

Embodied models and language grounding

Anna M. Borghi ^{◦Δ}

[◦] Department of Psychology, University of Bologna

^Δ Institute of Science and Technology of Cognition, CNR, Rome

Abstract

In the paper are reported experiments on the relationships between words and actions. Results show that language is grounded in perceptual and motor systems and are discussed within the framework of embodied cognitive science. I argue that models able to reproduce the experiments should be embodied organisms, whose brain is simulated with neural networks and whose body is as similar as possible to humans' body. I also claim that embodied models could help to solve some open issues. Finally, I discuss potential implications of this kind of models for theories of embodied cognition.

1. Introduction

This paper aims to argue that cognitive neurosciences need to broaden their objectives making a more extensive use of models that adequately reproduce the brain and bodily characteristics of human beings. In the introduction the main claims of embodied theories of cognition are described. Then some experimental studies on language grounding will be described, with response times or kinematics measures as dependent variables. While describing these studies I will argue that the experimental results obtained cannot be modelled without using embodied models, i.e. models that reproduce, at least in part, both the neural structure of the human brain and some of the crucial sensorimotor characteristics of human arms/hands. In addition, I will argue that embodied models might be crucial in order to disentangle some crucial issue for the field, and can help to formulate more precise predictions. In the conclusions I will discuss potential implications of the use of embodied models for embodied theories of cognition.

2. Theoretical framework: embodied theories of cognition

Within the field of cognitive science, traditional views intend concepts and words as abstract, amodal, and arbitrarily related to their referents (Fodor, 1975). In the recent years embodied views have provided a lot of evidence showing that concepts and words are rather grounded in sensorimotor processes, therefore in our bodily states. Among others, Barsalou (1999) has proposed that concepts should be conceived of as simulators. In other words, concepts are seen as re-enhancement of the pattern of neural activation recorded during perception and interaction with objects and entities. Thus, for example, the concept of “dog” would consist of the re-enhancement of the neural pattern that is active while observing a dog, caressing it, listening to its barking. This view entails that concepts are modal rather than amodal and that modality-specific systems play a central role in representing knowledge (Barsalou, 1999; Damasio, 1989; Gallese & Lakoff, 2005;

Martin, 2007). Therefore, concepts are directly grounded in the sensorimotor system, and different modality-specific areas are activated during conceptual processing.

3. Embodied models are necessary to reproduce results on language grounding

According to embodied views, not only concepts but also words are grounded, as words evoke their referents and re-enhance experiences with them. For example, the word “ball” refers to a ball, and evoke its bouncing sound, its visual and tactile properties, the experience of throwing it. According to this view, during language processing the same systems used for perceiving and acting would be activated. For example, while comprehending an action sentence we would “simulate” the situation it describes, activating the same neural substrate used in action (Barsalou, 2008; Gallese, in press; Jeannerod, 2007, Pulvermüller, 2003; Zwaan, 2004).

Borghini, Glenberg and Kaschak (2004) used a part verification procedure to explore the nature of the conceptual knowledge retrieved when people use words to think about objects. Participants read sentences (e.g. There’s a car in front of you) followed by nouns; their task consisted in deciding by pressing a key whether the noun referred to an object part or not. Parts were chosen so that actions directed toward them (on the real object) require movement upward (e.g., the roof of a car) or downward (e.g., the wheels of a car). Orthogonally, responding “yes” required an upward or a downward movement to a response button. Responding in a direction incompatible with the part’s location (e.g., responding downward to verify that a car has a roof) was slow relative to responding in a direction compatible with the part’s location.

This suggests that processing of words implies activating the motor system (see also Borghini, 2005). Along the same line, an increasing body of evidence indicates that during language comprehension we make use of the same neural systems used for perception, action and emotion. Since the seminal paper by Rizzolatti and Arbib (1998) on the relationship between language and motor system, a number of studies on canonical and mirror neurons have shown that these neurons

might provide the neural basis underlying the language comprehension mechanism (Gallese, in press).

In the previous study the effect of language on motor system was investigated using reaction times and movement times. In another study we *directly* addressed if the mere act of comprehending language affects the production of action, focusing on body kinematics parameters (Scorolli, Borghi & Glenberg, under review). Participants listened to sentences referring to the lifting of light or heavy objects (e.g., pillow or chest, respectively). Then they lifted one of two boxes that were visually identical, but one was light and the other heavy. After the box lifting, participants are asked a comprehension question (e.g. “Is the object on the table soft?”), in order to verify whether they have listened and comprehended the sentence. The kinematics of the body movements was recorded using a motion capture system. We focused on the kinematics of the initial lift (rather than reaching) because it is mostly shaped by proprioceptive features derived from weight that cannot be visually determined. If language did not have any effect on motor system, changes in biomechanical parameters should have been determined only by the actual object weight. Instead we found that listening to sentences activates a simulation which is sensitive to fine-grained object properties such as weight. The two most important results are based on analyses of hand and arm delay (latencies of first peak velocities) immediately after grasping the box. We found that participants’ time delay was larger when the weight implied by the sentence and the weight of the box they lifted were similar compared to when they were dissimilar. These results are consistent with the operation of the MOSAIC model of action control (Hamilton, Wolpert, & Frith, 2004). According to MOSAIC, the force used in an action arises from integrating the force parameters from several modules that might apply in the situation (e.g., modules for lifting a light vs. a heavy box). The integration is based on the estimated probability that a module applies in the situation. Modules may be rendered temporarily unavailable by simultaneous use in another task, and that this produces a type of repulsion effect. In our work language comprehension could serve as Task 1 and render modules unavailable when Task 2 consists of judging the weight lifted by another. For the

Light Box, the force generated after the light sentence is further from the required force than the force generated after reading a heavy sentence. Just the opposite obtains for the Heavy Box.

3.1. Which sort of models?

How could we model and reproduce the results of the experiment I have illustrated? Most common models of language comprehension, developed within the cognitivist tradition, are based on association frequency (for a more detailed critique of these models, see Borghi & Cimatti, submitted; Louwerse, & Jeuniaux, in press). One of the most influential models, Latent Semantic Analysis (Landauer & Dumais, 1997), explains word meaning in terms of the associations between one word and other words in large corpora. The higher the index of co-occurrence of words in similar texts, the higher their similarity in meaning. Even if it represents a very useful tool, this model fails when it claims to represent conceptual meaning formation. Namely, it does not take into account that words are grounded in our sensorimotor system, as it only considers the network of verbal associations in which words are embedded. For these reason a model like this cannot capture and cannot predict the fact that an upper part is processed faster when moving upwards than when moving downwards. Namely, these results cannot be simulated with a model that does not possess at least some features of human sensorimotor system – some form of proprioception, some kind of sensitivity to sensory inputs. In addition, the simulated organism should possess at least some features of the human motor system (e.g. reaching and grasping “devices” like an arm, a hand with fingers, etc.). For the experiment in which kinematics measures were used, the simulated body should be quite complex, in order to reproduce a lifting movement. In addition, this model should be endowed with the capacity to comprehend language by referring words to its sensorimotor experience.

Only an embodied model can reproduce the experimental results I described. Namely, reproducing the results of an experiment does not simply mean to model just a behavior performed

by a decontextualized brain simulated through a neural network, but to reproduce the behavior of an organism endowed not only with a brain but also with a body. In an ideal condition, this model should be able to reproduce learning through a mechanism of weight selection that avoids a-priori hardwiring any inhibitory or excitatory connections between or within modules. Consider that with the term embodied models I don't refer to generic mathematical models in which the output is obtained on the basis of equations between input and output variables. Rather, I refer to *neural networks models*, as these models could offer more possibilities than the traditional mathematical models to reproduce the functional and physiological aspects of the brain. Some well known characteristics of the real nervous system as robustness, flexibility, generalization, recovery based on the content, learning, parallel processing, can be reproduced by neural networks models and sometimes also by some mathematical models. However, using neural networks models it is possible to obtain an emergent behaviour that leads the artificial brain to auto-organize its different parts, exactly as it happens in the real brain. Finally, the elaboration of information in the real nervous system is distributed, as there are many neurons involved in the same operation and a single neuron can be involved in different operations at the same time or at different times. Neural networks models well reproduce this distributed elaboration of information. At the same time, feed-forward models probably do not provide an adequate formalization for embodied theories. Namely, a great lesson of embodied theories concerns the reciprocal and circular relationship between perception, action, and cognition. The problem with feed-forward model is that they risk reproducing the traditional "sandwich" of disembodied cognitive sciences, in which perception and action are intended merely as peripheral parts, having a scarce influence on the most tasty part, the inside, that is cognition.

Using an embodied model has two further advantages. First, it allows one to observe rather than to infer the results. Finally, it allows reproducing the real structure of the experiment we aimed to simulate – for example, the actual button reaching behaviors could be reproduced.

However, even if it is clear that using embodied models has many advantages, the recent literature on robotics shows that there are substantial differences in physical embodiment between robots and humans. Therefore, it is important that models have at least some basic requisites. They should be endowed with at least some characteristics of human reaching and grasping system, and their sensorimotor system should at least be roughly similar to the human one. In addition, it is important that models possess at least some characteristics of human neural systems. For example, studies on concepts and language grounding might have a neurophysiological basis in the recent discovery, first in monkey and then in humans, of two kinds of visuomotor neurons: canonical and mirror ones (see Gallese, Fadiga, Fogassi & Rizzolatti, 1996). In synthesis: an appropriate model of experiments such as the reported ones should be an embodied model, endowed with at least some crucial characteristics of human sensorimotor system and neural structure (neural network), and it should be able to replicate the behavioural results found (see as an example Caligiore, Borghi, Parisi & Baldassarre, in preparation).

4. Embodied models can help to formulate clearer predictions

So far I have shown that embodied models are necessary to model experimental results on language grounding. Now I intend to show that embodied models might be very useful in order to formulate more detailed predictions that might help to disentangle unsolved issues within cognitive science. In the last years a hotly debated issue concerns the kind of relationship existing between language and the motor system. I will label it the “Interference or Facilitation” (IF) issue. An example will help to clarify what it’s all about.

Consider the following study (Scorolli & Borghi, 2007; see also Borghi & Scorolli, in press). Participants were presented with pairs of nouns and verbs that could be referred to hand and mouth actions (e.g., to unwrap vs. to suck the sweet), or to hand and foot actions (e.g., to throw vs. kick the

ball). An equal number of non-sensible pairs were presented. Participants' task consisted of deciding whether the combination made sense or not: 20 participants responded by saying yes loudly into a microphone, 20 by pressing a pedal. Results suggest that sentence processing activates an action simulation. This simulation is quite detailed, as it is sensitive to the effector involved. Namely, it leads to a facilitation in responses to 'mouth sentences' and 'foot sentences' compared to 'hand sentences' in case of congruency between the effectors – mouth and foot – involved in the motor response and in the sentence.

However, why should we predict that, say, reading a foot sentence leads to facilitation in foot responses rather than to interference? If during language processing and action executing the same neural structures are activated, then this might slow down responses. In the literature studies performed with tasks that slightly differ from the one I described report an interference effect (Buccino, Riggio, Melli, Binkofski, Gallese, & Rizzolatti, 2005).

The IF issue consists in the fact that it is difficult predicting whether an interference or a facilitation effect will occur. The problem is further complicated by the fact that both interference and facilitation are compatible with an embodied account. Namely, they both indicate that language is grounded, and that reading sentences leads to an activation of the motor system. However, so far it is difficult to make accurate predictions about the direction of the interaction between language and motor system. Models can help to further detail the predictions, contributing to solve the IF issue.

5. Tentative conclusions and open issues

Aim of this paper is to show that embodied theories badly need models, and that these models need to be embodied. In the first part I illustrated some behavioral experiments indicating that reading words and sentences activates the motor system. The results of these experiments can only

be reproduced with embodied models, i.e. with embodied artificial organisms that are endowed with neural and sensorimotor structures that at least roughly reproduce human ones.

In the second part I illustrated behavioural experiments leading to ambiguous results. As an example, I illustrated the “interference vs. facilitation” issue. Embodied models can be a powerful means that helps to disentangle ambiguous issues and to formulate clearer predictions. The embodied cognition field has greatly expanded in the last years. In the last ten years much experimental evidence has been collected, but now it is crucial to formulate more detailed and precise predictions. Embodied cognition field badly needs for well specified theories, and models can help to formulate these theories.

One last issue is worth of notice. At a very general level, embodied models might provide a powerful way to test embodied theories of cognition. Namely, comparing models whose physical (neural, sensorial, motor) structure is more or less similar to the human one, will allow understanding to what extent possessing the same kind of “body” is necessary in order to understand the world and to comprehend and to use language.

From a theoretical point of view, a strong embodied view could predict that only models that share the same bodily characteristics with the entities they have to reproduce (human beings) can be adequate to explain them. From this claim might derive the choice to consider the artificial world as a parallel world, with its own laws, that should (and could) not be compared with the world of human beings. Here I adopt a milder embodied view. I argue that a certain degree of similarity between humans and their embodied models will probably allow capturing important aspects of human cognition and behaviour. One of the fascinating questions the research of the next years has to face is the following: to what extent do we need to be similar in body in order to share a common view of the world, and to communicate with others?

References

- Barsalou L. W. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences*, 22, 577-609.
- Barsalou, L.W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59, 617-645.
- Borghi, A.M. (2005). Object concepts and action. In D. Pecher & R.A. Zwaan (Eds). *Grounding Cognition: The role of perception and action in memory, language, and thinking*. Cambridge: Cambridge University Press.
- Borghi, A.M., & Cimatti, F. (submitted). Words are not words. For an extension of embodied theories of language and cognition.
- Borghi, A.M., Glenberg, A., & Kaschak, M. (2004). Putting words in perspective. *Memory and cognition*, 32, 863-873.
- Borghi, A.M., Scorolli, C. (in press). Language comprehension and hand motion simulation. *Human Movement Science*.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, 24, 355-63.
- Caligiore, D., Borghi, A.M., Parisi, D. & Baldassarre, G. (in preparation). Affordances and compatibility effect: a computational model.
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25–62.
- Fodor J. (1975). *The language of Thought*. Harvard University Press, Cambridge (MA).
- Gallese, V. (in press). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*.

- Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593-609.
- Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensorimotor system in conceptual knowledge. *Cognitive Neuropsychology*, *21*, 455-479.
- Hamilton, A., Wolpert, D., & Frith, U. (2004). Your Own Action Influences How You Perceive Another Person's Action. *Current Biology*, *14*, 493-498.
- Jeannerod, M. (2007). *Motor cognition. What actions tell to the self*. Oxford: Oxford University Press.
- Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.
- Louwerse, M., & Jeuniaux, P. (in press). Language comprehension is both embodied and symbolic. In M. De Vega, A.M. Glenberg, & A.C. Graesser (Eds.). *Embodiment and meaning: A debate*. Oxford: Oxford University Press.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25-45.
- Pulvermüller, F. (2003). *The neuroscience of language: On brain circuits of words and serial order*. Cambridge: Cambridge University Press.
- Rizzolatti, G., & Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*, 188-194.
- Scorilli, C., & Borghi, A. (2007). Sentence comprehension and action: Effector specific modulation of the motor system. *Brain research*, *1130*, 119-124.
- Scorilli, C., Borghi, A. & Glenberg, A. (submitted). Effects of language on bimanual object lifting.

Zwaan, R. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B.H. Ross (ed.), *Psychology of learning and motivation*, (vol.44, pp.35-62). New York: Academic.