Robots with language

Domenico Parisi*

Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

Trying to understand human language by constructing robots that have language necessarily implies an embodied view of language, where the meaning of linguistic expressions is derived from the physical interactions of the organism with the environment. The paper describes a neural model of language according to which the robot's behaviour is controlled by a neural network composed of two sub-networks, one dedicated to the non-linguistic interactions of the robot with the environment and the other one to processing linguistic input and producing linguistic output. We present the results of a number of simulations using the model and we suggest how the model can be used to account for various language-related phenomena such as disambiguation, the metaphorical use of words, the pervasive idiomaticity of multi-word expressions, and mental life as talking to oneself. The model implies a view of the meaning of words and multi-word expressions as a temporal process that takes place in the entire brain and has no clearly defined boundaries. The model can also be extended to emotional words if we assume that an embodied view of language includes not only the interactions of the robot's brain with the external environment but also the interactions of the brain with what is inside the body.

Keywords: emotional words, language, robots

STUDYING LANGUAGE BY CONSTRUCTING ROBOTS THAT HAVE LANGUAGE

If we want to construct *human* robots rather than just *humanoid* robots, that is, if we want to construct robots which actually behave like human beings rather than robots which only resemble human beings in their external morphology, it will be necessary for our robots to possess language because language is such a prominent feature of human beings. Some robots give us the impression of being able to use language but they do not actually understand the language they hear or produce. They are programmed to respond with specific actions to specific acoustic inputs and to generate specific sounds in specific circumstances but human language is much more than that. Of course, human language is a very complicated behavior and it will not be easy to construct robots that can be said to really possess language. But we can make some steps in that direction.

Studying language by constructing robots that have language implies a specific conceptual framework with which to look at human language. Robots are real or simulated physical artifacts. They have a body, they have sensors and effectors with which they interact with the physical environment, their behavior is controlled by a simulated "brain" (an artificial neural network), and their body contains (or should contain; cf. Parisi, 2004) not only a "brain" but also other internal organs and systems. Therefore, robots necessarily imply an "embodied" conception of cognition according to which cognition depends on, and is shaped by, the possession of a body and the movements of the body's different parts. Cognitive representations have been traditionally thought of as based on perception or as abstract representations that do not contain sensory-motor information. However, recent empirical findings and theoretical developments favor a different conception of cognitive representations according to which the body of the organism and the movements of the body's effectors play a critical role in shaping the organism's cognitive representations (Gibson, 1979; Clark, 1999; Barsalou, 2008a,b). Furthermore, if the robot's behavior is controlled by a neural network (as it should be since the brain is part of the body and, to be consistent, robots should be neurorobots), cognitive representations become neural representations, that is, patterns of activation or successions of patterns of activation in a set of units that simulate the brain's neurons. This usefully operationalizes the rather vague concept of cognitive or mental representation since, unlike cognitive or mental representations, artificial neural representations can be observed, measured, and compared with empirical data on the brain.

Constructing robots that have language necessarily extends the embodied conception of cognition to language. Empirical evidence in favor of an embodied conception of cognition has accumulated not only in experiments in which participants respond to the sight of objects but also when they respond to words that refer to those objects (for a recent review, see Fischer and Zwaan, 2008). In both cases, the sensory input activates the neural representation of the action with which participants usually respond to the object. However, a well-developed embodied theory of language should be able to answer many questions that remain still open, and constructing robots that have language should help us to answer these questions. Here are some examples. Are there differences in what happens in the brain when a participant responds to the sight of

an object and when he/she responds to a word which refers to the object? Do nouns evoke what are called the stable affordances of an object, i.e., the action with which one responds to the object and which is represented in the brain independently of the actual movements with which the action is physically realized in different circumstances, while seeing an object in a particular orientation also evokes the variable affordances of the object which specify at least some aspects of the movements one has to produce to physically realize the action (Borghi and Riggio, 2009)? Do verbs that refer to actions only or mainly activate the neural representation of the state of the environment which is produced by the movements of the effectors, i.e., the effect of the action, while the sight of an action activates a neural representation of both this state and the movements of the effectors that will produce it? Does possession of a language change how the brain responds to perceived objects and actions in that an individual tend to internally label those objects and actions and therefore he or she responds to both the perceived object and action and the self-produced linguistic signal? Are there differences between the neural representations evoked by verbs and by nouns, or by nouns that refer to tools (e.g., hammer) and nouns that refer to natural objects (tree) (Cangelosi and Parisi, 2001)? How can an embodied theory of language account for abstract words? Do abstract words imply going back and forth between the part of the brain that processes words as sounds and the part which constructs a meaning for the sound, while this is less true for concrete words? How is an embodied meaning for combinations of words (phrases and sentences) constructed? How can an embodied theory account for emotional words and for the emotional component of non-emotional words?

In this paper we describe a simple neural network architecture for language-using simulated robots living in simulated environments and we try to show how this architecture may explain a (very limited) number of linguistic behaviors, where to explain a behavior is to construct a robot that reproduces the behavior. We will refer to robots that have been actually constructed and we will indicate how these robots could be modified to account for other linguistic phenomena.

OBJECTS ARE INTERNALLY REPRESENTED IN TERMS OF THE ACTIONS WITH WHICH WE RESPOND TO THEM

Neurorobots develop internal (neural) representations of perceived objects which are based on the motor actions with which they respond to the objects rather than on the objects' perceptual properties. Imagine a robot whose neural network has sensory units encoding the visual properties of objects, motor units encoding the movements of the robot's arm, and an intermediate layer of internal units. The robot lives in an environment in which it may be exposed to one of four possible objects possessing two properties, color and shape, both with two values, blue and red, square and circle. The robot sees one object at a time and it has to respond by reaching with its arm one of two buttons, one on the right and the other on the left (Borghi et al., 2003, 2005; Di Ferdinando and Parisi, 2004). (The connection weights of the neural network of all the robots described in this paper are evolved using a genetic algorithm. See Mitchell, 1998). If the button on the right has to be reached when the robot sees a square object, independently of the object's color, and the button on the left when the robot sees a

circular object, again independently of the object's color, we find that the four objects are represented in the internal units of the robot's neural network in terms of the action that the robot has to do in response to the objects (go to the button on the right, go to the button on the left) rather than in terms of the perceptual properties of the objects as such. In fact we observe only two activation patterns in the neural network's internal units, one which controls the action of reaching the button on the right and the other one which controls the action of reaching the button on the left. The object's perceptual property which is critical to decide which action to do, in our case shape, determines the internal representation of the object, while the other property, color, is ignored. Notice that the internal representation of an action is abstract in the sense that it needs to be translated into a succession of specific movements of the robot's arm which vary as a function of the starting position of the arm. In fact, the robot's neural network includes an additional set of proprioceptive input units encoding the current position of the robot's arm which project directly to the motor units. The two activation patterns that constitute the internal representations of the two actions interact with this proprioceptive information from the arm so that, for any starting position of the arm, each of the two abstractly represented action can be translated in the appropriate succession of movements.

A NEURAL NETWORK'S ARCHITECTURE FOR LANGUAGE-USING ROBOTS

The robots we have described in the preceding Section do not have language but they only respond to objects with the appropriate non-linguistic action. We now ask: What is the basic architecture of the neural network that controls the behavior of a language-using robot? The robot's overall neural network is made up of two sub-networks, the non-linguistic sub-network (NL) and the linguistic sub-network (L) (Mirolli and Parisi, 2005). Both sub-networks include three layers of units: a sensory layer, a motor layer, and an intermediate layer of internal units. The sensory units of NL encode perceived objects and its motor units encode movements of the robot's effectors such as the robot's arm. The sensory units of L encode heard linguistic sounds and its motor units encode movements of the robot's phono-articulatory organs that result in the production of linguistic sounds. NL maps non-linguistic sensory input into non-linguistic actions. In fact, NL is identical to the neural network of our robots that had to reach with their arm one of two different buttons in response to a visually perceived object. L maps heard linguistic sounds into phono-articulatory movements. This is the network that controls the behavior of a robot which is able to imitate (repeat) heard sounds without associating any meaning to them. The robot hears a linguistic sound and it responds with movements of its phono-articulatory organs that reproduce the sound.

The two sub-networks remain functionally or perhaps even anatomically separate during an initial period of the robot's existence which corresponds to children's first year of life. During this period the robot learns to respond to non-linguistic sensory input (say, perceived objects) with movements of its non-linguistic effectors (arm, hands, legs, eyes) using its NL sub-network. In addition the robot uses its L sub-network to produce linguistic sounds with its phono-articulatory organs, either spontaneously or in response

of its own heard sounds (babbling) and, later on, by imitating the linguistic sounds produced by already speaking robots.

At the end of this period the two sub-networks begin to be connected together by two-way connections that go from the internal units of NL to the internal units of L, and vice versa, and the synaptic weights of these two-way connections are learned based on the co-variation of specific linguistic sounds with specific objects and actions in the robot's experience. From this point on our robot becomes a language-using robot. The robot still uses NL to produce non-linguistic actions in response to non-linguistic sensory input but, in addition, it begins to understand and to produce language. Language understanding consists in responding to heard linguistic sounds with the appropriate movements of the non-linguistic effectors while language production consists in responding to non-linguistic input with phono-articulatory movements that produce the appropriate linguistic sounds. In language understanding neural activation spreads from the sensory layer of L (heard words) to the internal units of L and from there to the internal units and to the motor layer of NL (non-linguistic actions). In language production activation spreads from the sensory layer of NL (perceived objects and actions) to the internal units of NL and from there to the internal units and to the motor layer of L (phono-articulatory movements that produce words). (We are talking here of overt responses to sensory input but activation can stop at the internal layer of the two sub-networks, where non-linguistic and linguistic actions are neurally represented, without producing overt responses, that is, without translating these actions into actual physical movements of either the non-linguistic or linguistic effectors).

INFLUENCE OF LANGUAGE ON THE ROBOT'S CATEGORIES

The network architecture described in the preceding Section allows us to (begin to) answer the question of what are the consequences of possessing a language for a robot's cognition, that is, for the functioning of the robot's NL sub-network. More specifically, in this Section we will see what are the consequences of possessing a language for the robot's categories.

As we have seen in Section "Objects are Internally Represented in Terms of the Actions with Which We Respond to Them", perceived objects that have to be responded to with the same action elicit an identical activation pattern in the network's internal units even if they are perceptually different, while perceived objects that have to be responded to with different actions elicit different activation patterns in the internal units. This is the basis for defining an action-based notion of categories. A category is an internal activation pattern elicited by different objects that have to be responded to with the same action. For the robots of Section "Objects are Internally Represented in Terms of the Actions with Which We Respond to Them", two objects with different color elicit the same activation pattern in the robot's internal units if they have to be responded to with the same action. Hence, for those robots square objects form one category and circular object form another category.

In the robots we have described categories correspond to a single activation pattern in the robot's internal units. As we consider more complex environments, however, we have to qualify this claim. Imagine a mobile (and armless) robot living in an environment that contains a large variety of perceptually different objects that have to be responded to with the same action (say, approaching and

reaching perceptually different edible mushrooms) and a variety of different objects which have to be responded to with another action (avoiding perceptually different poisonous mushrooms) (Cangelosi and Parisi, 1998). In these circumstances, we cannot expect that all the mushrooms that have to be responded to with the same action will evoke an identical activation pattern in the internal units of the robot's neural network, completely eliminating the differences among the individual mushrooms that have to be responded to with the same action. In fact, if we evolve a population of robots in this new environment and we examine the activation patterns elicited by the mushrooms in the internal units of the robot's neural network, we find that even perceptually different mushrooms that have to be responded to with the same action tend to elicit somewhat different activation patterns in the internal units. However, we can still maintain the basic assumption of the action-based conception of cognition, i.e., that different objects tend to be internally represented on the basis of the action with which they have to be responded to, rather than in terms of their purely perceptual characteristics. When the robots have (evolutionarily) learned to respond appropriately to the mushrooms, i.e., they eat the edible ones and avoid the poisonous ones, we discover that the mushrooms of one category do not evoke exactly the same internal activation pattern in the internal units of the robots' neural network. However, the internal activation patterns evoked by the mushrooms belonging to one action-defined category resemble each other and are very different from the activation patterns evoked by the mushrooms belonging to the other category. The two categories of mushrooms can be formally represented as two "clouds" of points in the abstract hyperspace of the internal units, where each point represents the internal activation pattern of an individual mushroom, each dimension of the space corresponds to one internal unit, and the position of the point on that dimension corresponds to the activation level of the unit. The points that correspond to one category of mushrooms form one cloud, and they are close to one another, and the points corresponding to the other category form another, separate, cloud.

What if when the robot encounters a mushroom, it does not only visually perceive the mushroom but it also hears the word that describes the category of the mushroom, for example the words "edible" and "poisonous"? Now the neural network that controls the robot's behavior includes both a NL sub-network and L sub-network. When the robot encounters an edible mushroom and it visually perceives the mushroom with its NL sub-network, it also hears the sound "edible" with its L sub-network, while when it encounters a poisonous mushroom it hears the sound "poisonous". (Notice that these two sounds can be produced by another robot or they can be self-produced by the robot. For the self-production of language as an important component of what we can call a robot's mental life, see Section "Robot That Talk to Themselves"). If we examine the clouds of points representing the two categories of mushrooms in the internal units of a robot's NL sub-network, we find that, compared to the robots without language, the two clouds have a smaller size and there in a greater distance between the centers of the two clouds. As a consequence, we find that the robots are better able to distinguish between the two categories of mushrooms and to avoid making errors by eating a poisonous mushroom or avoiding an edible mushroom (Mirolli and Parisi,

2005). The linguistic labeling of categories of objects makes these categories better able to support effective behavior. This appears to be an important consequence of possessing a language and may have had a crucial role in its evolutionary emergence.

Notice that in many neural network models of "semantic knowledge" (e.g., Rogers and McClelland, 2004), word meanings are identified with categories or concepts and therefore it is in principle impossible to ask the question of what might be the influence of language on categories. The neural network of our language-using robots is made up of two parts, one which is non-linguistic and the other one which is linguistic, and categories emerge in the non-linguistic part as a consequence of the non-linguistic interactions of the robot with the non-linguistic environment. Only when the linguistic part becomes operational (in children at around 1 year of age) one may pose the question of what are the consequences of possessing a language for the non-linguistic functioning of the organism.

THE EMERGENCE OF DIFFERENT TYPES OF WORDS IN THE ROBOTS' LANGUAGE

In the robots described so far the NL sub-network has a single layer of internal units that receive activation from the sensory units and send activation to the motor units, and we have seen that the pattern of activation appearing in the NL internal units encodes the action with which the robot has to respond to the sensory input and ignores the properties of the sensory input which are not relevant to decide the action. The edible mushrooms are all different from each other but this variation tends to be ignored (or minimized) by the internal units because all edible mushrooms, independently of their differences, have to be responded to with the same action: approaching and reaching the mushroom. What if NL has not one but two successive layers of internal units, with the sensory units sending their activation to the first layer, this layer sending its activation to the second layer, and the second layer sending its activation to the motor units? If we construct a robot with this type of NL, we find that, while the second layer of internal units specifies the action to be executed and ignores the properties of the perceived object which are irrelevant for the action (like in our robots with a single layer of units), the first layer of internal units preserve more of the properties of the perceived object, even if they appear not to be relevant for the action (Borghi et al., 2003).

Why might it be useful for our robots to have two layers of internal units and not only one? The robots described so far have only to approach and reach the edible mushrooms in order to eat them. But imagine another robot which has an arm and a hand and which to eat an edible mushroom has first to grasp the mushroom with its hand. The robots lives in an environment in which the edible mushrooms are of two sizes: small and large. The robot has to approach and reach both small and large mushrooms but then it has to grasp the mushrooms with its hand in order to bring them to the mouth. There are two actions of grasping. To grasp small mushrooms the robot has to produce a precision grip by using the thumb and index finger of its hand while to grasp large mushrooms the action has to be a power grip which uses all the fingers of the hand. This robot would find it useful to have two layers of internal units, one which specifies the action of approaching and reaching edible mushrooms or the action of avoiding poisonous mushrooms

and the other one which specifies the action of grasping small edible mushrooms with a precision grip or the action of grasping large edible mushrooms with a power grip.

These robots would make it possible to ask some interesting questions about the neural representation of language. As we know, the internal units of L are bi-directionally connected with the internal units of NL. But now NL has two layers of internal units, one specifying the actions of approaching and reaching the edible mushrooms and avoiding the poisonous ones, and the other one specifying the actions of producing a precision grip of the hand for small edible mushrooms and a power grip for large edible mushrooms. With which of these two different internal layers of NL will the internal layer of L be bi-directionally connected? Would the internal layer of L be bi-directionally connected with both the first and the second internal layers of NL, or would it be preferentially connected with the more perceptually abstract second layer? Notice that the answer to this question might depend on the robots' language. Small and large edible mushrooms might co-vary with two different sounds, i.e., two different nouns, that is, there might be one sound (name) for small edible mushrooms and a different sound (a different name) for large edible mushrooms. Or the robots' language might include a sound which co-varies with both small and large edible mushrooms and two other sounds which co-vary, respectively, with small and large mushrooms (and probably with other things that require a precision or a power grip). This would make it possible to being to recognize different types of words in our robots' language (say, nouns and adjectives).

A MORE SOPHISTICATED MODEL OF THE INTERNAL LAYER OF BOTH THE NL AND N SUB-NETWORKS

We have assumed so far that external input, from within the network or from the outside environment, evokes one single pattern of activation in the internal units of NL and L. But let us change the model and imagine that the internal units of both NL and L have internal (horizontal) connections that allow one activation pattern to elicit another activation pattern in the same set of units. In this manner, an external input will evoke not a single (static) activation pattern but a succession of activation patterns in the internal units of both NL and L. This is just one particular instance of a general property of brain activity which is not well captured in most neural network models: brain activity is made up of continuous processes, not states. Time is a crucial property of brain activity but it is not well captured by neural network models that conceive network activity as a succession of discrete time cycles. Objects and words should elicit processes in a neural network which at some point or another cause a response in the network's motor units, and this response will in turn cause other processes in the neural network. (We do not address here how this could be implemented in our language-using robots).

This new type of neural network for our language-using robots may help us explain one type of word associations, i.e., word-word associations. The sound of a word, represented as an action of the phono-articulatory effectors of the robot in the internal units of L, will evoke the sound of another word in the same internal units of L. Another type of word associations, based on the meaning of words and not just on their sound, can be reproduced with our robots if an activation pattern in the L internal units evokes an

activation pattern in the NL internal units which evokes a second activation pattern in the same NL internal units which in turn evokes an activation pattern in the L internal units. The first type of word associations requires a succession of activation patterns within the L internal units while this second type requires going back and forth between the internal layers of L and NL.

This more sophisticated (and, we believe, more realistic) neural network model may lead to a number of interesting conclusions concerning the nature of language, with particular reference to three issues: (a) how the meaning of words is represented in the brain; (b) the ambiguity of all words (and not only of ambiguous words); and (c) the idiomatic character of all multi-word expressions (and not only idiomatic expressions). Let us consider these three issues.

(a) There is no such thing as the meaning of a word in the brain

If an activation pattern in the L internal units evokes an activation pattern in the NL internal units which in turn evokes another activation pattern in the same NL internal units, and so on, one is led to the conclusion that there is nothing like the meaning of a word in the brain. Heard words are specific entry points to the NL sub-network but they elicit in the NL sub-network not a single activation pattern but a succession of activation patterns in many possible directions as a function of the current context and many other factors, and it is difficult to say where the process ends. The existence of a "semantic module" is assumed in many symbolic models and in many traditional, i.e., non-embodied, neural network models of language. But in the neural network that controls the behavior of our language-using robots there is no special "semantic module" which contains the "meanings" of words. The NL sub-network is the "rest of the brain" which is activated in many possible directions when one hears a word. Words do not have well-defined meanings but they are just entry points for activating the entire brain.

(b) All words are ambiguous

The more sophisticated neural network of our language-using robots should help us to explain the role of context in language understanding. We define context as any additional input, linguistic or non-linguistic, arriving from outside the brain or self-generated inside the brain, that may influence what activation patterns are sequentially elicited in the internal units of the NL sub-network. Among other things, context explains how the brain disambiguates ambiguous words. The context is an additional input that directs the activation process in the NL internal units in one direction or another. The ambiguous word "club" activates one activation pattern in the internal units of the NL sub-network in the context of golf and another activation pattern in the context of the social behavior of some people.

What is more interesting is that our model can explain the less well-recognized fact that all words are to some extent ambiguous and there is no clear dividing line between ambiguous and non-ambiguous words. For all words, the context in which a word is used directs the understanding of the word, i.e., the succession of activation patterns in the internal units of the NL sub-network, in one or another direction. For example the

word "water", which is not normally considered to be an ambiguous word, may elicit different activation patterns in the NL sub-network as a function of the particular context in which the word is used. This may be extended to the metaphorical use of words and to words that have both a literal meaning and a metaphorical meaning, such as the verb "to grasp".

(c) All multi-word expressions are idiomatic

Not only there is no clear separation between ambiguous words and non-ambiguous words but there is no clear separation between idiomatic expressions and non-idiomatic expressions. Idioms are defined as multi-word expressions whose meaning cannot be derived from the meanings of the component words (Cacciari and Tabossi, 1993). Idiomatic expressions are considered as different from non-idiomatic expressions in that non-idiomatic expressions are sequences of words which elicit an overall pattern of activation in the NL sub-network which is made up of the activation patterns elicited by the words that make up the sequence (phrase or sentence) according to some general rules (syntax). We claim that all multiple-word expressions, when they are actually used, are to some extent idiomatic, that is, they elicit an activation pattern in the NL internal units is something more than, and more specific, than the sum of the component words. (This applies even to what appear to be the simplest multi-word expressions, i.e., verb-noun expressions). Our model can provide an explanation both for idioms and for the fact that all multi-word expressions possess some degree of idiomaticity (Wray, 2002). The model should explain these different degrees (and types; cf. Wray, 2002) of idiomaticity because the overall activation pattern which is activated in the internal units of the NL sub-network when the robot arrives to the end of the sequence of heard words may be related to the activation patterns elicited by the single words of the sequence in a variety of different and unique ways.

ROBOT THAT TALK TO THEMSELVES

The neural network of our language-using robots allows us to (begin to) explain an important aspect of mental life, that is, mental life as talking to oneself (Parisi, 2007). The simple network architecture described in Section "A Neural Network's Architecture for Language-Using Robots" appears to be generally appropriate to capture how language can influence cognition, providing the basis for a Vygotskyan robotics (Mirolli and Parisi, 2009, 2010). But if we assume that the internal units of the robots' neural network have horizontal connections and these connections can produce a succession of activation patterns in both the NL and L internal units, we can see how the reciprocal connections linking the NL and L internal units can explain mental life as talking to oneself. We have seen the role of these reciprocal connections in explaining language understanding and language production. But when a non-linguistic input arrives to the sensory units of NL, for example the robot sees a cat, and the activation spreads to the internal units of NL and then to the internal units of L, two different things can happen. One is that the activation reaches the motor units of L and the robots pronounces the word "cat". The other is that the activation pattern in the internal units of L elicits another activation pattern in the same set of units, for

example the activation pattern corresponding to word "dog". This activation pattern in turn will elicit in the internal units of NL the activation pattern (or rather the succession of activation patterns) that gives a meaning to the word "dog". This is already talking to oneself. The robots hears the self-produced word "dog" and understands the word. But what is interesting is that the process can go back and forth between NL and L. An activation pattern in the internal units of NL can elicit another activation pattern in the same units and this other activation pattern can elicit an activation pattern in the internal units of L. The process can go on an indefinite number of times, as when one is immersed in his or her thoughts.

Talking to oneself really takes off when the process of going back and forth between the L and NL internal units interacts with the process of generating a succession of activation patterns in many possible directions in the NL internal units. The result of the interaction between these two processes is that the activation patterns evoked in the L internal units (words) influence and control the succession of activation pattern evoked in the NL units. This is an important component of talking to oneself as thinking.

LANGUAGE PRODUCTION DETERIORATES MORE THAN LANGUAGE UNDERSTANDING WITH AGE

A robot that has language must also be able to exhibit the rich phenomenology of pathological linguistic behaviors. By lesioning in different places and in different ways the neural network that controls both the non-linguistic and the linguistic behavior of the robot, we should be able to reproduce a variety of linguistic disorders. We will only mention here a phenomenon which is not considered as really pathological but still involves some malfunctioning of language. With old age many people find it difficult to find the word that expresses something they appear to have in their mind. This difficulty in producing the word is not normally accompanied by a parallel difficulty in understanding the word. Our model of language might be able to reproduce this asymmetry if we assume that in old age there is a gradual but diffuse loss of neurons or of connections between neurons. The model can explain both facts if we make the very reasonable assumption that the internal layer of the NL sub-network contains many more units (neurons) than the internal layer of the L sub-network. If there is a diffuse loss of units or of connections between units (including the two-way connections between the two layers of units) it might be easier for the network to go from a pattern of activation in the L sub-network (internal representation of the heard word) to the appropriate pattern of activation in the NL sub-network (understanding the word) than to go from a pattern of activation in the NL sub-network to the appropriate pattern of activation in the L sub-network (finding the word to express something one has in mind) simply because the larger sub-network is more robust than the smaller sub-network (Mirolli et al., 2007).

THE EMOTIONAL MEANING OF WORDS

Words do not only have a cognitive or informational meaning but they also have an emotional meaning. Can the neural network of our language-using robots be modified so that our robots might be able to appreciate the emotional meaning of words?

As we have said at the beginning of the paper, robots indicate the importance of the body and its movements in determining cognition so that robotics naturally converges with theories of cognition as embodied and as action-based. But both current robots and these theories have two related limitations that need to be overcome if we want to construct a more complete theory of the human mind. The first limitation is due to the fact that an organism's body does not only have an external morphology and sensory and motor organs but it also includes internal organs and systems which exist inside the body beyond the brain. The second limitation is that the mind is not only cognition but also motivation and emotion. The two limitations are related because while cognition mainly results from the interactions of the brain with the external environment, motivations and emotions mainly result from the interactions of the brain with the other organs and systems that exist inside the body.

An embodied conception of the entire mind (not just cognition) assumes two levels of functioning of the behavioral system of an animal, a strategic or motivational level and a tactical or cognitive level. All animals have many different motivations that are generally impossible to satisfy at the same time. Therefore these different motivations necessarily compete with one another for the control of the animal's behavior and at any given time the strategic level of functioning of the animal has to decide which motivation the animal should pursue with its behavior. The decision is taken on the basis of the current intensity of the different motivations, which is determined by many different factors, both intrinsic (the overall adaptive pattern of the animal and the specific environment in which the animal lives) and contextual (sensory input from the body and from the external environment). Once a decision is taken at the strategic level, the cognitive level executes the activity which will hopefully satisfy the motivation decided at the strategic level. Emotions operate at the strategic or motivational level by increasing the current intensity of one or another motivation so that the strategic level may function more effectively (fewer errors, faster decisions, increasing the persistence of important motivations, etc.). The tactical level is mostly implemented through the interactions of the animal's brain with the external environment, while the strategic level is mostly implemented through the interactions of the brain with what is inside the body. If robots should help us to develop a complete embodied theory of the mind, what is needed is an internal robotics, that is, the construction of robots that do not have only the external morphology of an animal's body and a "brain" which interacts only with the external environment but also have internal (artificial) organs and systems and a "brain" which interacts with these internal organs and systems (Parisi, 2004).

How are this more complete conception of the mind and this more complete robotics related to the construction of robots that have language? Not only so-called emotional words but all words have an emotional component that plays a role in their use and, unless we are able to endow the words used by a robot with this emotional component, we are not authorized to say that we have constructed robots that have language.

How should we proceed? The first step is to construct robots that have many different motivations and have to choose which

one of these different motivations will control the robot's behavior at any given time. Current robots tend to have just one single motivation, and this motivation is not chosen by them but by their users, that is, by us. The second step is to endow our robots with emotions (not just with the capacity to express emotions that they do not actually have, as in most current "emotional" robots). This can be done by adding an "emotional circuit" to the neural network that controls the robot's behavior, where the function of this emotional circuit is to enable the robot to make more effective and more efficient motivational choices. The emotional (neural) circuit can be activated by input from the body (e.g., hunger or thirst) or from the external environment (e.g., a predator or a possible mate) and it sends activation to the rest of the robot's neural network, influencing the motivational decision taken by the neural network and therefore the actual behavior exhibited by the robot. The emotional circuit also interacts with the rest of the robot's body, sending and receiving activation to and from internal organs (e.g., heart and gut) and systems (e.g., endocrine and immunological systems).

The first steps in this direction have already been made by constructing robots that to survive and reproduce have to both eat and drink, or to both eat and avoid being killed by a predator, or to both eat and approach a mate. The results indicate that in all cases adding an emotional circuit to the neural network that controls the robot's behavior leads to more effective behaviors and, therefore, to longer lives and more offspring. (For a detailed description of these robots, see Parisi and Petrosino, in press).

How can we extend our model of language-using robots so that our robots can understand and produce words that have an emotional component? The answer is to add an emotional circuit to the NL sub-network of our robots so that the emotional circuit can also be activated when the robot understands or produces a word. The use (understanding and production) of some words will more directly and more extensively involve the activation of the emotional circuit of the NL sub-network but all words will in one manner or other and to a greater or smaller degree activate the circuit. Adding an emotional circuit to the NL sub-network of our language-using robots will be necessary if the motivational/emotional level of behavior of our robots should be influenced by hearing both words produced by other robots and self-produced words. If we further assume that exercising our emotions in safe conditions such as those implied in exposing oneself to artistic artifacts leads to a more sophisticated motivational/emotional functioning, our emotional language-using robots might also be able to understand and enjoy poetry and other forms of verbal art as humans do. Poems and novels are verbal stimuli that to be really understood and enjoyed should activate the emotional circuit of the NL sub-network of our language-using robots.

CONCLUSION

A crucial step toward the construction of really human, and not simply humanoid, robots is to construct robots that have language. In this paper we have described a simple neural network architecture that controls the behavior of a language-using robot and we have illustrated a number of language-related phenomena that can be explained (reproduced) with our language-using robot. However, most of the work to construct robots that can be said to have language has still to be done since human language is such a complex and multi-faceted phenomenon. Language has emerged from animal-like non-linguistic communication systems, is culturally transmitted, and it changes historically. Language is learned through a succession of specific stages. Linguistic expressions are made up of simpler expressions, from morphemes to words, from phrases to sentences. Language is a crucial ingredient of human social life and it is used to accomplish a large number of different social goals. We think that all these aspects of language which are studied by a variety of scientific disciplines might be illuminated by a well-developed linguistic robotics. (For a description of the different goals of such a linguistic robotics, see Parisi and Cangelosi, 2002).

REFERENCES

Barsalou, L. W. (2008a). Grounded cognition. *Annu. Rev. Psychol.* 50, 617–645.

Barsalou, L. W. (2008b). Grounded cognition: past, present, and future. *Top. Cogn. Sci.* 2, 716–724.

Borghi, A., Di Ferdinando, A., and Parisi, D. (2003). "The role of perception and action in object representation," in *Connectionist Models of Cognition and Perception*, eds J. A. Bullinaria and W. Lowe (Singapore: World Scientific), 40–50.

Borghi, A., Parisi, D., and Di Ferdinando, A. (2005). Action and hierarchical levels of categories: A connectionist perspective. *Cogn. Syst. Res.* 6, 99–110.

Borghi, A. M., and Riggio, L. (2009). Sentence comprehension and simulations of objects' temporary, canonical, and stable affordances. *Brain Res.* 1253, 117–128.

Cacciari, C., and Tabossi, P. (eds) (1993). *Idioms: Processing, Structure, and Interpretation*. New York, Psychology Press.

Cangelosi, A., and Parisi, D. (1998). The emergence of language in an evolving population of neural networks. *Connection Sci.* 10, 83–97.

Cangelosi, A., and Parisi, D. (2001). "How nouns and verbs differentially affect the behaviour of artificial organisms," in *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, eds J. D. Moore and K. Stenning (London: Erlbaum), 170–175.

Clark, A. (1999). An embodied cognitive science? *Trends Cogn. Sci.* 3, 345–351.

Di Ferdinando, A., and Parisi, D. (2004). "Internal representations of sensory input reflect the motor output with which organisms responds to the input," in *Seeing and Thinking*, ed. A. Carsetti (Dordrecht: Kluwer), 115–141.

Fischer, M. H., and Zwaan, R. A. (2008). Embodied language. A review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol.* 61, 825–850.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Mirolli, M., Cecconi, F., and Parisi, D. (2007). A neural network model for explaining the asymmetries between linguistic production and linguistic comprehension. in *Proceedings of the 2007 European Cognitive Society Conference*, eds S. Vosniadou, D. Kayser, and A. Protopapas. Hove: Lawrence Erlbaum, 670–675.

Mirolli, M., and Parisi, D. (2005). "Language as an aid to categorization: A neural network model of early language acquisition," in *Modeling Language, Cognition, and Action. Proceedings of the 9th Neural Computation and Psychology Workshop*, eds A. Cangelosi, G. Bugmann, and R. Borysyuk (Singapore: World Scientific), 97–106.

Mirolli, M., and Parisi, D. (2009). Language as a cognitive tool. *Minds Mach.* 19, 517–528.

Mirolli, M., and Parisi, D. (2010). Towards a Vygotskyan cognitive robotics. The role of language as a cognitive tool. *New Ideas Psychol.* 1–14.

Mitchell. M. (1998). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Parisi, D. (2004). Internal robotics. *Connection Sci.* 16, 325–338.

Parisi, D. (2007). "Mental robotics," in *Artificial Consciousness*, eds A. Chella and R. Manzotti (New York: Imprint-Academic).

Parisi, D., and Cangelosi, A. (2002). "A unified simulation scenario for language development, evolution, and historical change," in *Simulating the Evolution of Language*, eds A. Cangelosi and D. Parisi (London: Springer), 255–276.

Parisi, D., and Petrosino, G. (in press). Robots that have emotions. *Adapt. Behav.*

Rogers, T. T., and McClelland, J. L. (2004). *Semantic Cognition: A Parallel Processing Approach*. Cambridge, MA: MIT Press.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.