

# The Sensorimotor Bases of Linguistic Structure: Experiments with Grounded Adaptive Agents

Angelo Cangelosi

Adaptive Behaviour & Cognition Research Group  
School of Computing, Communication and Electronics  
University of Plymouth (UK)  
acangelosi@plymouth.ac.uk  
www.plymouth.ac.uk/staff/acangelosi

## Abstract

This research uses grounded adaptive agents for investigating the evolutionary origins of syntactic categories, such as nouns and verbs. To analyze the sensorimotor bases of linguistic structure, the techniques of categorical perception and of synthetic brain imaging are employed. The simulation uses two different architectures for the adaptive agent's neural controller. Analyses show that the neural processing of verbs is consistently localized in the regions of the networks that perform sensorimotor integration, while nouns are associated with sensory processing areas. The general implications of such model and of the analysis techniques for adaptive behavior and language evolution research are discussed.

## 1. Modeling language evolution in grounded adaptive agents

Investigations on the emergence of language, both in evolutionary and developmental terms (MacWhinney 1999; Cangelosi & Parisi 2002), have greatly benefited from the use of computational models. This has resulted in an alternative approach to the nature vs. nurture (and biological vs. cultural) dilemma in language emergence. Models demonstrate that language is a complex system that emerges from intricate interactions between various biological and environmental processes. Adaptive behavior and artificial life provide useful modeling methodologies (Kirby 2002; Wagner et al. 2003) for dealing with such a complex system vision of language. In adaptive behavior models, populations of autonomous agents interact via language games to exchange information about the environment. Their coordinated communication system emerges from the direct interaction between agents.

Amongst the various adaptive behavior approaches, some provide a more integrative vision of language and treat it as an integral part of the whole cognitive system. The agent's linguistic abilities are strictly dependent on, and grounded in, other behaviors and skills. Various sensorimotor, cognitive, neural, social and evolutionary factors contribute to

the emergence and establishment of communication and language. For example, in these models there exists an intrinsic link between the communication symbols (words) used by the agent and its own cognitive representations (meanings) of the perceptual and sensorimotor interaction with the external world (referents). We call this grounded adaptive agent modeling of the emergence of language. This is consistent with the psychologically-plausible theories of the grounding of language in the organism's perceptual and action systems (Barsalou 1999; Glenberg & Kaschak in press; Joyce *et al.* 2003). Such an approach is in opposition to other adaptive modeling systems that view language as an independent and autonomous capability of the agent, and are subject to the symbol grounding problems (Harnad, 1990).

There are various language models based on grounded adaptive agents. Some use real (or realistic) robots interacting in physical environments, while others use simulated adaptive agents. In robotic models, communication results from the dynamical interaction between the robot's physical body, its cognitive system and the external physical and social environment. Some studies stress the grounding in sensorimotor processes, such as Marocco's *et al.* (2003) model of robotic arms in a physics dynamics simulator and Vogt's (2002) mobile robots. Other robotic models highlight the grounding through social interaction, such as Steels' (2002) talking heads and Kaplan's (2000) interactive toy robot. On the other hand, some studies are based on simulation adaptive agents. They model the agent and its environment with a good degree of detail upon which emergent meanings can be directly constructed. These simulation models have focused on grounding in perceptual experience (Cangelosi *et al.* 2000; Dyer 1994) and in cognitive representations and sensorimotor interactions (Cangelosi & Harnad 2000; Hazlehurst & Hutchins 1998).

Both robotic and simulation models have been used to study the emergence of simple communication systems, and that of more complex syntactic languages. Of the models focusing on syntax, some have specifically addressed the evolutionary transition from non-compositional

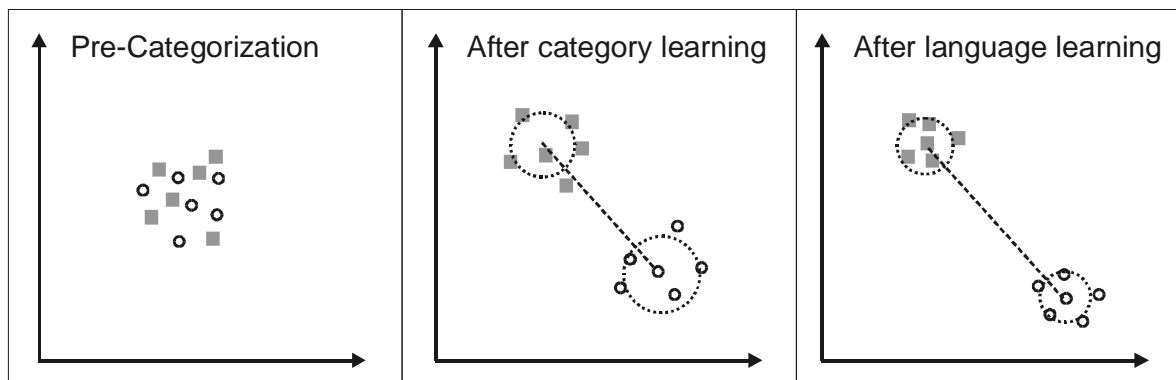


Figure 1: Typical formation of clusters of points (i.e. square and circle categories) during category and language learning. Before category learning (Left), points corresponding to different categories overlap. After categorization (Centre) and language learning (Right) points group in distinct areas. Notice the enhancement of compression and separation (CP effects) between simple categorization and language learning. These diagrams do not correspond to real data, but only provide a sketch of CP effects observed in modeling experiments (e.g. Cangelosi et al. 2000).

communication to syntactic language. This is an important issue in the field of language origins, because it can shed light on the evolutionary emergence and diversification of word categories, such as nouns, verbs and function words (e.g. prepositions and quantifiers). This topic is also very important for studies on the ontogenetic development of language based on the emergentist/constructivist approach (Tomasello & Brooks, 1999; MacWhinney 1998).

This paper will focus on the use of grounded adaptive agents for investigating the evolutionary origins of syntactic categories, such as nouns and verbs. It will use grounded adaptive agents in which neural networks control all behaviors, including sensorimotor, cognitive and linguistic abilities. In order to better understand how all these behaviors interact, two techniques will be adopted for examining the agents' internal (neural) representations. These analyses can identify and highlight the sensorimotor bases of linguistic structure, such as the role of perceptual and motor knowledge in differentiating the syntactic categories of verbs and nouns. The first methodology considers the phenomena of categorical perception in the hidden representations of neural networks. The second technique applies neuroimaging principles to artificial neural networks for comparing neural representations in different linguistic and behavioral tasks. The results of such analyses will also be related to our current knowledge of the neural basis of language processing.

## 2. The neural processing of language and cognition

The neuropsychological and neurocognitive literature on language processing in the brain is quite extensive (Pulvermuller, 2003). This has been possible through the use of a variety of scientific methods, such as neuropsychological experiments of patients with

psycholinguistics deficits, brain imaging studies on normal adults and language-impaired patients and artificial neural network models. These methods support detailed investigations of the neural principles and mechanisms behind various linguistic abilities. For example, a recent review paper has analyzed numerous studies on the neural processing of the syntactic categories of nouns and verbs (Cappa & Perani, 2003). This review explicitly includes and compares brain imaging experiments on verb and noun processing (e.g. Martin *et al.*, 1995; Perani *et al.*, 1999). Cappa and Perani (2003) report an overall agreement on the fact that the left temporal neocortex plays a crucial role in lexical-semantic tasks related to the processing of nouns. The processing of verbs, instead, involves additional regions of the left dorsolateral prefrontal cortex.

Such data show that language processing is strictly interconnected with other sensorimotor and cognitive abilities. For example, the prefrontal cortical areas implicated in verb processing are those also involved in motor control. This finding is consistent with other work that illustrates a strong relationship between language and motor development (Greenfield, 1991; Rizzolatti & Arbib, 1998). More generally, in cognitive psychology and neuropsychology it is widely accepted that language is not an autonomous function of the organisms, but it is strictly dependent on/from other cognitive abilities (Gazzaniga, 2000).

In parallel to neuroscientific investigations, neural network models have extensively been used to study the neural control of linguistic behavior (Christiansen & Chater, 2001; Just et al. 1999). However, most of these models tend to study language in isolation. Only a few studies have directly and purposely simulated in the same model linguistic and motor functions (e.g. Reilly, in press). These include grounded adaptive agent models (e.g. Cangelosi 2001; Marocco et al. 2003), where by definition linguistic abilities emerge from sensorimotor and cognitive skills.

## 2.1 Categorical perception

Category learning has been hypothesized to be one of the core capabilities supporting the evolution and development of language and cognition. In particular, categorical perception (CP) is proposed as the basic mechanism sustaining our ability to build discrete and hierarchically-ordered representations of the environment (i.e. categories). Language and other higher-order cognitive abilities are then grounded upon such categories (Harnad, 1987; Cangelosi & Harnad 2000). CP refers to the effect of “warping” the similarity space of internal categorical representations. This results in the compression of within-category differences between members of the same category and the expansion of between-category distances amongst members of different categories. Such phenomena can be graphically represented through the process of the formation of clusters of points in the similarity space of categories (Figure 1). The first diagram (Fig. 1 Left) shows the undifferentiated similarity space before category acquisition. The other two diagrams represent the formation of quite distinct clusters (categories) after category learning has occurred without (Fig. 1 Centre) and with (Fig. 1 Right) language. These diagrams show the presence of two distinct categories (cluster of squares vs. cluster of circles) in an abstract two-dimensional similarity space. Each dimension may correspond to some classification component (e.g. presence of a feature) or to the hidden unit activation of a neural network. Relative distances in this space can be calculated using Euclidean measures between points. The two dotted circles in each diagram represent the within-category distances, corresponding to the standard deviation of the Euclidean distances between each point and the center of its cluster. The continuous straight line represents the between-category distance, e.g. the Euclidean distance between the centers of the two clusters.

CP has been shown to occur in animals (Zentel *et al.* 1986) and human subjects (Goldstone 1994; Andrews, Livingstone & Harnad, 1998). The compression effects have also been analyzed in real neural systems (Kosslyn *et al.* 1989) and in artificial neural networks (Tijsseling & Harnad 1997; Cangelosi, Greco & Harnad 2000; Nakisa & Plunkett 1998). CP effects have been reported in models performing both category and language learning. Specifically, language appears to enhance the CP warping effects. The acquisition of categories via linguistic instruction produces stronger compression effects than those obtained in categories acquired through direct experience with stimuli. This is sketched out in the comparison of Figure 1 Left (after simple category learning) and Right (after category learning via language).

More recently, CP effects have been investigated in different syntactic categories. In a grounded model of the evolutionary acquisition of verbs and nouns (Cangelosi & Parisi 2001), it was reported that the

strength of language-enhanced CP effects varies depending on the word type used by agents. Verbs produce stronger warping effects than nouns (Figure 2). These distinct internal representation patterns for the processing of verbs and nouns are also shown to generate differential benefits on the agent. This is because more similar representations of different situations help the agents to respond with the same action, albeit they are used in tasks with more variable environmental contexts (e.g. different objects/category may require the use of the same action, and its corresponding verb). Nouns, on the other end, tend to co-vary with the features of input stimuli (categories) and require less differentiated categorical representations (Cangelosi & Parisi, 2001).

In this paper further investigations of such differential verb/noun effects will be presented. Specifically, CP measurements will be performed in two neural architectures where the integration of sensorimotor information varies depending on the input of motor-related (proprioceptive) signals.

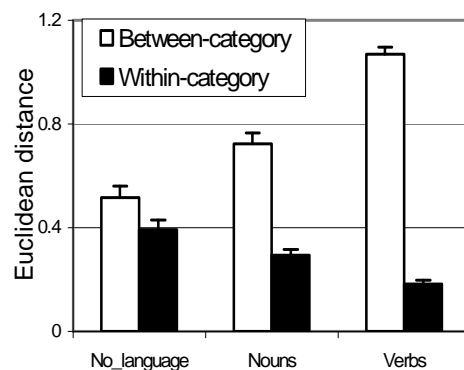


Figure 2: Categorical perception (CP) effects with verbs and nouns. Notice the enhanced CP effects in verbs, compared to the other conditions (adapted from Cangelosi & Parisi 2001)

## 2.2 Synthetic brain imaging

Experimental brain imaging techniques (also known as neuroimaging) permit the visualization and identification of neural activity in accurate regions of the brain in response to controlled tasks. Neuroimaging is based on comparisons of differences in brain activity (rCBF, regional Cerebral Blood Flow) between tasks. The two most commonly used techniques are Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI).

Synthetic brain imaging (SBI) (Arbib *et al.* 2000; Horwitz *et al.* 1999; Cangelosi & Parisi, *in press*) is a modeling technique used to highlight the regions of an artificial neural network that are functionally active in response to specific stimulations and tasks. SBI can be used to compare directly PET and fMRI imaging data in empirical studies and in computational models. This permits the development of more detailed and neurally-plausible models of behavioral functions. In analogy with experimental

neuroimaging, SBI is based on differences in artificial neural “activity” between tasks. For example, following Arbib et al.’s (2000) assumptions, in synthetic PET the involvement of a region in a specific task is correlated with local integrated synaptic activity. Synaptic activity is computed by summing together the synaptic inputs arriving to the artificial neurons of a specific area (ignoring the positive/negative sign of the connections). Arbib and colleagues (2002) used synthetic PET to test their FARS model of parietal-premotor interactions in the grasping control of primates. In a study on language processing, Just *et al.* (1999) applied synthetic neuroimaging in a computational model of sentence comprehension. They found good agreement between the number of activated voxels in human fMRI experiments and the predictions of their model for sentence of different complexity.

In this paper SBI techniques will be used to analyze the internal activations in the neural controllers of adaptive agents. Various linguistic conditions will be compared, in particular to show differences in the neural processing of verbs and nouns.

### 3. The evolution of nouns and verbs: An adaptive behavior model

A grounded simulation model of the evolution of nouns and verbs was employed. This uses a population of object manipulation agents that are able to understand the names of objects (nouns) and the names of actions (verbs)<sup>1</sup>. The simulation extends a previous language evolution model (Cangelosi & Parisi, 2001, in press), by employing different architectures in the agent’s neural controller.

#### 3.1 The agent and the tasks

The simulated agent (Figure 3) consists of an organism with a retina and a two-segment arm (e.g. Schlesinger & Barto, 1999). The arm moves in a 2D environment in front of the agent. One of two different objects (object A = vertical bar; object B = horizontal bar) can appear at any time in the environment. The object occupies either three vertical cells or three horizontal cells and is centered in one of the 9 central cells of a 5x5 grid. The pixel values (0 for empty cell, 1 for cell occupied by a portion of the object) will be projected to the 25 units of the agent’s retina. The agent’s arm is transparent and absent from the retinal image.

During each task, an agent begins by grasping an object with its hand and it has either to pull the object toward itself or to push the object away from itself. The agent may respond to visual stimuli, linguistic instructions (nouns, verbs), or both. There are 11

different tasks which combine these two sensory modalities (Figure 4). The agent always receives proprioceptive input regarding the current position of the two segments of its arm.

In the first task (No\_Language), the retina receives a 5x5 image of the object location and shape. No linguistic input is present. When object A is present in the scene, the agent must push it. Instead, object B must be pulled towards the agent’s shoulder.

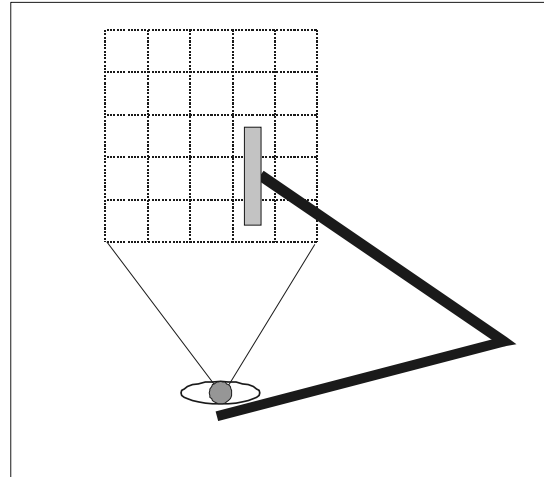


Figure 3: Agent interaction with environment. The two-segment arm grasps the object and has to pull it toward itself or push it away from itself. This depends on the type of object and the linguistic command that the agent may receive.

	OBJECT A		OBJECT B	
	retina	language	retina	language
No_language	■ +	* *	■ +	* *
Noun only	■ +	* A	■ +	* B
" "	* +	* A	* +	* B
Verb only	■ +	Push *	■ +	Pull *
" "	* +	Push *	* +	Pull *
" "	■ +	Pull *	■ +	Push *
" "	* +	Pull *	* +	Push *
Noun Verb	■ +	Push A	■ +	Pull B
" "	* +	Push A	* +	Pull B
" "	■ +	Pull A	■ +	Push B
" "	* +	Pull A	* +	Push B

Figure 4: List of the retina and language input for the 11 tasks. The \* symbol indicates the absence of input (i.e. a series of 0s)

The remaining 10 tasks involve language understanding. In 5 Vision+Language tasks, in addition to the retina and proprioceptive input, the following linguistic instructions are received by the agent: (1) “A” or “B”, as the nouns of objects A and B; (2) “Push” or “Pull” for the verbs describing the default action (push for object A; pull for B); (3) “Push” or “Pull” for the verbs describing the opposite action (pull for object A; push for object B); (4) “Push A” or “Pull B” for describing the default actions by combining a noun and a verb; (5) “Push B” or “Pull A” for describing the opposite actions. When

<sup>1</sup> This is a model of language comprehension only. A related model that deals with both linguistic comprehension and production is described in Cangelosi (2001).

the verb is received as input, the agent must do whatever the verb meaning indicates, thus overriding the default action if the verb describes the opposite action. In the 5 Language-only tasks, only the linguistic instruction and the proprioceptive input are provided.

The use of the terms “noun” and “verb” is not intended to correspond to the full blown grammatical categories of verbs and nouns. Nevertheless, the nouns and verbs of the present simulation capture some of the fundamental properties of nouns and verbs of real human languages (Parisi *et al.*, 2002). Nouns co-vary with the input stimuli and can act as attentional cues. Verbs co-vary with the motor action performed by the agents. These simple properties may have been those of primitive proto-nouns and proto-verbs which, through cultural and linguistic evolution, have given way to the complex, full blown grammatical categories. As suggested by Pinker (1994) the use of mental labels for different types of objects and actions has clear adaptive advantages.

### 3.2 The neural networks

The behavior of the agent is controlled by a feedforward neural network. Two different architectures will be used (Figure 5). They differ in the connection pattern between some input and hidden nodes.

The input layer contains 33 units. The proprioceptive input is encoded in 4 units. These encode the position of the arm through the input of the angles of the two pairs of muscles (extensor and flexor) of each of the two arm’s segments (shoulder and elbow). The visual object is perceived through a retina of 25 units corresponding to the 5x5 cells of the environment grid. Four localist-encoding linguistic units are used for language input, two respectively for the nouns “A” and “B” and two for the verbs “Push” and “Pull”.

The output layer contains four motor units. These control the extension/contraction of the four arm muscles (a pair of extension/contraction muscles per arm segment). The output activation corresponds to the force that is applied to each muscle.

The network has two layers of hidden units, whose organization varies in the two different architectures. In the first network configuration (Figure 5 top), the first hidden layer has two separate modules: a group of 5 units for the preprocessing of proprioceptive information and a group of 10 units for the preprocessing on linguistic and visual stimuli. In the other network architecture (Figure 5 bottom), all 14 first-layer hidden nodes preprocess the input coming from all input units. In all networks, the second hidden layer contains 5 units with connections to the previous layer of hidden nodes. The motivation behind such manipulations of the neural architecture is to create a first network with a full modular organization, where separate sensory processing modules for motor information (proprioceptive input)

and visual information (retina input) exist, in addition to the sensorimotor integration module of the second hidden layer. The second architecture, instead, uses a fully distributed approach where proprioceptive and visual information are integrated since the first hidden layer.

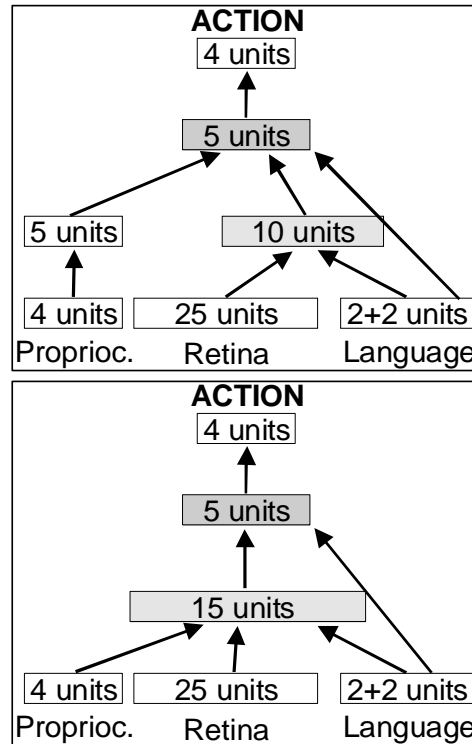


Figure 5: The two neural network architectures used in the simulation: Modular (Top) and distributed (Bottom) architectures. The connection patterns between the proprioceptive input units and the two hidden layers (shaded rectangles) are varied in the two networks.

The two architectures resulting from the different organization of the hidden layers will be used to compare the categorical perception effects in two experimental conditions: (1) use of a modular architecture, with separated sensory-processing and sensorimotor integration modules; (2) use of a fully connected network directly based on sensorimotor integration. These two architectures permit the analysis of the functional organization of verb and noun processing in the networks in relationship to their sensorimotor abilities. In addition, the first architecture will be used to make qualitative comparisons with brain imaging data on modular language processing in humans. In all networks, the input language units are always connected to both hidden layers in the two networks. This is to better control the specific effects of the differential sensorimotor processing levels on linguistic input.

### 3.3 The evolutionary algorithm

A genetic algorithm is used to evolve the connection weights of the agents’ neural networks. During the

agent’s lifetime, the weights do not change. The motor behavior and language understanding improves (i.e. is learned) during evolution.

During each generation, a population of 80 agents is used. At the beginning of the simulation, 80 genotypes (connection weights) are randomly generated in the range of  $\pm 1$ . Weights are encoded as real numbers. A single task in the lifetime of an individual consists of a total of 18 subtasks (2 objects x 9 positions). Each subtask lasts for 20 input/output cycles, necessary for the arm to move an object from its initial position to the target. The fitness formula computes the total number of subtasks successfully completed by each agent in all conditions. A subtask is considered successful when (a) object A is pulled to a distance of 5 points or less from the agent’s shoulder, or (b) object B is pushed to a distance of 45 points or more from the agent’s shoulder.

At the end of the generation, agents are ranked on the basis of their fitness formula. The 20 agents with highest fitness are selected for reproduction. Each individual asexually generates 4 offspring with the same genotype of its single parent except for the addition of random changes (mutations) to some of the weights. Five percent of weights are mutated by adding a random quantity between  $\pm 1$ .

In the first 1000 generations, agents only perform the No\_language task. These initial generations are needed for evolving a set of connection weights that allows agents to respond appropriately to the two objects (before language is introduced). From generation 1001 to 2000, agents have a longer lifetime and are exposed to all 11 tasks. These are experienced by each individual in a random order. In the fitness formula for the 10 tasks with language, a subtask is successful if the agent pushes or pulls the object according to the linguistic input.

### 3.4 Fitness results

For each network architecture, 20 replications were performed using different random initial populations. In all replications, agents evolved an ability to respond appropriately to the different input conditions. At the end of generation 2000, over 90% of the objects are responded to with the appropriate pushing or pulling behavior in all tasks.

In this paper the focus is on the neural processing of nouns and verbs, rather than on evolutionary and fitness results. Therefore the following sections will discuss only the categorical perception analyses and the synthetic brain imaging data. For a more detailed discussion of fitness and evolutionary data in this type of model, see Cangelosi & Parisi (2001).

## 4. Analyses on language processing

### 4.1 Categorical Perception (CP) results

In previous simulations of this language evolution model (Cangelosi & Parisi, 2001), it was shown that

categorical perception (CP) varied in the word categories of verbs and nouns. Verbs produced an enhancement of between-category distances and reduction of within-category sizes when compared with nouns (Figure 2). These results refer to the activation patterns of the second hidden layer where sensorimotor integration happens<sup>2</sup>. The proposed explanation for such phenomenon is that verbs are better than nouns in shaping the internal representations (i.e. enhanced CP effects) because they “prepare the motor output with which the organism must respond to the input” (Cangelosi & Parisi 2001: 174).

This explanatory hypothesis, based on CP data taken from the hidden layer preceding the output units, can have two possible interpretations. The first hypothesis assumes that verbs distinguish themselves from nouns in the hidden layers next to the motor layer. As a consequence, the closer a hidden layer is to the output motor layer, the more action-oriented the representation is (which is the case of verbs that co-vary with the action to be performed). The alternative interpretation is that verbs produce CP effects whenever hidden units play a direct role in shaping the type of motor response through the integration of internal sensory (visual and linguistic) representations and motor (proprioceptive) information. The analyses on the two architectures will resolve such an ambiguity and shed light on the relationship between sensorimotor processing and linguistic categories.

In this study the CP data were computed for both hidden layers<sup>3</sup> of each of the two architectures. Figure 6 reports the between-category data for the modular (top) and the fully distributed (bottom) architecture. Data in Figure 6 were computed and averaged using the hidden unit activations of the best agent in each of the 20 replications. To compute the between-category distances, the sets of  $N$  hidden activation values  $h$  for the object A ( $h1_A, h2_A \dots hn_A$ ) and B ( $h1_B, h2_B \dots hn_B$ ) are necessary. These constitute the coordinates of two points in a  $N$ -dimensional Cartesian plane. Then the Pythagoras theorem is applied to obtain the Euclidean distance  $d_{A-B}$  between the two points:

$$d_{A-B} = \sqrt{(h1_A - h1_B)^2 + (h2_A - h2_B)^2 + \dots + (hn_A - hn_B)^2} \quad (1)$$

Results clearly indicate that in the modular network (Figure 6 Top) verbs differ from nouns only in the second hidden layer, i.e. where motor (proprioceptive) information integrates with sensory (visual) input. Instead, in the distributed architecture (Figure 6 Bottom) verbs differentiate from nouns starting from the first sensorimotor hidden layer.

<sup>2</sup> In Cangelosi & Parisi (2001), the network architecture is similar to that of Figure 5 (top). The only difference was in the linguistic input solely connected to the second hidden layer.

<sup>3</sup> The CP data on the proprioception preprocessing layer of the first architecture were not computed because they do not receive linguistic input.

These data supports the hypothesis that verbs produce enhanced CP effects whenever hidden units play a direct role in shaping the motor response through sensorimotor integration. This is independent from the relative proximity to the motor output layer.

The localization of the CP effects also highlights the specific sensorimotor basis of verbs, and the sensorial characterization of nouns. The linguistic structure behind the differentiation of the word categories of verbs and nouns is strictly interconnected with sensorimotor representations.

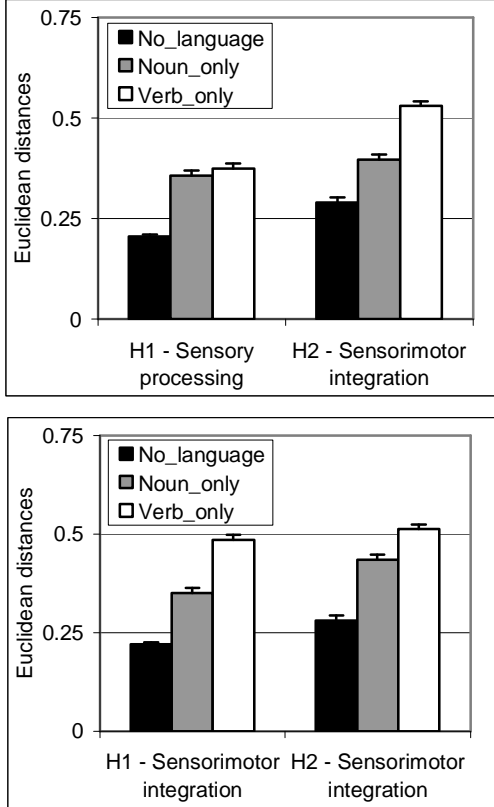


Figure 6: Between-category distances for both layers of the modular network (Top) and the fully-distributed architecture (Bottom). The Euclidean distances in the Y axes are normalized to permit direct comparisons between layers with different number of units.

## 4.2 Synthetic Brain Imaging (SBI) results

Synthetic brain imaging analyses were performed only on both neural architectures. SBI will be used to make some qualitative comparisons between these artificial models and known data on syntax processing in the human brain.

The SBI technique presented here will compare the neural activity of the two hidden layers in different linguistic tasks. This method is called Integrated Neural Activity (INA) and generally follows Arbib *et al.* (2000) synthetic fMRI procedure. A related synthetic brain imaging method, called Integrated Synaptic Activity (ISA), has also been developed to make comparisons with PET data. ISA

will not be described and applied here, but for a comparison of INA and ISA analyses in grounded neural networks see Cangelosi & Parisi (in press).

The overall value of the integrated neural activity  $INA_A(1/2)$  reflects the difference of activation between the two tasks 1 and 2 in the region A (e.g. hidden layer). It can be computed using the following formula:

$$INA_A(1/2) = |rINA_A(1) - rINA_A(2)| \quad (2)$$

where the regional Integrated Neural Activity for each task  $rINA_A$  is

$$rINA_A = \sum_{i=1}^N a_i \quad (3)$$

with  $N$  being to the number of units in the hidden layer, and  $a_i$  the activation of each individual neuron in the region A.

In small networks like the ones used here, it is possible to consider the  $INA_A(1/2)$  as the average of differences between individual neuron activations in the two tasks. In this case, the following formula can be used:

$$INA_A(1/2) = \frac{\sum_{i=1}^N |a_i(1) - a_i(2)|}{N} \quad (4)$$

Formula (4) will be used here, due to the small number of units in each hidden layer. Only the first activation value produced at the beginning of each task is used. Observations of the behavior showed that agents produce distinct push/pull movements since the first activation cycle. The INA values for all possible pairs of task subtractions were calculated separately for each hidden layer. Subsequently, all subtractions were averaged for the two linguistic classes of verbs and nouns. For example, the average of the values No\_language–Noun\_only and Verb\_only–Noun\_verb contribute to the INA value for Noun processing.

Figure 7 contains the average INA values over the 80 individuals of the 20 populations at generation 2000. Means and standard errors were computed for both the modular architecture (Figure 7 Top) and the fully distributed architecture (Figure 7 Bottom). In general, higher columns correspond to more activity in the specific hidden layer and linguistic task to which they refer.

For each network, an analysis of variance with repeated factors (MANOVA) was computed to compare the average differential activities in the two hidden layers and in the two linguistic tasks of noun and verbs. A sample size of 20 was considered, i.e. one data point for each of the 20 populations. The dependent variable was the INA value averaged over the 80 networks of each population. In computing the INA values for each individual network, only the values when the individual agent correctly manipulates the two objects (over 90% of the cases)

were used. The two independent variables used in the MANOVA were HIDDEN\_LAYER (2 levels: 1<sup>st</sup> and 2<sup>nd</sup>) and NOUN\_VERB (2 levels: noun tasks and verb tasks).

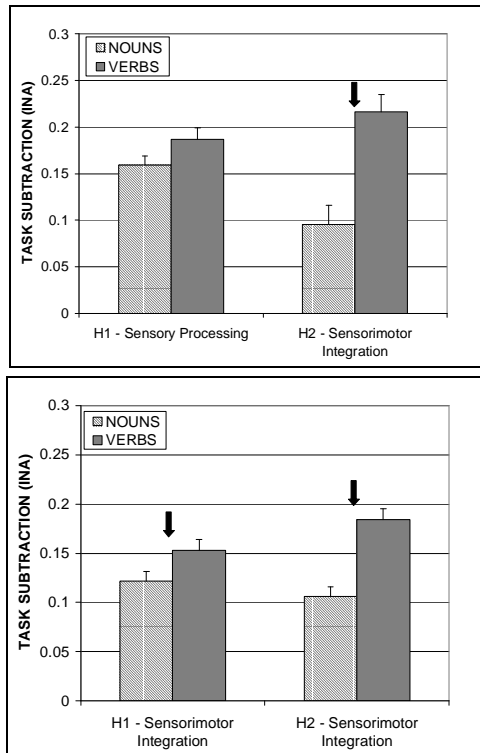


Figure 7: Data on the synthetic brain imaging of verb and noun processing in the two networks. Black arrows indicate significant statistical difference between the two verb/noun columns. The modular network (Top) has no significant differential activity for nouns in the first hidden layer. Verbs are significantly more active in the second hidden layer. In the network with fully distributed architecture (Bottom) verbs are significantly more active in both layers.

Two individual MANOVAs were carried out, one for the modular network and one for the distributed architecture. In both analyses, only the NOUN\_VERB factor and the interaction between the two factors were significant. In particular, to identify the specific differences between the 4 means of the interaction, some post hoc test were carried out. In the first network (Figure 7 Top), there is no significant difference of activity between nouns and verbs in the first hidden layer ( $p=0.281$ ). Instead, verbs produce significantly more neural activity than nouns in the second hidden layer ( $p=0.009$ ). In the second network (Figure 7b) verbs are significantly more active than nouns in both layers ( $p=0.017$  in the first layer, and  $p<0.001$  in the second).

Overall, this pattern of results confirms that verb processing is “localized” in the regions of the network specialized for sensorimotor integration. This is the case of only the second hidden layer in the modular network, and of both layers in the fully distributed architecture. In addition, the results are consistent

with the categorical perception analyses in the two neural architectures.

## 5. Discussion and Conclusions

This research was based on grounded adaptive agents for simulating the evolutionary acquisition of simple lexicons based on verbs and nouns. The use of neural networks that control all behaviors (sensorimotor, cognitive and linguistic) has permitted a better understanding of the interaction between such abilities. The manipulation of the connectivity in the two networks had the scope of designing agents with different neural processing strategies for the control of sensorimotor and linguistic behavior. The CP and SBI analyses have highlighted that the linguistic structure differentiating the two categories of nouns and verbs is strictly dependent on the perceptual and sensorimotor representations evolved by the agents. For example, tasks involving the use of verbs require the activation of the same neurons (and hidden layers) that play the function of integrating sensorial information (e.g. vision for identifying the shape and position of the object) with motor signals (proprioception for planning arm movement) to produce action.

In section 4.1, two alternative mechanisms were hypothesized for explaining differences in the neural processing of verbs and nouns. The first hypothesis suggests that the closer the hidden layer is to the motor output units, the more it will be specialized for verb representation and processing (e.g. verbs will produce enhanced CP effects and higher SBI activity). The alternative hypothesis is that a hidden layer specializes for verb processing when it integrates sensory information, such as vision and language, with proprioceptive information. Both CP and SBI analyses support the second hypothesis. For example, verbs produce enhanced CP effects in the hidden layers that play a direct role in determining the motor response through sensorimotor integration. This is the case of the second hidden layer in the modular network and of both layers in the fully-distributed network. This also contradicts the first interpretative hypothesis on the correlation between specialization for verb processing and relative proximity to the motor layer. In fact, in the distributed architecture, both layers show enhanced CP effects, with stronger effects in the layer more distant from the output units.

The value of this model in highlighting the sensorimotor grounding of the linguistic categories is strengthened by the similarity between the model performance and data on the neuroscience of language processing. In section 2.2 it was illustrated that experimental brain imaging studies have consistently established that nouns activate more the posterior areas of the brain related to sensory and associative processing, while verbs activate more the anterior motor areas (Cappa & Perani, 2003). Although the aim of this simulation and the SBI



analyses is not that of building a neurally-plausible model of language processing in the brain, it is nevertheless remarkable that such a simple neural network model can be constrained to exhibit some functional and architectural similarities with the neural processing of language. This outcome further supports the use of grounded adaptive agent models for testing language evolution hypothesis. For example, the current model could be easily expanded to manipulate known constraints on the animal and human brains that have been hypothesized to affect the evolution of cognitive and linguistics abilities (Deacon, 1997). Simulations would then investigate the validity of such hypotheses.

From a more general perspective on language origins, one of the most important contributions of this simulation is the demonstration that syntax evolution is directly grounded in sensorimotor and cognitive representations. The CP and SBI analyses show that verbs systematically “follow” the sensorimotor integration processing units, regardless of their position in the neural network. Therefore we can say that there is some kind of functional and neural equivalence between understanding the name of an action and performing the action (also regardless of any linguistic input). This coupling of language and sensorimotor processing abilities is consistent with cognitive grammar literature (Langacker 1987). According to such a theoretical approach, the formation of linguistic constructs leads back to cognitive constraints and functions.

This work also has more general methodological implications for research in adaptive agents. The two techniques of CP and SBI have been used here to visualize and analyze the internal representation used by agents in various tasks and behavioral contexts. With these methodologies, the neural and categorical representations can be not only visually represented, but can also be compared using quantitative methods (e.g. Figure 7 with statistically comparisons of means in different hidden layers and linguistic tasks). Information visualization has become an important issue in adaptive behavior and artificial life research (Smith *et al.* 2002; Marocco *et al.* 2003). Current adaptive behavior models are becoming increasingly more complex in terms of evolutionary, behavioral and structural properties. Simulations and robotic studies can produce huge sets of data that are often difficult to analyze using existing methodologies. The development of information visualization tools, widespread in other areas such as genomic and neuroscience research (Stuart *et al.* 2001), can be fruitfully adopted in adaptive behavior. The two techniques used here can be easily adapted and used to visualize and interpret the internal representations that adaptive agents use to control complex behaviors.

The model is currently being extended to study the emergence of syntactic languages. The lexicon will not be provided by the experimenter, but rather it will self-organize through the communicative interaction

between agents (e.g. Cangelosi 1999). In addition, agents will learn language from their own peers and parents, so that the simulation can combine both the evolutionary and developmental acquisition of language. New simulations will focus on the transition from a non-symbolic and non-compositional lexicon to the emergence of combinatorial and syntactic languages. Further manipulations of the neural network architectures will be conducted to determine the modular organizations favoring the transition to syntactic languages.

## Acknowledgements

The research was supported by the UK Engineering and Physical Research Council (EPSRC Grant GR/N01118). The author thanks Domenico Parisi for valuable discussions on this model.

## References

- Andrews J., Livingston K., Harnad S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24: 732-753
- Arbib M.A., Billard A., Iaconi M., Oztop E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Networks*, 13: 975-997
- Arbib M.A., Fagg A.H., Grafton S.T. (2002). Synthetic PET imaging for grasping: From primate neurophysiology to human behavior. In F. Sommer, A. Wichert (Eds.), *Explorative analysis and data modelling in functional neuroimaging*. Cambridge, MA: The MIT Press.
- Barsalou L. (1999) Perceptual symbol systems. *Behavioral and Brain Sciences*, 22: 577-609.
- Cangelosi A. (2001). Evolution of communication and language using signals, symbols, and words. *IEEE Transactions on Evolutionary Computation*. 5(2): 93-101
- Cangelosi A., Greco A., Harnad S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, 12(2): 143-162
- 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
- Cangelosi A., Parisi D. (in press). The processing of verbs and nouns in neural networks: Insights from Synthetic Brain Imaging. *Brain and Language*.
- Cangelosi A., Parisi D. (2002). *Simulating the Evolution of Language*. London: Springer.
- Cangelosi A., Parisi D. (2001). How nouns and verbs differentially affect the behavior of artificial organisms. In J.D. Moore, K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, London: Lawrence Erlbaum Associates, 170-175.
- Cappa S.F., Perani, D. (2003). The neural correlates of noun and verb processing. *Journal of Neurolinguistics*, 16(2-3): 183-189

- Christiansen M.H., Chater N. (Eds.) (2001). *Connectionist Psycholinguistics*, Westport: Ablex.
- Deacon T.W. (1997). *The Symbolic Species*. London: Penguin.
- Dyer M.G. (1994). Grounding language in perception. In V. Honavar, L. Uhr (Eds.), *Artificial Intelligence and neural networks: Steps toward principled integration*. Boston: Academic Press.
- Gazzaniga M.S. (Ed.) (2000). *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press.
- Glenberg A.M., Kaschak M. P. (in press). Grounding language in action. *Psychonomic Bulletin & Review*
- Goldstone R. (1994). Influences of categorization of perceptual discrimination. *Journal of Experimental Psychology: General*, 123: 178-200
- Greenfield P. (1991). Language, tool and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14: 531-595.
- Harnad S. (Ed.) (1987). *Categorical Perception: The Groundwork of Cognition*. New York: CUP
- Harnad S. (1990). The symbol grounding problem. *Physica D*, 42: 335-346
- Hazlehurst B., Hutchins E. (1998). The emergence of propositions from the co-ordination of talk and action in a shared world. *Language and Cognitive Processes*, 13: 373-424
- Horwitz B., Tagamets M.-A., McIntosh A.R. (1999). Neural modeling, functional brain imaging, and cognition, *Trends in Cognitive Science*, 3: 91-98.
- Joyce D., Richards L., Cangelosi A., Coventry K.R. (2003). On the foundations of perceptual symbol systems: Specifying embodied representations via connectionism. *Proceedings of the 5<sup>th</sup> Intl. Conference on Cognitive Modeling (ICCM 2003)*. Bamberg
- Just M.A., Carpenter P.A., Varma S. (1999). Computational modeling of high-level cognition and brain function. *Human Brain Mapping*, 8: 128-136.
- Kaplan F. (2000). Talking AIBO: First experimentation of verbal interactions with an autonomous four-legged robot. In A. Nijholt, D. Heylen, K. Jokinen (Eds.). *Learning to Behave: Interacting agents. CELE-TWENTE Workshop on Language Technology*, 57-63.
- Kirby S. (2002). Natural language and artificial life. *Artificial Life*, 8: 185-215.
- Kosslyn S.M., Koenig O., Barrett A., Cave C.B. et al. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception & Performance*, 15(4): 723-735.
- Langacker R.W. (1987). *Foundations of Cognitive Grammar, Vol.1: Theoretical Prerequisites*. Stanford, Stanford University Press.
- MacWhinney B (Ed.) (1999). *Emergence of Language*. Hillsdale, NJ: LEA.
- MacWhinney B. (1998). Models of the emergence of language. *Annual Review of Psychology*, 49: 199-227.
- Marocco D., Cangelosi A., Nolfi S. (2003), The emergence of communication is evolutionary robots. *Philosophical Transactions of the Royal Society London – A*, 361: 2397-2421
- Marocco D., Stuart E.J., Cangelosi A. (2003). Visualisation techniques in artificial life: The method of parallel coordinates. *1st Italian Workshop on Artificial Life*, Universita' della Calabria, Italy.
- Martin A., Haxby J.V., Lalonde F.M., Wiggs C.L., Ungerleider L.G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, 270: 102-105.
- Nakisa R.C., Plunkett K. (1998) Evolution of a rapidly learned representation for speech. *Language and Cognitive Processes*, 13: 105-127
- Parisi D., Cangelosi A., Falcetta I. (2002). Verbs, nouns and simulated language games. *Italian Journal of Linguistics*, 14(1): 99-114
- Perani D., Cappa S.F., Schnur T., Tettamanti M., Collina S., Rosa M.M., Fazio, F. (1999). The neural correlates of verb and noun processing: A PET study. *Brain*, 122: 2337-44.
- Pinker S. (1994). *The Language Instinct*. Penguin.
- Pulvermuller F. (2003) *The neuroscience of language. On brain circuits of words and serial order*. Cambridge: Cambridge University Press.
- Reilly R.G. (in press). The relationship between object manipulation and language development in Broca's area: A connectionist simulation of Greenfield's hypothesis. *Behavioral and Brain Sciences*.
- Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience*, 21: 188-194.
- Smith T., Bullock S., Bird J. (2002). Beyond fitness: Visualising evolution - Workshop overview. *Artificial Life VIII*, Sydney.
- Steels L. (2002). Grounding symbols through evolutionary language games. In A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*. London: Springer-Verlag.
- Stuart L., Walter M., Borisyuk R. (2002). Visualisation of synchronous firing in multi-dimensional spike trains. *BioSystems*, 67: 265-279.
- Tijsseling A., Harnad S. (1997). Warping similarity space in category learning by backprop nets. In M. Ramscar et al. (Eds.). *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization*, Edinburgh University, 263-269.
- Tomasello M., Brook P.J. (1999). Early syntactic development: A Construction Grammar approach. In M. Barrett (Ed.), *The Development of Language*. Philadelphia, PA: Psychology Press, 161-190
- Vogt P. (2002). The physical symbol grounding problem. *Cognitive Systems Research*, 3: 429-457
- Wagner K., Reggia J.A., Uriagereka J., Wilkinson G.S. (2003). Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1): 37-69.
- Zentall T.R., Jackson-Smith P., Jagielo J.A., Nallan G.B. (1986). Categorical shape and color coding by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 12(2): 153-159.