that this is necessarily so. It is possible that even a very simple form of proto-language, consisting of words (or holophrases) that correlate with relevant experiences, can give important individual advantages if it is used not only for communication but also for talking to oneself. In this paper we use artificial neural network simulations for exploring the influence that a very simple form of proto-language can have on the categorization abilities of an individual when the language is used to talk to oneself, in private speech or in inner speech.

1.2. *Related work*

The idea that language is not only a system of communication but also a very powerful aid to cognition dates back at least to the 1930s, with the work of Russian psychologist Lev Vygotsky. Vygotsky (1978) considered language as a cognitive tool, whose development in the child and whose evolution in the species is the prerequisite to the development of all specifically human high-level psychological functions. Despite the fact that this idea has been somehow neglected in mainstream psychology, recently it has been raising increasing interest in the cognitive sciences (Gentner & Goldwin-Meadow, 2003) and is at the core of the interesting speculations on the evolution of human mind of such philosophers as Daniel Dennett (1991) and Andy Clark (2005). Indeed, there is

a growing body of empirical evidence demonstrating the importance of language for a number of cognitive functions including, for example, learning (Nazzi & Gopnik, 2001), memory (Gruber & Goschke, 2004), analogy making (Gentner, 2003) and problem solving (Diaz & Berk, 1992).

From the computational modeling point of view, there have been a few attempts to study the possible individual advantages that language can confer to those who possess it. Lupyan (2005) has demonstrated, with artificial neural networks, that learning to name perceived stimuli can facilitate category learning, especially the learning of those categories which are more difficult to learn. Using an artificial life framework, Cangelosi and Harnad (2000) have studied the adaptive advantage of what they call 'symbolic theft', that is the learning of new categories by hearsay from combinations of labels describing them. The results of their simulations show that organisms which employ symbolic theft outperform those which learn categories only by 'sensory-motor toil', that is, through trial-and-error and direct experience with objects. In previous work (Mirolli & Parisi, 2005b), we have shown that the use of a simple signalling system not only for communicative purposes but also for talking to oneself can facilitate the evolutionary emergence of the communication system itself. In particular, we have demonstrated that linguistic signals that benefit the hearer but not the speaker do not evolve if they are only used for communication while they do evolve if the hearer has to repeat the signals to himself or herself as an aid to memory. Finally, in a previous model on which the present work is based (Mirolli & Parisi, 2005a), we have shown that the coupling between the linguistic and the sensory-motor systems of an organism can improve categorization even after learning has taken place. In fact, we have shown that the internal representations of objects are improved if the objects are perceived together with the labels that designate them. In the previous model the linguistic input was always social, coming from another individual. In the present work we develop the model by studying the effects of language on categorization when the individual is all alone and he or she talks to herself either externally, with private speech, or internally, with inner speech.

2.   A simple neural network model of language acquisition

We assume that we can divide this process of language acquisition in two stages. In the first stage the learner's brain functions as two functionally separated modules, the sensory-motor module and the linguistic module, which are modeled as two independent feed-forward three-layer neural networks (see figure 1a). In the sensory-motor module the input units encode the sensory

properties of perceived objects that can belong to four different categories[a]. In the first, pre-linguistic stage the sensory-motor module learns (by a standard back-propagation learning algorithm) to classify perceived objects by producing in the output units a specific activation pattern which represents the action appropriate to the category of the object. On the other hand, the input units of the linguistic module encode the acoustic properties of heard linguistic sounds and the linguistic module learns (by back-propagation) to imitate heard sounds, that is, to produce in its output units the same activation pattern that has been received in the input units[b]. Overall, the first stage of learning is meant to model the first year of life of the child, when the child acquires both non-linguistic sensory-motor mappings and the ability to imitate heard sounds but there still is no learning of the mapping between linguistic sounds and their meanings.

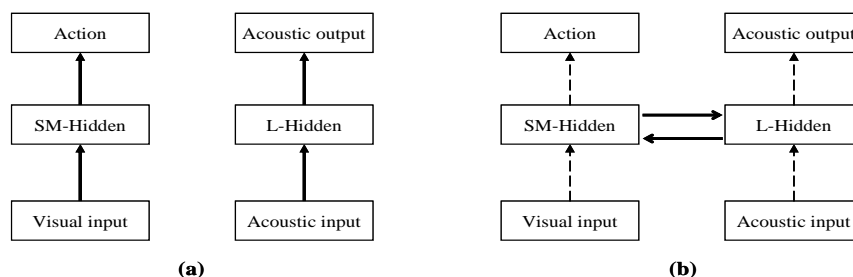**(a)**   **(b)**

Figure 1. The neural network. During the first stage of learning (a) the two modules learn their respective tasks independently of one another. During the second stage of learning (b) new connections linking the two modules appear or become functional and their synaptic weights are progressively adjusted so that the overall network learns to map objects into the linguistic sounds that designate them (speaking) and linguistic sounds into actions (understanding).

In the second stage, in which language learning in the proper sense begins, the two modules become functionally connected by the development of two new sets of synaptic connections, one linking the hidden layer of the sensory-motor module to the hidden layer of the linguistic module and the other linking the hidden layer of the linguistic module to that of the sensory-motor module (figure 1b). The second stage of learning, which is meant to model language learning in the proper sense, consists in adjusting the weights of these two new sets of connections (through a delta-rule learning algorithm) in such a way that (a) the internal representation evoked in the sensory-motor hidden units by a

---

[a] There are 120 different objects for each category, created by changing the category's prototypical perceptual pattern.

[b] For symmetrical reasons, there are also 120 real instantiations of the same words, each differing slightly from the word's prototype.

perceived object will tend to produce the internal representation evoked in the linguistic hidden units by the linguistic sound normally co-occuring with the object in the learner's experience, and vice versa, (b) the internal representation of a heard sound in the linguistic hidden units will tend to evoke its 'meaning', that is, the internal representation of the object normally experienced together with the sound. This results in the ability to speak by mapping non-linguistic inputs into the appropriate phono-articulatory movements (e.g., saying "apple" in response to the sight of an apple) and in the ability to understand language by mapping heard linguistic sounds into the appropriate non-linguistic actions (e.g., approaching the speaker in response to the linguistic sound "come here!").

3. How language affects categorization

In order to study how language affects categorization, we analyzed the internal representations of perceived objects in five different conditions, which we call (a) no-learning, (b) no-language, (c) social language, (d) private speech, and (e) inner speech. In our model, the internal representation of a perceived object is the activation pattern that is evoked in the sensory-motor hidden units by the perceived object. Activation patterns in a set of units can be represented as points in an abstract space with the same number of dimensions as the number of units in the corresponding layer of units and with each dimension of the space representing the level of activation of the corresponding unit, say, from 0 to 1. A particular internal activation pattern will be represented by a specific point located in the appropriate position with respect of each dimension, reflecting the activation level of the corresponding unit in the pattern. All the different members of a specific category of objects, say, all different apples, will be represented as a cloud of points. At the beginning of learning in the first stage, since all the connection weights are random, the cloud of points representing the internal activation patterns evoked by the different apples will be very large and it will largely overlap with the cloud of points representing other categories of objects. This is what is actually observed (see figure 2a). At the end of learning, the internal representations of objects belonging to the same category will form much better clouds, that is, clouds which are much smaller, do not overlap with other clouds, and have a considerable distance between their respective centers (figure 2b). Notice that the clouds in both the no-learning and no-language conditions are calculated by giving to the network only the perceptual properties of the object (figure 3a).

In order to test whether social language affects the internal representations of objects we calculated our clouds (categories) by making the network perceive

objects accompanied by instances of the linguistic sounds that normally accompany the objects in the learner's experience (figure 3b). The results, shown in figure 2c, demonstrate that language improves categorization in that the clouds of different categories of objects are now both smaller and more distant from one another than the clouds of objects unaccompanied by language.
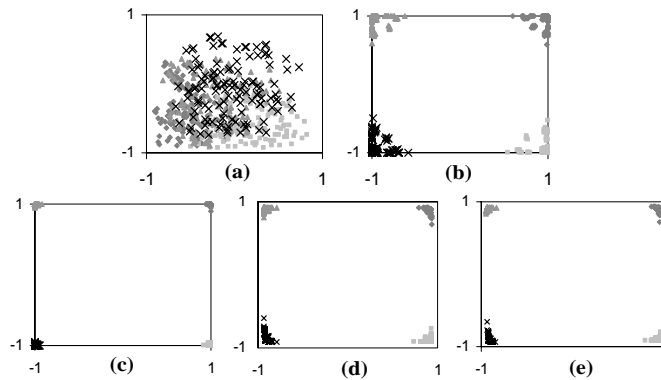

Figure 2. Internal representations of objects in the conditions of no-learning (a), no-language (b), social language (c), private speech (d), and inner speech (e). See text for details.

In order to study whether language can improve categorization not only if it is used socially but also if it is used individually, we simulated two ways in which humans can talk to themselves: externally, by speaking aloud but to themselves (private speech), and internally, by using language inaudibly (inner speech). In the private speech condition (figure 3c) the individual encounters an object and prior to responding to the object the individual produces the linguistic sound that designates the object. The individual hears the sound that he or she has produced and responds to an internal representation of the object which is influenced by the self-produced heard sound.

In the inner speech condition (figure 3d) when the individual sees the object he or she does not produce any externally audible sound. However, the sight of the object evokes in the individual not only the internal representation of the object but also, through the connections leading from the sensory-motor module to the linguistic module, the internal representation of the sound that designates the object. Through the connections leading from the linguistic module to the sensory-motor module, this internal representation of the sound that designates the object can influence the internal representation of the object in the sensory-motor module. The results of the both the private language and the inner language conditions (figure 2d and 2e, respectively) confirm the prediction that categorization is improved not only by social language, but also by talking to

oneself. In fact, both self-produced and internally-thought labels improve internal representations of perceived objects more or less to the same extent as external speech.
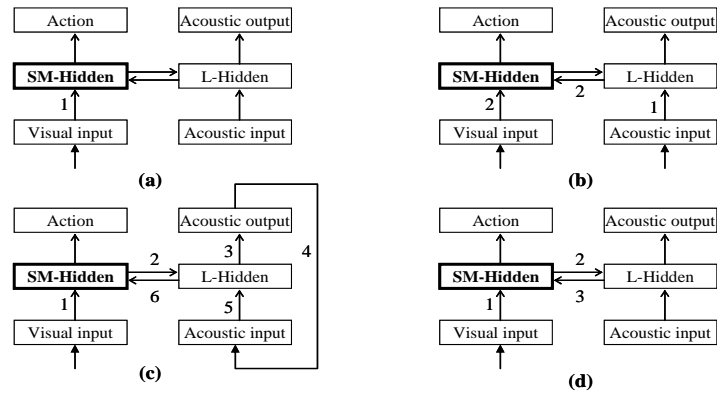
Figure 3. Experimental conditions for testing internal representations. The process of activation is divided up into a sequence of discrete time steps. Numbers indicate connections that are involved in each time step. See text for details.

4. Conclusions

Using a very simple neural network model of language acquisition it has been possible to demonstrate that using language for oneself can improve an individual's categorization of the world. We conclude that human language may have emerged not only due to social pressures, i.e., because language made it possible to have more sophisticated forms of social communication and coordination, but also due the advantages that language conferred on the cognitive functioning of the single individual when it is used to talk to oneself, either aloud or internally. Since these advantages can be demonstrated even with the very simple "language" of our simulations, we conclude that it is not necessary for language to be as sophisticated and complete as present-day language to provide individual cognitive advantages but that these advantages were probably already present in the very early stages of the evolutionary emergence of human language.

Acknowledgements

References

Bickerton, D. (1990). *Language and Species*. Chicago: University of Chicago Press.

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

Chomsky, N. (2002). *On Nature and Language*, Cambridge, MA: Cambridge University Press.

Clark, A. (2005). Beyond the Flesh: Some Lessons from a Mole Cricket. *Artificial Life*, 11 (1-2), 233-244

Dennett D.C. (1991). *Consciousness Explained*. New York: Little Brown & Co.

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

Gentner, D. (2003). Why we're so smart. In Gentner & Goldin-Meadow (2003) (pp.195-235).

Gentner D., & Goldin-Meadow S. (Eds.) (2003). *Language in mind*. Cambridge, MA: MIT Press.

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

Hurford, J. (2002). The roles of expression and representation in language evolution. In A. Wray (Ed.), *The transition to language* (pp. 311-334). Oxford: Oxford University Press.

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

Lupyan, G. (2005). Carving nature at its joints and carving joints into nature: How labels augment category representations. In A. Cangelosi, G. Bugmann and R. Borisyuk (Eds.) *Modelling Language, Cognition and Action* (pp. 87-96). Singapore: World Scientific.

Mirolli, M., & Parisi, D. (2005a). Language as an Aid to Categorization: A Neural Network Model of Early Language Acquisition. In A. Cangelosi, G. Bugmann and R. Borisyuk (Eds.) *Modelling Language, Cognition and Action* (pp. 97-106). Singapore: World Scientific.

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

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

Pinker, S., & Jackendoff, R. (2005) The faculty of language: What's special about it? *Cognition*, *95*(*2*), 201-236.

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