Published as: Mirolli Marco, Parisi Domenico (2005): 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. Proceedings of the 9th Neural Computation and Psychology Workshop. Singapore, World Scientific: 97-106

LANGUAGE AS AN AID TO CATEGORIZATION: A NEURAL NETWORK MODEL OF EARLY LANGUAGE ACQUISITION

MARCO MIROLLI [§†] DOMENICO PARISI [§]

[§] *Institute of Cognitive Sciences and Technologies,*
National Research Council,15 Viale Marx , 00137, Rome, Italy
d.parisi@istc.cnr.it

[†] *Philosophy and Social Sciences Department,*
University of Siena, 47 Via Roma, 53100, Siena, Italy
mirolli2@unisi.it

The paper describes a neural network model of early language acquisition with an emphasis on how language positively influences the categories with which the child categorizes reality. Language begins when the two separate networks that are responsible for nonlinguistic sensory-motor mappings and for recognizing and repeating linguistic sounds become connected together at 1 year of age. Language makes more similar the internal representations of different inputs that must be responded to with the same action and more different the internal representations of inputs that must be responded to with different actions.

1. Introduction

Understanding the relationship between language and cognition is one of the most difficult and important challenges facing cognitive science. Language has not only a communicative (social) function, but also a cognitive (individual) function (Vygotsky 1962, Jackendoff 1996, Clark 1998, Carruthers 2002). The view of language as something that transforms all human psychological processes dates back as early as the 1930s, with the work of Russian scholar Lev Vygotsky (1962, 1978), but it is a view which has been rather neglected in mainstream cognitive science. Nonetheless, the idea has not completely disappeared and recently has been raising increasing interest in philosophy and cognitive science (Carruthers & Boucher, 1998; Gentner & Goldwin-Meadow, 2003). Indeed, in recent years, language has been proposed to have an influence on, and to improve, many human psychological functions, among them categorization (Clark, 1998; Gentner, 2003), learning (Waxman & Markov, 1995; Nazzi & Gopnik, 2001), selective attention (Jackendoff, 1996; Clark, 1998), memory (Gruber & Goschke, 2004), voluntary control (Diaz & Berk 1992), perspective taking (Tomasello, 2003), analogy making (Gentner 2003), and reflexive thinking (Dennett 1991, 1996; Carruthers 2003). This paper

describes some very simple neural network simulations that (a) model the very earliest stages of language acquisition and (b) investigate the effects of language on categorization.

During her first year of life, the child learns to control her movements, to make appropriate sensory-motor mappings, to categorize perceptual experiences, and to reproduce her own sounds and the linguistic sounds of her environment. Notwithstanding all these progresses, this phase of child development is called 'pre-linguistic' because in her first 10-12 months the child does not show any strictly linguistic competence, that is, she is able neither to understand nor to meaningfully produce words. It is only around the end of her first year that the child learns to connect the linguistic sounds that have become familiar to her with their meanings as indicated by the fact that she reacts correctly to linguistic stimuli and she produces words in the appropriate circumstances.

Language acquisition can be considered to involve three sub-tasks(Kit, 2002): the acquisition of linguistic forms, the acquisition of non-linguistic sensory-motor mappings, and the association between linguistic forms and specific sensory-motor mappings, which become the meanings of the linguistic forms. Behavioral evidence (Bloom, 1994) suggests that the acquisition of linguistic forms and the acquisition of sensory-motor mappings run quite independently until the end of the first year and that only after the child has acquired a certain ability to map sensory inputs into motor outputs and to categorize experiences, on one side, and to recognize and produce linguistic forms, on the other side, the third task, the association of linguistic forms with specific sensory-motor mappings, can begin. The model of language learning presented here is based on this kind of behavioral evidence.

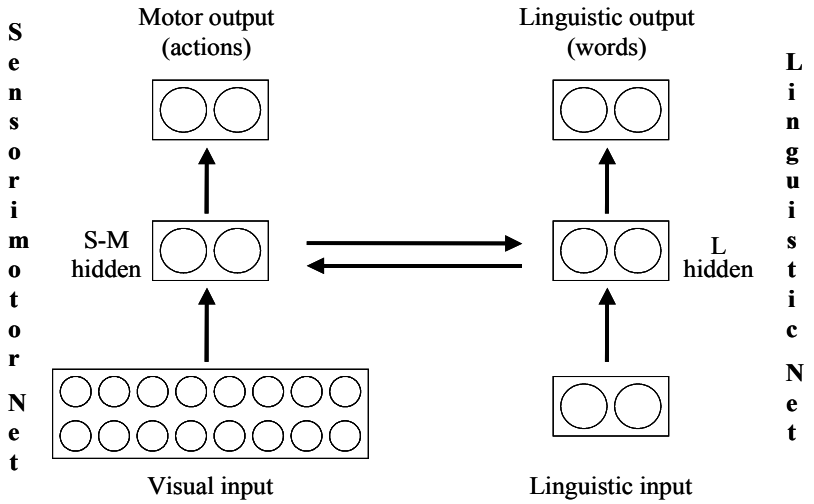
2. Method

2.1. *The neural network and the environment*

The neural network used in our simulations is modular. It is constituted by two sub-networks with three layer each, which we call the sensory-motor sub-network and the linguistic sub-network. The hidden layers of the two sub-networks are reciprocally connected by two matrices of connection weights so that the linguistic and the sensory-motor systems can interact with the each other (figure 1).

The sensory-motor sub-network has 16 bipolar input units (each unit's activation can be either 1 or –1) which encode the properties of perceived

objects, 2 hidden units (with continuous activation in the interval [–1; 1]), and 2 output units which encode the action performed by the network in response to each object. The activation of the two output units is thresholded to be either 1 or – 1, so that there are only four possible 'actions': <-1; -1>, <-1; 1>, <1; -1>, <1; 1>. The network's environment consists of 480 objects, belonging to 4 categories of 120 exemplars each. There are four prototype vectors, one for each category[a], and the perceptual properties of objects are generated by flipping 4



bites of the prototype to which the object belongs.

Figure 1. The neural network.

The linguistic sub-network has 2 input units encoding incoming linguistic signals, 2 hidden units, and 2 output units which represent emitted sounds. All the units of the linguistic sub-network have continuous activations in the interval [– 1; 1]. The linguistic environment is constituted by 4 words, which can be interpreted as the names of the four kinds of objects or of the appropriate actions to be performed upon them. Since words are pronounced in different ways by different persons and by the same person at different times, the acoustic inputs are created by changing slightly 4 prototype vectors, one for each word (<-0.5; -0.5>, <-0.5; 0.5>, <0.5; -0.5>, <0.5; 0.5>). There are 120 instances of each word and each instance is produced by changing both values of the

[a] The prototype vectors are: <-1; -1; -1; -1; -1; -1; -1; -1; 1; 1; 1; 1; 1; 1; 1; 1>; <-1; -1; -1; -1; 1; 1; 1; 1; -1; -1; -1; -1; 1; 1; 1; 1>; <-1; -1; 1; 1; -1; -1; 1; 1; -1; -1; 1; 1; -1; -1; 1; 1> and <-1; 1; -1; 1; -1; 1; -1; 1; -1; 1; -1; 1; -1; 1; -1; 1>

corresponding prototype vector by a amount randomly chosen in the range [–0.25; 0.25] .

2.2. *The two stages of learning*

The entire network goes through two successive stages of learning which are meant to correspond approximately to child's learning from birth to 1 year and to her learning from 1 year on, respectively. In the first stage of learning the two sub-networks are trained independently to accomplish two different tasks. The sensory-motor network learns to categorize objects and the linguistic network learns to repeat or imitate words. At the end of this first stage the connections between the hidden layers of the two sub-networks become functional and in the second stage of learning the network learns to associate words with the categories of objects to which they refer.

At the beginning the network is initialised with random connection weights in the range [-0.5; 0.5]. Then, 1000 cycles of backpropagation learning (with a learning rate of 0.2) are run in the following way. For the sensory-motor sub-network, one of the 480 different objects is randomly chosen, its perceptual properties are encoded in the input units of the sensory-motor network, the response of the network is calculated, and the correct action is provided to the network as teaching input (correct responses are <-1; -1>, <-1; 1>, <1; -1> and <1; 1> for objects belonging to category A, B, C and D, respectively). For the linguistic sub-network, one of the 480 possible instances of words is randomly chosen and encoded in the linguistic input units, the sound produced by the linguistic network in response to this heard sound is calculated, and the same sound heard as input is given as teaching input (in other words, the linguistic network has to accomplish an autoassociative task).

During this first stage of learning, the connections that link the hidden layers of the two sub-networks are non-functional. After 1000 cycles of back-propagation learning, the inter-network connections become functional and in the second stage of learning their weights are modified so that the network learns to associate the internal representations of objects (the vectors of activation of the sensory-motor hidden layer) with the internal representations of the appropriate words (the vectors of activation of the linguistic hidden layer), and vice versa. This second phase of learning runs as follow. One of the 480 objects is chosen randomly, together with one of the 120 instantiations of the word that designates the category to which the object belongs. The object and the word are given as input to the sensory-motor network and to the linguistic network, respectively. Then, the vectors of activation of the two hidden layers

are calculated and two cycles of the delta rule learning algorithm are applied by considering the two hidden layers with the connection weights in between them as two distinct perceptrons: one perceptron has the sensory-motor hidden layer has its input layer and the linguistic hidden layer as its output layer, and the other perceptron has the linguistic hidden layer as its input layer and the sensory-motor hidden layer as its output layer. The delta rule is applied by using the vector of activation observed in the linguistic hidden layer as teaching input for the first perceptron and the vector of activation observed in the sensory-motor hidden layer as teaching input for the second perceptron. The procedure, like that of the first stage of learning, is applied 1000 times with a learning rate of 0.2.

3. Results

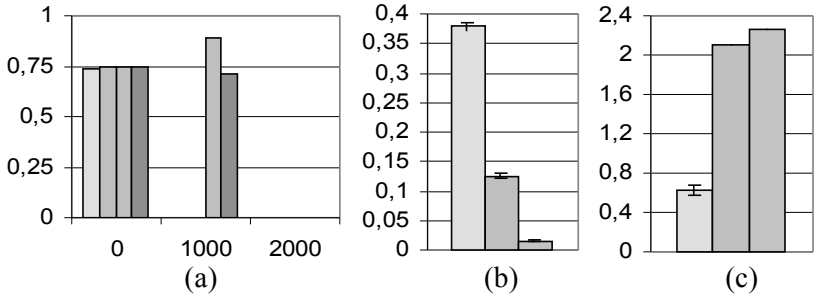
We have run two kind of tests, behavioral tests and internal representations tests. The behavioral tests have the function of assessing the performance of the network, while with the internal representation tests we analyse the effects of language on the categorization abilities of the network.

3.1. *Behavioral tests*

There are 4 behavioral tests: categorization, linguistic imitation, naming, and comprehension. In the categorization test the sensory-motor network is given an object as input and its performance is calculated as the difference between its action response and the right one (according to the object's category). In the linguistic imitation test the linguistic network is given the instantiation of a word as input and performance is how much the network is able to reproduce the heard sound in its sound output. In the naming test we give the network one of the 480 objects as input, activation spreads to the sensory-motor hidden units, then to the linguistic hidden units, and finally to the linguistic output units. We calculate the network's performance with respect to the vector prototype of the word that designates the category of the perceived object. The comprehension test is symmetrical to the naming one. In this test, we give the network an instantiation of a word as input and let the activation spread from the linguistic input units to the linguistic hidden units, then to the sensory-motor hidden units and, finally to the motor output units of the sensory-motor networks. The network's performance is calculated with respect to the response that is appropriate to the meaning of the word, that is, to the action to be produced in response to objects of the category to which the heard word refers.

All the behavioral tests are conducted three times: at the beginning of the

simulation (cycle 0), at the end of the first stage of learning (cycle 1000), and at the end of the second stage (cycle 2000). Figure 2a reports the percentage of errors in the four behavioral tests averaged on all the 480 inputs (objects in the categorization and naming tests and instances of words in the linguistic imitation and comprehension tests)[b]. The network learns both to perform the correct action in response to a perceived object and to imitate the linguistic sounds it hears: for both tasks the percentage of error decreases from around the 75% at the beginning of the simulation to 0%, that is, no error at all, at the end of the first stage of learning. Even the second stage of learning is very successful: while at the end of the first stage of learning the network gives



wrong responses almost 90% of the time in the naming test and around 70% of the time in the comprehension test, at the end of the second stage the network has acquired a perfect linguistic competence in that it always names objects correctly and responds appropriately to all the linguistic signals present in its environment.

Figure 2. (a): Percentage of errors in the categorization (left), linguistic imitation (middle-left), naming (middle-right) and comprehension (right) tests with random weights (0) and after the first (1000) and the second stage of learning (2000). (b): Average clouds' dimension before learning (left) and after learning in the 'vision' (middle), and 'language' (right) conditions. (c): Average distance between clouds before learning (left) and after learning in the 'vision' (middle), and 'language' (right) conditions. See text for details.

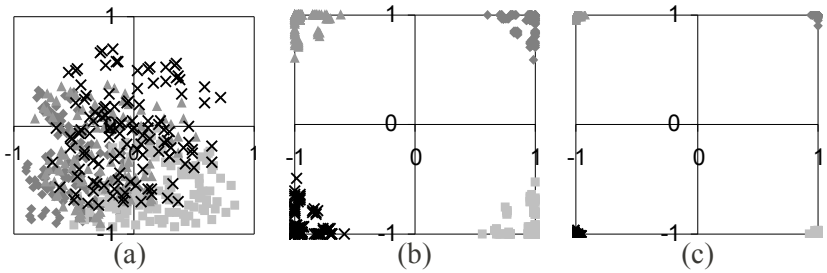
3.2. *Internal representations tests*

Given the perceptual properties of an object as input, we can look at the activation vector of the two sensory-motor hidden units (in the range [-1; 1]) and consider this activation vector as the network's internal representation of that object. Therefore, the internal representation of an object can be viewed as a point in a bi-dimensional space, with each dimension of the space

[b] All the results are averaged over 10 replication of the simulation with different initial conditions (random connection weights).

corresponding to the activation state of one hidden node; consequently, the internal representation of one category of objects is the "cloud" of points which represent all the objects belonging to that category (see figure 3). A good categorization is one in which (1) objects belonging to the same category are represented similarly, that is, the size of their cloud is small, and (2) objects belonging to different categories are represented differently, that is, the different clouds are distant from each other. So, for assessing goodness of categorization we use two measures: (1) cloud size, calculated as the average distance between all the points belonging to a cloud and the cloud's geometric center, averaged over the four clouds, and (2) distance between clouds, calculated as the average distance between the centers of the four clouds. We analyse the categorization performance of the trained neural network in two conditions: 'vision' and 'language'.

The vision condition corresponds to the situation described above, in which the clouds are the network's internal representations given the objects as visual input. In the language condition the network perceives the objects together with



their names (random instances of the four words corresponding to the four categories). In this condition we present both kinds of input, then we calculate the linguistic hidden unit activations given the linguistic input, and then we measure the sensory-motor hidden unit activations given both the visual input and the activation arriving from the linguistic hidden units.

Figure 3. Network's internal representations (clouds) before learning (a) and after learning in the vision (b) and language (c) conditions. See text for details.

Figure 3 shows the internal representations of objects before learning (a), and after learning in the vision (b) and language conditions (c). While before learning the four clouds are big and highly overlapping and so the network cannot discriminate between different kinds of objects (fig. 3a), after learning objects' categories are clearly distinguished in that each class of objects occupies one of the four corners of the network's representational space (fig.

3b). In other words, during training the network has learnt to categorize objects by decreasing the average size of the clouds (fig. 2b, left and middle bars) and by increasing the average distance between the clouds (fig 2c, left and middle bars). This is just another example of the well-known categorical perception effect (Harnad et al.1995). What is more interesting is the effect that language has upon categorization: as you can see comparing the internal representations in the vision (fig. 3b) and the language (fig. 3c) conditions, categorization is consistently facilitated when objects are perceived together with their 'names'. In fact, the four clouds are both significantly smaller (fig. 2b middle and right bars) and more distant from each other (fig. 2c middle and right bars).

4. Discussion and conclusions

We have described a neural network model of early language acquisition. The model assumes that during her first year of life the child separately learns various sensory-motor mappings (reaching, manipulating, categorizing objects) and various sound-related abilities (recognizing sounds, repeating her own sounds and the sounds produced by others). It is only at the end of the first year that language learning begins. The two separate networks that are responsible, respectively, for the sensory-motor mappings and for the sound-related abilities become functionally connected and language learning consists in learning the connection weights linking the two networks. These weights allow the child to produce linguistic sounds in the appropriate circumstances and to react appropriately to the linguistic sounds produced by other individuals.

But language also changes the way in which the child categorizes reality. Input-output mappings require categorization. Categorization is to make the internal representations (patterns of activation in the network's internal units) of different inputs more similar if the different inputs must be responded to with the same action, and more different if the inputs must be responded to with different actions (Di Ferdinando & Parisi 2004). All organisms categorize reality. And infants categorize reality during their first year in order to generate the appropriate input/output mappings. But our model shows that language can influence cognition by inducing better categorization of non-linguistic inputs.

Our model can also be considered as a general model of the integration of multiple sensory-motor mappings in the brain. Much activity in the brain consists in exploiting the co-variations observed in experience to establish correlations between different sensory-motor mappings. For example, visual-motor mappings are correlated with proprioceptive-motor and tactile-motor mappings in such a way that given one type of input the brain can predict the

other one. Language is but one example of this type of integration between various non-linguistic sensory-motor mappings and the sensory-motor mapping from heard sound to pronounced sounds. Given a visual input from an object, the brain generates the internal representation of the word (sound) that designates the object. Or, given a word (sound), the brain generates the internal representation of the visually or haptically perceived object designated by the word.

However, language is somewhat different from other types of integration between different sensory-motor mappings. While other correlations between different sensory-motor mappings are not arbitrary, language is arbitrary - and this may require a species-specific genetically inherited basis for language learning. Given a visual input from a new object, the brain can predict how the visually perceived object would feel if touched. In contrast, given a new object the brain cannot predict what is the word that designates the object or, given a new word, the brain cannot predict which object is designated by the word. Furthermore, variations in how the same word is pronounced do not predict variations in the object designated by the word. Another important difference between language and other types of integration between different sensory-motor mappings is that the overall space of variation of heard or pronounced linguistic signals appears to be much smaller than the space of variation of the objects or actions that are designated by the linguistic signals. This is particularly important for categorization because it might contribute to explaining why language makes our sensory-motor categories more distinctive and more compact. Finally, language may be different because the sensory-motor mapping from heard to pronounced words can easily become a circular or closed circuit. When one sees an object, the visual input from the object does not only generate the internal representation of the word that designates the object but it can generate the actual sound of the word through private or inner speech.

In fact, unlike animal communication systems, human language is not only used to communicate with other individuals but it is also used to communicate with oneself, either externally (private speech) or internally (inner speech). In the simulations described in the paper, the language which arrives to the neural network is produced by another individual. We intend to further explore how language influences categorization by applying our neural network model to private speech and inner speech, trying to answer the following question: How an individual categorizes reality when he or she is exposed to reality and at the same time to language stimuli that the individual him- or herself generates,

either externally or internally?

Acknowledgments

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. Thanks also to an anonymous referee for his useful comments.

References

Bloom P. (Ed.) (1994). *Language acquisition: Core readings*. Cambridge, MA: MIT Press.

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

Carruthers P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, 25, 657-726

Carruthers P., Boucher J. (Eds) (1998). *Language and thought: Interdisciplinary themes*. Cambridge: Cambridge University Press

Clark A. (1998). Magic words: How language augments human computation. In Carruthers and Boucher (Eds) (1998).

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

Dennett D.C. (1996). Reflections on language and mind. In Carruthers and Boucher (Eds) (1998).

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

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

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

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

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

Kit C. (2002). How does lexical acquisition begin? A cognitive perspective. *Cognitive Science*, 1, 1-50

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

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

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

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