Meaning in Words, Gestures, and Mental Images

Paolo Bernardis (paolobernardis@units.it)
Department of Psychology, University of Trieste
Via S. Anastasio, 12. 34134 Trieste Italy

Nicoletta Caramelli (nicoletta.caramelli@unibo.it)
Department of Psychology, University of Bologna
Viale Berti Pichat, 5. 40127 Bologna Italy

Abstract

In this study, we provide evidence for a cross-modal interaction between the meaning of pantomimes and words when the visuo-spatial and perceptual information of these last is enhanced. We recorded behavioral and electrophysiological responses with a cross-modal repetition priming. Pantomimes of objects and actions were used to prime visually presented nouns and verbs with an image formation task. The behavioral results showed that the image formation times of words primed by a preceding gesture were faster in the matching meaning condition than in the mismatching one. Electrophysiological results confirmed the interaction between gesture and word meanings showing a N400 localized all over the scalp with a peak on the left anterior hemisphere. Overall, these results support the idea of a tight interplay between the meaning of pantomimes and words when perceptual information is enhanced in words at both the behavioral and neurophysiological levels.

Keywords: Image formation; Gestures; Words; Semantic priming; Event-related potentials.

Introduction

Recently, researchers' interest has focused on the facilitation effect of iconic and co-verbal gestures on a number of cognitive tasks. Considerable evidence has been collected showing that the function of the gesture system cannot be reduced only to a mere support to the verbal system, providing lexical access in a tip-of-the-tongue state, nor simply to an aid in understanding language (see Butterworth & Hadar, 1989; Hadar & Butterworth, 1997; Krauss, Chen, & Chawla, 1996). Instead, it has been convincingly shown that gesture is deeply rooted in overall cognition, as it helps to highlight stages of learning (Alibali, & DiRusso, 1999; Church & Goldin-Meadow, 1986; Perry, & Elder, 1996; Pine, Lufkin, & Messer, 2004), problem solving strategies (Alibali, Bassok, Solomon, Syc, & Goldin-Meadow, 1999), how attention is directed (Goodwin, 2000) and memory improvement (Feyereisen, 2006).

Developing hints by McNeill's (1992) that gesture reflects the imagistic mental representation activated at the moment of speaking, Kita (2000) advanced the Information Packaging Hypothesis. In this perspective, the specific characteristics of the two systems complement each other. The visuo-spatial and holistic character of gesture and the segmented and linear character of language are combined in conveying complex meaning. Gesture helps speakers package spatial information into units appropriate for verbalization. In this view, the gesture system has been shown to cooperate with the language one in the conceptual planning of the message to be verbalized (Alibali, Kita, & Young, 2000; Hostetter, Alibali, & Kita 2006).

This tight interaction between the gesture and language systems was also found to characterize language comprehension and evidence thus far collected suggests that speech and gesture establish a highly integrated system of communication (Beattie & Shovelton, 1999; Cassell, McNeill, & McCullough, 1999; Clark, 1996; Goldin-Meadow, Wein, & Chang, 1992; Kelly, Barr, Church, & Lynch, 1999; Kelly & Church, 1998; Krauss, 1998; Krauss, Morrel-Samuels, & Colasante, 1991; McNeill, 1992).

This conclusion attained on behavioural grounds has been confirmed and complemented with converging electrophysiological data. The full understanding of a message in the brain is the result of a qualitatively similar elaboration of integrated types of information in a dynamic large scale neural network. This integration process involves information from world-knowledge, co-speech gestures, pictures, speaker's identity derived from voice characteristics and information from a preceding discourse (for a review, see Willems & Hagoort, 2007).

The Event Related Potentials (ERPs) technique has been largely used to investigate the processing of meaning. It is widely agreed that one particular component, the N400, is a general index of semantic integration that is yielded by verbal and pictorial stimuli presented both uni-modally and cross-modally (e.g. Willelms et al. 2008). The N400 distribution over the scalp can change depending on the type of the stimuli. Verbal stimuli usually produce a parietal distribution (Kutas and Hillyard, 1980, 1984) and pictorial stimuli a frontal N400 distribution (West and Holcomb's 2002; Ganis & Kutas, 2003). Hamm, Johnson and Kirk (2002), instead, found that with word-picture pairs the d-N400 was localized in the parietal areas (the d-N400, or difference wave, is the result of a subtraction between semantically congruent and incongruent ERPs).

In this framework, a previous study aimed at showing the interactions between iconic gestures (pantomimes) and visually presented words showed an interference effect between pantomimes and words with matching meanings instead of the expected priming effect at the behavioral level (Bernardis, Salillas & Caramelli, 2008). On the contrary, the pattern of the electrophysiological activation clearly showed a frontally distributed N400 highlighting that the meaning of pantomimes and words interacted. The same lack of priming between gestures and words at the behavioral level and their integration indexed by the N 400 was also shown with different tasks in a recent study similar to this one (Wu and Coulson, 2007).

This study will focus on this difference between the behavioral and neurophysiological results obtained in the study of the interaction between gestures' and words' meanings.

The rationale behind the present research is that at the behavioral level a priming effect should be found in the matching meaning condition when pantomimes and words access the meaning system with the same type of information, i.e. visuo-spatial and perceptual information. Accordingly, it is possible to suppose that the interplay between the meanings of gestures and words can be better highlighted on both behavioral and neurophysiological grounds when perceptual information is enhanced in words.

In order to increase the activation of the visuo-spatial and perceptual information about the referent of the target word, in the following experiments participants had to form a mental image of it before responding. In this experiment, we used the same materials and conditions as those in Experiment 2 of the Bernardis et al. (2008) study. Unlike in the previous study, participants had to watch the priming gesture, read the target word and form a mental image of its referent. We expected that in the same meaning condition the pantomime and the visually imagined word would activate the same visuo-spatial information, thus producing a repetition priming effect.

Experimental sessions

In order to correctly check for the priming effect, we started with the preliminary assessment of the baseline image formation times of all the stimuli (nouns, verbs) subsequently used in the experiment. In addition, the imageability value of both nouns and verbs was assessed. Then the experiment was carried out to collect both behavioral and neurophysiological data.

Preliminary assessment

We collected baseline image formation times for the all the stimuli (40 nouns and 40 verbs). Nineteen students volunteered their participation sitting in front of a computer screen. The participants had to silently read the words, which were presented one at a time, and form a mental image of their content. When the image was clear in their mind, the participants had to press the response button. Their image formation times were measured from the appearance of the word on the screen to the moment when they pressed the button. Mean image formation times for nouns and verbs were respectively 2633 ms (s.e. 106 ms). and 2976 ms (s.e. 132 ms).

The imageability value of the 40 target words in the same meaning condition was collected from another group of 23 participants. They had to rate on a 7 point scale how easily each word presented on a sheet of paper in 2 different random orders was imaged. The median imagery value was 5.8 (s.e. 0.35) with nouns (6.85, s.e. 0.31) more easily imaginable than verbs (4.75, s.e. 0.38) [FriedmanAnova (20, 1) = 17, $p < 0.001$].

Experiment

The experiment was aimed at providing behavioral and electrophysiological evidence of the interaction between the meanings of pantomimes and words while recording brain activity with the ERPs technique.

Method

Participants. Fifteen students at the University of Trieste participated for course credit. All of them were Italian mother tongue and did not take part in any of the other studies in this research. Because of a large number of artefacts, the data from one participant were excluded from the analyses. Thus, the final data set for the analyses was collected from 14 participants.

Materials. The materials were the same as those in the Bernardis et al. (2008) study (for the selection criteria of both gestures and words and concreteness, familiarity and length assessments, see Bernardis et al., 2008).

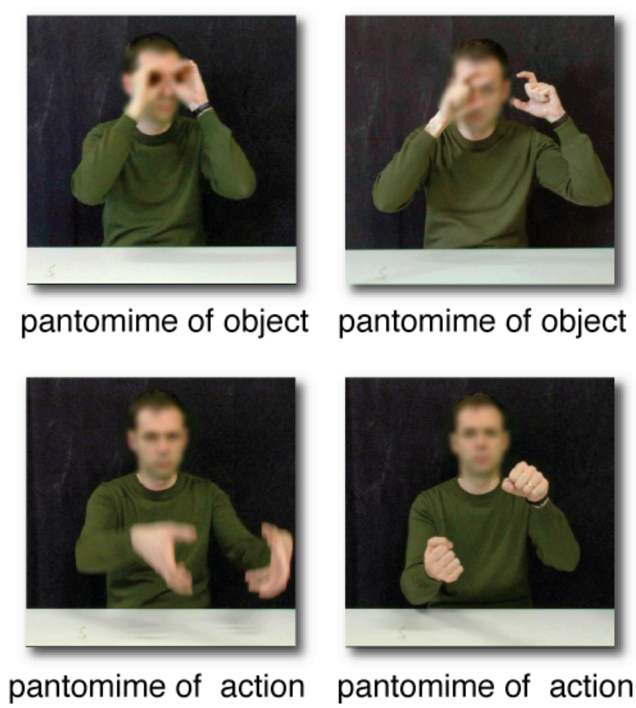

Figure 1. Four frames extracted from the pantomime video-clips of object nouns (binocular and camera) and action verbs (to lift and to drive).

To summarise there were 2 categories of stimuli:

- 40 pantomimes out of which there were:
  - 20 pantomimes of objects expressed by nouns, e.g. "binoculars" (see fig. 1);
  - 20 pantomimes of actions expressed by verbs, e.g. "to knock" (see fig. 1).
- 80 words, already checked for in the preliminary assessmen, out of which there were:
  - 40 words (nouns and verbs) to be used as targets in the matching meaning condition. These words were the same used to name the pantomimes in a preliminary study (see Bernardis et al., 2008);
  - 40 words (nouns and verbs) to be used as targets in the mismatching meaning condition. These words were unrelated in meaning to the 40 words in the matching meaning condition.

Procedure. Each participant, sitting in front of a computer screen, was asked to watch the pantomime video-clip, to silently read the word that followed, to form a mental image of its referent, and then to press the response button. The stimuli were displayed on a 19" LCD screen placed 80 cm in front of the participants. A central fixation cross (500 ms) preceded the pantomime video-clips used as prime (average duration 3627 ms). Then, a black screen (200 ms), used to record the baseline for the following ERPs analysis, preceded the presentation of the words (500 ms). The inter-trial interval was 1000 ms. The random presentation of the stimuli and blocks were controlled automatically with Presentation Software (Neurobehavioral System, Inc.).

The pairs of video-clip and word were arranged in a balanced way so that each participant was presented with all the video-clips in the two matching and mismatching meaning conditions. Each participant was presented with the stimuli twice. The order of blocks was counterbalanced across participants and the list of the stimuli was randomised within each block for each participant. Each block of trials lasted approximately 5 minutes and there were short breaks between the blocks. The experimental session was preceded by a practice trial to familiarise the participants with the task, the equipment and the materials (4 pairs of gesture-word in the two conditions).

ERPs Data Acquisition. EEG was recorded from 28 scalp electrodes mounted on an elastic cap. Following the standard International 10-20 system, the electrodes were located at the midline (Fz, Cz, Pz, Oz), medial (FP1, F3, FC3, C3, CP3, P3, O1, FP2, F4, FC4, C4, CP4, P4, O2), and lateral brain areas (F7, FT7, T3, TP7, T5, F8, FT8, T4, TP8, T6). These recording sites plus an electrode placed over the right mastoid were referenced to the left mastoid electrode.

The data were recorded continuously throughout the task by a SynAmps (NeuroSoft) amplifier and software NeuroScan 4.3.1. Each electrode was re-referenced off-line to the algebraic average of the left and right mastoids. Impedances of these electrodes never exceeded 5kΩ. The horizontal electro-oculogram (HEOG) was recorded from a bipolar montage with electrodes placed 1 cm to the left and right of the external canthi. The vertical electro-oculogram (VEOG) was recorded from a bipolar montage with electrodes placed beneath and above the right eye to detect blinks and vertical eye movements. The EEG and EOGs were amplified by a SynAmps amplifier with a band pass of .01-30 Hz, filtered for 50 Hz and digitised at 500 Hz. Trials containing ocular or movement artefacts or amplifier saturation were corrected from averaged ERP waveforms.

For the analysis of the lateral electrodes, five different electrodes were chosen for each of four ROI (Region of Interest). Left Anterior: F3, F7, FC3, FT7, C3, Left Posterior: CP3, TP7, P3, T5, O1, Right Anterior: F4, F8, FC4, FT8, C4, and Right Posterior: CP4, TP8, P4, T6, O2. For the analysis of the midline electrodes, there were four different electrodes (Fz, Cz, Pz, Oz).

We assessed the effects of words' meaning between prime and target (matching vs. mismatching) and of their grammatical class (nouns vs. verbs) by measuring the mean amplitude of ERPs. The ERPs were time-locked to word onset and based on a 100 ms pre-stimulus baseline. The time window was chosen according to both visual inspection of the distribution of the waves on the scalp and literature (Pulvermüller, Lutzenberger, & Preissl, 1999; Federmeier, Segal, Lombrozo, & Kutas, 2000; Willems, Özyürek, Hagoort, 2008). Segments were averaged for each condition for each participant at each electrode site.

Data Analysis and Results

Behavioural data analysis. After removing outliers from each participant's data (less than 1% of all the responses), the mean response times were calculated. Two ANOVAs were performed, one with subjects and the other with materials as random factors. Their results are presented together. Both the analyses included one between-subject factor (Experiment: baseline vs priming) and two within-subjects factors (Relation (matching vs. mismatching meaning) between pantomime and word; Grammatical Class (noun vs. verb). In both the analyses we found a significant effect of the interaction between the factors Relation and Experiment [F (part) (1,36) = 6.26, MSE = 5209430, p = .017; F (mat) (1,36) = 13.10, MSE = 4238191, p = .0008]. Post-hoc analysis (Neuman-Keuls, p<.05) revealed that in the priming experiment the image formation time was faster in the matching meaning condition (2165 ms; SE = 48 ms) than in the mismatching meaning condition (2316 ms; SE = 47 ms). No differences were found between the matching and the mismatching condition in the baseline assessment, and between the baseline assessment and priming experiment in the mismatching condition. In addition, nouns were imagined faster than verbs [F (part) (1,36) = 14.87, MSE = 2347028, p = .0005; F (mat) (1,36) = 7.22, MSE = 3340044, p = .011].

ERPs data analysis. Several repeated-measures ANOVAs were performed on the mean activity from the lateral electrodes. The factors were: Relation between pantomime and word (matching vs. mismatching meaning), Grammatical Class (nouns vs. verbs), Hemisphere (left vs. right), Localization (anterior and posterior ROIs), and

Electrodes. On the ERPs data from the midline electrodes, different ANOVAs were performed in the same time window, the factors of which were Relation between pantomime and word (matching vs. mismatching meaning), Grammatical Class (nouns vs. verbs), and Electrodes. In all the ANOVAs performed on the ERPs data, the Greenhouse-Geisser correction was applied when the sphericity assumption was violated.

The visual inspection of the grand-average waveforms (see Fig. 2) clearly showed a N1, P2 and N300 components. In the N1 component, there was no clear difference between the conditions, in the P2 time window the difference between the conditions (matching vs mismatching meaning) started to be detectable. The N300 component was present in the central and frontal electrodes, with major peaks on the frontal ones. At 300 ms a negativity component resembling the N400 started. This negative component, present only in the mismatching meaning condition, reached the peak near 450 ms and finished at 550 ms with the highest peaks in the central and anterior electrodes. Conversely, in this time interval the matching meaning condition was positive or almost at zero. Lastly, we observed a Late Negativity component. For the purposes of this study we will focus only on the N400 time window. The latency range chosen for the analyses was 300-550 ms.

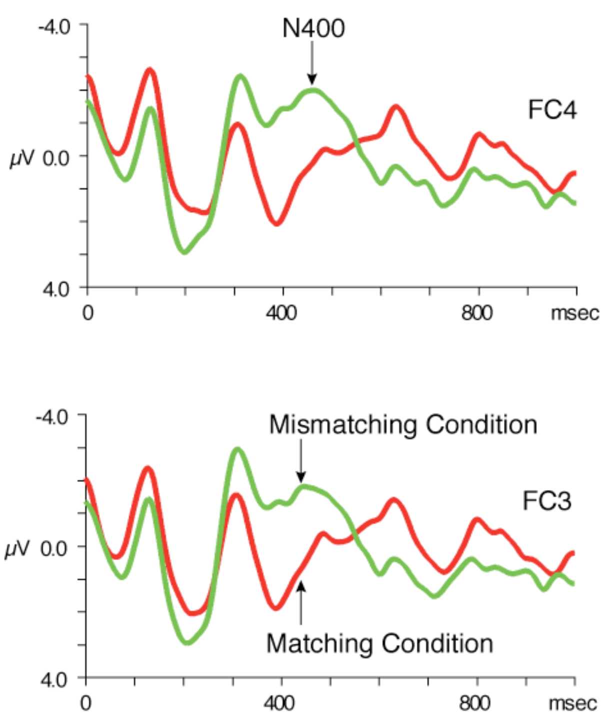

Fig. 2. Grand average ERPs for the conditions with matching and mismatching meaning pantomimes at electrodes FC3, FC4. ERPs were time locked to the word onset. Negativity is plotted upward.

The N400 time window (300-550 ms) . The ANOVA on the data from the midline electrodes showed a significant main effect for the factor Relation [F (1,13) = 75.84; MSE = 288.19; p < .001], with mismatching meaning words producing the N400 component. The ANOVA on the data from the lateral electrodes showed a significant main effect for the factors Relation [F (1,13) = 103.72; MSE = 1353.21; p < .001)] and Localization [F (1,13) = 23.30; MSE = 292.87; p < .001], and the interaction between Relation, Hemisphere and Localization [F (1,13) = 4.76; MSE = 1.37; p = .048]. The largest values of the N400 for the mismatching meaning condition was localised in the anterior quadrants with no significant difference between the hemispheres (see Fig. 2, 3).

The distribution over the scalp of the N400 (see Fig. 3) was assessed with an ANOVA on the values resulting from the difference between the waves elicited by the matching and mismatching meaning words to the preceding gesture. The factors were Grammatical Class, Hemisphere, Localization, and Electrodes. A significant effect was found for the interaction of Hemisphere x Localization [F (1,13) = 4.76; MSE = 2.75; p = .048]. As shown in Figure 5, a pair wise comparison in the anterior quadrants revealed a higher N400 in the right hemisphere compared to the left [F (1,13) = 5.35; MSE = 10.52; p = .037].

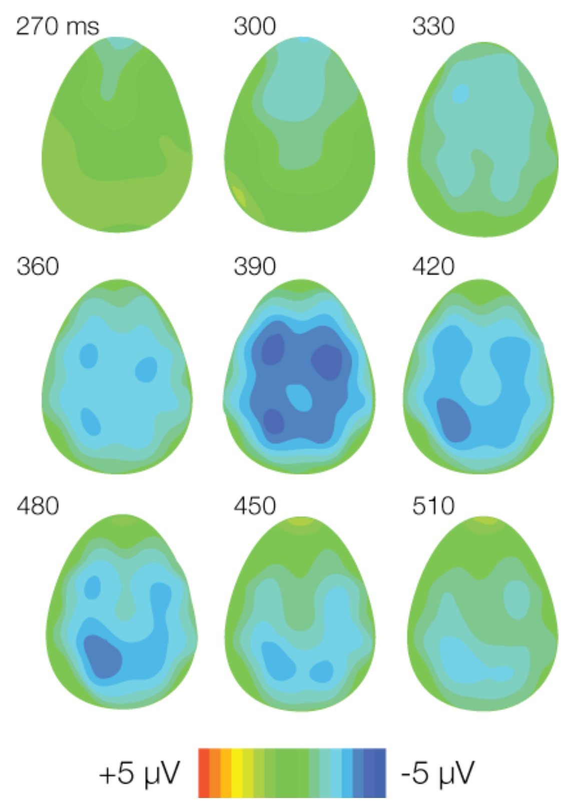

Fig. 3. Distribution over the scalp of the N400 component, obtained subtracting the matching from the mismatching meaning condition.

Discussion

The behavioral results showed the expected repetition priming effect of pantomimes on the image formation time of the words' referents in the matching meaning condition.

Accordingly, we can conclude that the meanings of words and gestures can interact, yielding a repetition priming effect. This happens when the visuo-spatial and perceptual information is enhanced in words, as is the case in the pantomimes in the matching condition.

This finding was also confirmed by the electrophysiological results. As in Experiment 2 of the Bernardis et al. (2008) study, the highest peak appeared in the right anterior quadrant. However, in this study, the distribution over the scalp of the N400 component was wider than that obtained previously, reaching also posterior regions, as found in the classical research with verbal stimuli. This may also suggest that the interaction does not depend only on the formed image per se, but on the enhanced perceptual information in the words' meaning, as imagery processes are related to activity in the occipital areas with an overall involvement of the left hemisphere (Farah, Weisberg and Monheit, 1989).

In addition, this enhancement modulated the distinction between nouns and verbs, which was not found in our previous experiment with a naming task.

This evidence, along with the electrophysiological results of our previous study, support the idea of a tight interplay between the meaning of pantomimes and words when the visuo-spatial and perceptual information of words' meaning is enhanced.

Acknowledgments

This research was supported by a research grant of the University of Bologna (FRO-2007) to the second author. Many thanks to Carlo Semenza for allowing the use of his equipment for recording of ERPs.

References

Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E., & Goldin-Meadow, S. (1999). Illuminating mental representations through speech and gesture. *Psychological Science*, *10*, 327–333.

Alibali, M., & DiRusso, A. (1999). The function of gesture in learning to count: more than keeping track. *Cognitive Development*, *14*, 37-56.

Alibali, M., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*, 593-613.

Beattie, G., & Shovelton, H. (2000). Iconic hand gestures and the predictability of words in context in spontaneous speech. *British Journal of Psychology, 91*, 473-91.

Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology*, DOI: 10.1080/02643290801921707.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: a reply to McNeill. *Psychological Review, 96*, 168-74.

Cassell, J., McNeill, D., & McCullough, K. E. (1999). Speech–gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition, 7*, 1–34.

Church, R., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*, 43-71.

Clark, H. (2005). Coordinating with each other in a material world. *Discourse Studies, 7*, 507-525.

Farah, M., Weisberg, L., Monheit, M., & Peronnet, F. (1989). Brain Activity Underlying Mental Imagery: Event-related Potentials During Mental Image Generation. *Journal of Cognitive Neuroscience, 1*, 302-316.

Federmeier, K., Segal, J., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain: a Journal of Neurology, 123*, 2552-66.

Feyereisen, P. (2006). How could gesture facilitate lexical access? *Advances in Speech-Language Pathology, 8*, 128 - 133.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research , 16*, 123-44.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction, 9*, 201–219.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*, 1489–1522.

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica , 115*, 147-172.

Hamm, J., Johnson, B., & Kirk, I. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology, 113*, 1339-50.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, *22*, 313-336.

Kelly, S. D., Barr, D., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language, 40*, 577–592.

Kelly, S., & Church, R. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development, 69*, 85-93.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.) *Language and Gesture* (pp. 162–185). Cambridge, UK: Cambridge University Press.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological* Science, 7(2), 54–60.
Krauss, R., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology, 28*, 389-450.
Krauss, R., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology, 61*, 743-54.
Kutas, M. & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203-5.
Kutas, M. & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161-3.
McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.
Perry, M. & Elder, A. D. (1996). Knowledge in transition: Adults' developing understanding of a principle of physical causality. *Cognitive Development*, *12*, 131–157.
Pine, K., Lufkin, N., & Messer, D. (2004). More gestures than answers: children learning about balance. *Developmental Psychology, 40*, 1059-67.
Pulvermüller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex, 9*, 497-506.
West, W., & Holcomb, P. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13*, 363-75.
Willems, R. M. & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, *101*, 278-89
Willems, R. M., Ozyürek, A. & Hagoort, P. (2008). Seeing and Hearing Meaning: ERP and fMRI Evidence of Word versus Picture Integration into a Sentence Context. *Journal of Cognitive Neuroscience, 20*, 1235-1249.
Wu, Y.C., & Coulson, S. (2007). Iconic gestures prime related concepts: An ERP study**,** *Psychonomic Bulletin & Review, 14* (1)*,* 57-63.