Sentence comprehension and action: Effector specific modulation of the motor system.

Claudia Scorolli & Anna M. Borghi

Claudia Scorolli, Dipartimento di Discipline della Comunicazione, Bologna

Anna Borghi, Dipartimento di Psicologia, Bologna

Text pages: 14 pages.

Figures: 2.

Address for correspondence:

Claudia Scorolli

Dipartimento di Discipline della Comunicazione

Via Azzo Gardino, 23

40122 Bologna

Phone: +39-051-2092218

Email: cscoroll@dsc.unibo.it.

Abstract

The purpose of the paper is to study whether sentence comprehension modulates the motor system. Participants were presented with 24 pairs of nouns and verbs that could be referred to hand and mouth actions (e.g., to unwrap vs. to suck the sweet), in the first block, or, in the second block, to 24 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 support embodied theories of language comprehension, as they suggest that sentence processing activates an action simulation. This simulation is quite detailed, as it is sensitive to the effector involved. Namely, it lead 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.

Section

Sensory and Motor Systems

Keywords

Effectors, modulation, motor system, language, embodiment.

Abbreviations

Response Times: RT

Analysis of Variance: ANOVA

1. Introduction

Recent proposals in cognitive science and neuroscience claim that cognition is embodied. The embodied view claims that knowledge is not abstract but grounded in sensorimotor experiences, and that there is a deep unity among perception, action and cognition (Thelen & Smith, 1994; Pecher & Zwaan, 2005). This view of cognition contrasts with the classical perspectice, according to which the mind is a mechanism for manipulating arbitrary and amodal symbols. In the classical cognitive view, concepts are seen as being inherently non-perceptual: perceptual states arise in sensory-motor systems, but perceptual experience will be transduced in a completely new representational language. The resulting symbols do not correspond with the perceptual states that produced them. They are, therefore, amodal, and the link between the concept and the perceptual state is just arbitrary.

Instead, in the embodied view concepts are not conceived of as being given by arbitrary and amodal symbols but rather by perceptual symbols. These perceptual symbols are neural representations located in sensory-motor areas in the brain: there is no transduction process (Barsalou, 1999). More precisely, concepts consist of the reactivation of the same neural activation pattern that is present when we perceive the objects or entities they refer to and when we interact with them (Barsalou, 1999, Gallese & Lakoff, 2005, Glenberg, 1997). For example, the concept of *dog* refers to a real or an imagined dog and, when encountered, reactivates any previous experiences with this extra-linguistic entity. In this view, object attributes are thought to be stored near the same modality-specific neural areas that are active when objects are being experienced (Martin, Ungerleider & Haxby, 2001). Moreover, symbols, according to the embodied view, are not amodal, but multimodal – for example, they refer both to the tactile experience of caressing a dog as well as the auditory experience of hearing a dog bark (Barsalou, 1999; Gallese & Lakoff, 2005).

Contemporary neuroscience provides evidence in support of the claim that concepts make direct use of sensory-motor circuits of the brain (Gallese & Lakoff, 2005). There is much neural evidence to indicate, for example, that the same areas are involved when forming motor imagery and when activating information on objects, particularly on tools. For example, evidence gathered through Positron Emission Tomography indicates that the naming of tools, as opposed to the naming of animals, differentially activates the left middle temporal gyrus, which is also activated by action generation tasks, and the left premotor cortex, generally activated when participants imagine themselves grasping objects with their dominant hand (Martin, Wiggs, Ungerleider, & Haxby, 1996). Along these same lines, fMRI studies have shown that the premotor left cortex responds selectively to images of tools, but not to images of animals and houses (Chao & Martin, 2001; see also Grafton, Fadiga, Arbib & Rizzolatti, 1997).

An important consequence of this embodied view concerns language, as it makes use of concepts. According to the embodied theory there is no ‘language module’. Instead, language makes direct use of the same brain structures used in perception and action. Understanding language implies forming a “simulation”, that is the recruitment of the very neurons that would be activated when actually acting or perceiving the situation, action, object or entity described by language (Barsalou, 1999; Gallese & Lakoff, 2005; Gibbs, 2003; Glenberg, 1997; McWhinney, 1999; Zwaan, 2004). There is much behavioural evidence in support of the role simulation plays in sentence comprehension (e.g., Borghi, 2004; Glenberg & Kaschak, 2002; Kaschak, Madden, Therriault, Yaxley, Aveyard, Blancerd & Zwaan, 2005; Zwaan & Taylor, 2006). For example, Zwaan, Stanfield, and Yaxley (2002) presented participants with two kinds of sentences (*the ranger saw the eagle in the sky* vs. *the ranger saw the eagle in the nest*), followed by a picture of an object. Participants were required to indicate whether or not the object in the picture was the same object mentioned in the sentence by pressing a different key on a keyboard. The authors found an advantage in the congruent condition, in contrast with the predictions of a classical amodal vision. Stanfield and Zwaan (2001) found similar results when they investigated the effect of the orientation of the objects in visual images presented to subjetcs participating in an experiment investigating the role of simulation in sentence comprehension. When participants were presented with a sentence like *John put the pencil in the drawer*, their response times were faster when recognizing a horizontally-oriented pencil than when recognizing the same pencil presented vertically. The opposite was true in the case of a sentence like *John put the pencil in the cup*. Glenberg and Kaschak (2002) asked participants to indicate whether or not a sentence made sense by pushing one of two buttons whose position entailed either moving toward the body or away from the body. Response times were longer when responding by pushing the button that required a movement in the opposite direction from that implied by the sentence. For example, participants were faster in responding that *Close the drawer* made sense when pushing the proper button entailed moving away from the body rather than toward it. The simulation activated while processing a sentence that referred to objects’ movement seems to be quite detailed, as it contains directional information. Recently Kaschak and Borraggine (in press) replicated the study by Glenberg and Kaschak (2002) in order to investigate timing effects during sentence processing. They manipulated the delay between the acoustic sentence presentation and the visual cue that triggered the response. This cue indicated whether that the participant should press a button located near or far from the body (towards vs. away movement) in order to respond “yes”. The visual cue could come at the beginning of the sentence presentation or after it (delay of 0, 50, 500, or 1000 ms). The compatibility effect between action and sentence (ACE) was present only when the motor instruction was presented simultaneously with the beginning of the sentence rather than after the sentence presentation. This suggests that the simulation process takes place when participants can plan their motor response while processing the sentence.

Even though the reported evidence suggests that during sentence comprehension we activate simulations, the extent to which these simulations are specific is still a matter of debate. In our work we sought to investigate the degree of specificity of these simulations. More specifically, we sought to understand whether reading sentences related to actions to be performed with different effectors (mouth and foot) activates the same neural systems activated during the effective execution of these actions. Though behavioural in nature, our study has relevant implications for physiological and neural models of the relationships between language and the motor system.

Participants were presented with pairs of nouns and verbs that referred to 'hand actions' and 'mouth actions', in the first block, or to 'hand actions' and 'foot actions', in the second one. They were asked to decide whether the combination made sense. Half of them indicated their responses by using a microphone, half by pressing a pedal. 'Hand sentences' were used as a baseline.

The rationale is as follows: if the simulation is specific, that is, if the same neurons are recruited while understanding an action sentence as while performing an action with a specific effector, then 'mouth sentences' should be processed faster than 'hand sentences' when responding with the microphone than with the pedal. Similarly, 'foot sentences' should be processed faster than 'hand sentences' when responding with the pedal than with the microphone.

2. Results

All incorrect responses were eliminated. As the error analyses revealed that there was no speed-for-accuracy tradeoff, we focused on the RT analyses. To screen for outliers, scores 2 standard deviations higher or lower than the mean participant score were removed for each participant.

The remaining response times were submitted to two different mixed factor ANOVAs, one for each block ('hand sentences' vs. 'mouth sentences'; 'hand sentences' vs. 'foot sentences'). The factors of each ANOVA were Sentence Modality ('hand' vs. 'mouth' for the first analysis; 'hand' vs. 'foot' for the second one) and Response Modality (microphone vs. pedal), with Response Modality as a between participants variable.

In the block 'mouth-hand', participants responded 84 ms more quickly with the pedal than with the microphone, F (1, 38) = 12.39, MSe = 11322.55, p < .001. The advantage of the pedal over the microphone (87 ms) was present also in the block ‘foot-hand’, F (1, 38) = 14.74, MSe = 10167.86, $p < .0005$. In addition, in the block ‘foot-hand’ we also found a significant effect of the main factor Sentence Modality, with ‘foot sentences’ 21 ms faster than ‘hand sentences’, F = 17.98, MSe = 482.52, $p < .0001$.

Insert Figure 1 and 2 about here.

Further analyses were performed in order to better understand the results. We performed four separate ANOVAs, one for each Response Modality (microphone vs. pedal) and for each block (‘hand sentences’ vs. ‘mouth sentences’; ‘hand sentences’ vs. ‘foot sentences’).

The first two ANOVAs performed on participants who responded with the microphone confirmed the hypotheses advanced. As predicted, participants using the microphone responded with significantly greater speed to ‘mouth sentences’ than to ‘hand sentences’, F (1,19) = 8.28, MSe = 377.65, $p < .009$. The difference between ‘foot sentences’ and ‘hand sentences’ was less marked, F (1,19) = 5.45, MSe = 405.84, $p < .05$. Even though the last difference also reached significance, the marked difference between the effect sizes (p <.009 vs. p <.05) confirms that the simulation is effector-specific.

The ANOVAs performed on participants who used the pedal as their responding device showed that there was no significant difference between ‘mouth sentences’ and ‘hand sentences’, that is, between sentences referring to effectors not involved while using the device, 2 ms, F (1,19) = 0.0056, MSe = 559.41, $p < .81$. Instead, as predicted, we found that response times were significantly faster, 26 ms, for ‘foot sentences’ than for ‘hand sentences’, F (1,19) = 12.84, MSe = 559.21, $p < .002$.

3. Discussion

The results support the view that the act of comprehending sentences leads to the creation of an internal simulation of the action read. This simulation seems to be fairly specific, as it leads to a different modulation of the motor system depending on the effector (hand, mouth, foot) necessary for performing the actions described by the sentence. This suggests that the same motor areas are recruited whether a person is understanding action sentences or actually performing the action. Importantly, this modulation occurred even with a task in which the information related to the involved effector was really irrelevant, such as the evaluation of the sensibility of sentences. Our

results clearly show that ‘mouth sentences’ were processed faster than ‘hand sentences’ when participants were responding with the microphone rather than with the pedal. The same facilitation effect was obtained with ‘foot sentences’ compared to ‘hand sentences’ when participants were responding with the pedal rather than with the microphone. Even though our study clearly suggests that an internal simulation occurs, our results do not permit us to definitively determine when this process takes place because we recorded reaction times after the appearance of the noun. Namely, the motor resonance effect could occur either during sentence comprehension or after the sentence has been understood in order to prepare for action. Data from Borreggine and Kaschak (in press) suggest that the ACE effect, at the very least, was due to the simultaneous occurrence of a motor preparation phase and sentence comprehension. However, to our knowledge there has been no systematic study on the influence of timing on effector-specific effects in sentence comprehension. In order to solve this complex matter, more detailed studies on the relationship between timing and effector specific effects on sentence comprehension are needed. Evidence on timing could provide stronger support to the idea that a simulation process is necessary and not just epiphenomenal in order to understand the sentence (Boulenger, Roy, Paulignan, Deprez, Jeannerod, Nazir, in press).
Our results are also in line with previous fMRI studies showing that listening to sentences expressing actions performed with the mouth, the hand and the foot produces activation of different sectors of the premotor cortex, depending on the effector used in the listened sentences (Tettamanti, Buccino, Saccuman, Gallese, Danna, Scifo, Fazio, Rizzolatti, Cappa, & Perani, 2005). Of particular significance, they represent a behavioural extension of these results.
Some may object to the results of our study on the grounds that the advantage of ‘foot sentences’ over ‘hand sentences’ is significant not only with the pedal but also with the microphone. However, this effect does not go against our main hypothesis – that is, that the effector used to respond facilitates responses to sentences implying the same effector – for a number of reasons. First, the effect is much stronger with the pedal than with the microphone, as the comparison of the effect sizes demonstrates. Second, ‘foot words’ have wider cortical distributions compared to ‘mouth words’, that have a more narrow distribution (Pulvermüller, 2005). This can easily account for the slight asymmetric result we found.
Our results are in line with recent neurophysiological and behavioural evidence. Pulvermüller, Härle and Hummel (2001) recorded neurophysiological (they calculated event-related current source densities from EEG) and behavioural responses (reaction times and errors) to verbs referring to actions performed with the face, the arms and the legs. They found topographical differences in the brain activity patterns generated by the different verbs in a lexical decision task, starting from 250 ms after word presentation. The behavioral experiment indicated that response times were

shorter for face-related words compared to leg-related words, whereas the arm-related words were in the middle.

Our study represents both an extension and a modification of the results attained by Pulvermüller et al. First of all, the study by Pulvermüller et al. focused on verb comprehension, whereas the purpose of our research is to study whether understanding simple sentences composed of a transitive verb and a noun activates the motor system. In addition, the kind of task we used implied access to semantic knowledge, unlike the study by Pulvermüller et al., who used a lexical decision task on verbs. Importantly, however, we used a task for which the information pertaining to the kind of effector involved in the action described was not relevant. Given that Pulvermüller et al. found a significant difference between face-related and arm-related verbs on the one hand and leg-related verbs on the other – with manual responses –, we decided to compare 'mouth sentences' and 'foot sentences' and to use the 'hand sentences' as a baseline. Moreover, instead of employing a manual response, we used either a 'mouth response' or a 'foot response' (microphone and pedal). Namely, our purpose was to directly test whether or not understanding a sentence directly involves the motor system, affecting motor responses with the effector referred to by the sentence.

Another recent study using both transcranial magnetic stimulation and a behavioral paradigm provides evidence for a modulation of the motor system depending on the effector referred to by action sentences. Buccino, Riggio, Melli, Binkofsky, Gallese and Rizzolatti (2005) presented three kinds of sentences: hand action, foot action and abstract content related sentences. Participants were required to respond with the hand or the foot if the verb was concrete and had to refrain from responding if the verb was abstract. Results showed that if subjects responded with the same effector necessary for executing the action described by the sentence, they were slower than if they had to respond with the other effector. Although this study shows that the meaning of the sentence modulates motor system activity, the authors found an inhibition rather than a facilitation. Even though our study investigates the difference between 'foot actions' and 'mouth actions' and Buccino et al. (2005) study the difference between 'foot actions' and 'hand actions', further differences between the two behavioural studies may account for the result. The first is the modality used to deliver the stimuli. In our experiment, participants had to read the sentences, whereas in the study by Buccino et al. (2005), stimuli were acoustically presented. Furthermore, the stimuli were not the same. More importantly, Buccino et al., also used a task that implied a higher depth of processing than lexical decision, as we did, but they required the participants to evaluate the action described rather than the meaning of the whole sentence. This is clearly implied by the fact that they gave the "go" signal to respond in coincidence with the second syllable of the verb, when the noun hadn't yet been presented. On the contrary, we recorded response times from the noun presentation, and focused on comprehension of the sentence rather than of the verb alone. This explanation is in line with recent experiments on language and motor resonance that have shown that the timing between linguistic stimulus and motor response is crucial (e.g., Borraggine & Kaschak, in press; Zwaan & Taylor, 2006). In addition, in our task the information relating to the effector is really irrelevant, given that we asked participants to evaluate whether the sentence made sense and didn't require them to focus on the verb meaning.

In conclusion, our results clearly show that understanding action sentences implies an effector specific modulation of the motor system, suggesting that a simulation effect takes place. This modulation leads to a facilitation of responses in case of congruency between the effector – mouth and foot – involved in the motor response and the effector involved in the sentence.

4. Experimental procedure

Participants

Forty students of the University of Bologna took part in the experiment. All were native Italian speakers, right-handed, and all had normal or corrected-to-normal vision. They all gave their informed consent to the experimental procedure. Their ages ranged from 18 to 29 years old.

Materials

Materials consisted of word pairs (sentences) composed of a transitive verb and a concept noun. There were two different blocks: hand – mouth sentences, hand – foot sentences. For each block, we chose 12 nouns which refer to objects of daily use, each preceded by an action verb. In the first block (block mouth – hand sentences), verbs could refer either to an action usually performed with the mouth (e.g., *to suck the sweet*), or with the hand (e.g., *to unwrap the sweet*). In the second block (block foot – hand sentences), verbs could refer to an action usually performed with the foot (e.g., *to kick the ball*) or to an action typically performed with the hand (e.g., *to throw the ball*).

We decided to use two blocks because of the difficulty in finding triads of verbs that could be combined with the same noun, referred to actions with the three different effectors and had the same association rate. For example, we usually act with an object like an *ice cream* with the hand or the mouth, but not with the foot; similarly we typically interact with an object like a flower, *daisy*, with the hand or the foot, but not with the mouth. For this reason, the first block contained nouns that could be combined with both ‘mouth verbs’ and ‘hand verbs’, while the second block contained nouns that could be combined with both ‘foot verbs’ and ‘hand verbs’.

A pre-test was performed before the experiment in order to be sure that the verb-noun pairs had the same association rate in the two conditions. We required 18 subjects to produce the first five nouns they associated to each verb. Then we checked whether the noun we had chosen to associate with the verb of the critical pairs was present among the nouns they produced, and in which position it was produced. Then we calculated the weighted means for each participant, taking into account whether the noun was produced or not and, if it was produced, its production order. The weighted means of the productions for each participant were submitted to two different mixed factor ANOVAs, one for each block of sentences (‘hand sentences’ vs. ‘mouth sentences’; ‘hand sentences’ vs. ‘foot sentences’). The results showed that there was no significant difference in production means between ‘mouth sentences’ and ‘hand sentences’, F (1,11) = 0.22, MSe = .09, p = .65, and between ‘foot sentences’ and ‘hand sentences’, F (1,11) = 0, MSe = .05, p = 1. This means that our results could not be explained on the basis of the degree of association between verb-noun pairs that we had chosen.

At last we obtained 48 verb-noun pairs balanced for association rate. In addition to the critical pairs, we added 272 filler pairs. 40 were sensible verb-noun pairs – abstract sentences (e.g. *to dream the summer*). The remaining 232 were non sensible verb-noun pairs – false sentences (eg. *to switch off the shoe*) –. Each pair was presented four times in one of the two blocks.

Procedure

Participants were randomly assigned to one of two groups. Members of both groups were tested individually in a quiet laboratory room. They sat on a comfortable chair in front of a computer screen and were instructed to look at a fixation cross that remained on the screen for 500 ms. Then a verb appeared on the screen. After 200 ms it was substituted by a noun, which was preceded by a determinative article. For each verb-noun pair, participants were instructed, if the combination made sense, to say *yes* loudly (first group) or to press a pedal with the right foot (second group), and to avoid responding if the combination did not make sense. Each noun was presented in the two different combinations, that is, preceded by a verb of ‘mouth action’ or ‘hand action’ in the first block, and by a verb of ‘foot action’ or ‘hand action’ in the second block. The timer started operating when the concept noun appeared on the screen, in order to avoid problems related to length and frequency of the noun, so in the response times analyses we compared each noun (e.g., *candy*) with itself. All participants were informed that their response times would be recorded and were invited to respond as quickly as possible while still maintaining accuracy. Stimuli were presented in a random order. Sixteen training trials preceded the experimental trials, in order to allow the participants to familiarize with the procedure.

Acknowledgments

Thanks to Art Glenberg, Lucia Riggio, Giovanni Buccino, Roberto Nicoletti and Stefano Ghirlanda for useful discussions of these data. Thanks to Ann Gagliardi for help with English.

References

1. Barsalou, L. W., 1999. Perceptual Symbol System. Beavioral and Brain Sciences, 22, 577-609.

2. Borghi, A.M., Glenberg, S.M., Kaschak, M.P., 2004. Putting words in perspective. Memory and cognition, 32, 863-873.

3. Borreggine, K.L., Kaschak, M.P., (in press). The Action-sentence Compatibility Effect: It's All in the Timing. Cognitive Science.

4. Boulenger, V., Roy, A. C., Paulignan, I., Deprez, V., Jeannerod, M., Nazir, T.A. (in press). Cross talk between language processes and over motor behaviour in the first 200 msec of processing. Journal of Cognitive Neuroscience.

5. 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.

6. Chao L.L., Martin A. Representation of manipulable man-made objects in the dorsal stream. NeuroImage, 12, 478-484, 2000.

7. Gallese, V. & Lakoff, G, 2005. The brain’s concept: the role of the sensorimotor system in conceptual knowledge. Cognitive Neuropsychology, 21, 455-479.

8. Gibbs, R.W.,2003. Embodied experience and linguistic meaning. Brain and Language. 84, 1-15.

9. Glenberg, A. M., 1997. What Memory is for. Behavioral and Brain Sciences, 20, 1-55.

10. Glenberg, A. M., & Kaschak, M. P., 2002. Grounding language in action. Psychonomic Bulletin & Review, 9, 558-565.

11. Grafton, S. T., Fadiga, L., Arbib, M. B., & Rizzolatti, G., 1997. Premotor Cortex Activation during Observation and Naming of Familiar Tools. Neuroimage 6, 231-236.

12. Kaschak, M.P., Madden, C.J., Therriault, D.J., Yaxley, R.H., Aveyard, M., Blanchard, A.A., Zwaan, R.A., 2005. Perception of Motion affects Language Processing. Cognition, 94, B79-B89.

13. Martin A, Chao, L.L. Semantic memory and the brain: Structure and processes. Current Opinion in Neurobiology (Cognitive Neuroscience). 11, 194-201, 2001.

14. Martin, A., Ungerleider L.G., Haxby J. V., 2000, Category-specificity and the brain: The sensory - motor model of semantic representations of objects. In The Cognitive Neurosciences, 2nd Edition. Gazzaniga MS (Ed), Cambridge: MIT Press, pp. 1023-1036.

15. Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V., 1996. Neural correlates of category-specific knowledge. Nature 379, 649-652.

16. McWhinney, B. 1999, The emergence of language from embodiment. In The emergence of language, B. MacWhinney, London: Erlbaum, pp. 213-256.

17. Pecher, D., & Zwaan, R.A., 2005, Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking, Cambridge University Press.

18. Pulvermüller, F., 2005. Brain mechanism linking language and action. Nature Publishing Group, 6, 576-582.

19. Pulvermüller, F., Härle, M., Hummel, F., 2001. Walking or talking? Behavioral and electrophysiological correlates of action verb processing. Brain and language, 78, 143-168.

20. Stanfield, R.A. & Zwaan, R.A., 2001. The effect of implied orientation derived from verbal context on picture recognition. Psychological Science, 12, 153-156.

21. Tettamanti, M., Buccino G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S.F., Perani, D., 2005. Listening to action related sentences activates fronto-parietal motor circuits. J Cogn. Neurosci., 17, 273-281.

22. Thelen, E, Smith, L.B., 1994, A dynamic systems approach to the development of cognition and action. Bradford Book series in cognitive psychology. MIT Press, Cambridge.

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

24. Zwaan, R.A., Stanfield, R.A., Yaxley, R.H., 2002. Language comprehenders mentally represent the shape of objects. Psychological Science, 13, 168- 171.

25. Zwaan, R.A. & Taylor, L.J., 2006. Seeing, Acting, Understanding: Motor Resonance in Language Comprehension. Journal of Experimental Psychology: General, 1, 1-11.