Semantic priming between words and iconic gestures

Paolo Bernardis (bernardis@psico.units.it)
Scuola Superiore di Studi Umanistici
Bologna, Italy

Nicoletta Caramelli (nicolletta.caramelli@unibo.it)
Dipartimento di Psicologia
Bologna, Italy

Abstract

The interaction between words and gesture meanings was highlighted in two experiments with a priming paradigm. The results converge in showing the peculiarities of the two meaning systems, thus supporting the Information Packaging Hypothesis (Kita, 2000) against the Lexical Retrieval Hypothesis (Butterworth & Hadar, 1989).

Introduction

Since the pioneering studies of Bellugi and Brown (1964) and Kendon, (1982), the variety of the dimensions underlying the complex structure of gesture has provided a rich taxonomy according to which each type of gesture is characterized by specific properties depending on the function it accomplishes.

The present research deals with a particular type of gesture, the so-called ‘iconic gesture’ (McNeill, 1992), or, better, ‘pantomime’ (Kendon, 1982) and, more generally, ‘representational gesture’. This type of gesture is produced when people move their arms and hands to produce a dynamic visual representation of the properties of the objects or events they want to communicate. Thus, for instance, when somebody raises his/her right arm bringing his/her closed hand, eventually with the thumb and the little finger protracted in opposite direction, close to his/her ear, we understand that s/he is referring to something about a telephone call. While shared knowledge and contextual cues aid the unambiguous identification of the entire message (e.g., ‘I’ll call you’ or ‘Call me up, please’), this gesture unambiguously refers to a telephone call in absence of speech.

Iconic gestures can be produced both with and without speech and can vary in their degree of conventionality, nonetheless they are straightforwardly understood. The transparency of the meaning of iconic gestures raises the problem of their relationship to language with which they often co-occur. Even if the language and the gesture systems greatly differ, because speech is segmented and linear while gestures convey information all at once as they rest on the visuo-spatial medium instead of the verbal one, they share the function of conveying meaning. Since McNeill’s (1992) idea that iconic gestures undergo semantic processing by listeners, as both language and gesture provide complementary meanings, much supportive evidence has been provided that has developed his view mainly in two complementary directions. One, pragmatically oriented, holds that gesture reflects the variations in speech acts (Kendon, 2000) and that the gestural and the linguistic systems interact in a flexible manner depending on the communicational intent of the speaker (Holler, & Beattie, 2003). The other, instead, holds that language and gesture dynamically interact in shaping thought, expanding on McNeill’s (1992) suggestion that gesture reflects the imagistic mental representation activated at the moment of speaking. Gestures are no more conceived of as a mere communication device subservient the production and comprehension of language. Gesture, instead, has been acknowledged to help modulating cognition as shown by the fact that, for example, it can 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), and how attention is directed (Goodwin, 2000). Thus, recent research has focused on the deep interplay between gesture and cognitive activity stressing the pervasive influence of gesture on cognition and thought, overcoming the Lexical Retrieval Hypothesis. According to this hypothesis, iconic gestures do not carry meaning by their own, but derive it from the lexicon (Butterworth & Hadar, 1989) or simply facilitate accessing lexical entries that incorporate syntactic and semantic information (Krauss, Chen, & Chawla, 1996). Instead, in the Information Packaging Hypothesis (Kita, 2000), gesture helps speakers package spatial information into units appropriate for verbalization and, thus, plays a relevant role in the conceptual planning of the message to be verbalized (Alibali, Kita, & Young, 2000; Hostetter, Alibali, & Kita, 2006). Stressing the interaction between language and gesture in shaping thought, this view grants gesture’s meaning a greater autonomy than that offered by the Lexical Retrieval Hypothesis, according to which gestures play a direct role only in the process of speaking as they simply embody spatially encoded knowledge. In the Information Packaging Hypothesis, gestures convey meanings that can be independent of those conveyed by language, as shown by the finding that incongruent concurrent gestures can negatively affect the processing of speech (Kelly, Kravitz, & Hopkins, 2004). Moreover, studying the effects of representational gestures on memory, Feyereisen (2006) has highlighted that it is the meaningfulness of gesture that is responsible for memory improvement and not gesture by
itself, as gestures that are incongruent with the sentence meaning, do not facilitate recall.

The studies reported thus far, however, have considered gestures produced, or understood, concomitantly with language production, or understanding. They did not consider gesture’s meaning outside a linguistic context where gestures assume the burden of expressing meaning by themselves. As Golding-Meadow (2005) has pointed out, gesture changes its function and form when it is produced as a complement to language or on its own. As she has remarked, in this last case gesture takes on the discrete and segmented form characteristic of all linguistic systems becoming ‘language-like’ as she puts it. That is, gestures have a different structure depending on the linguistic functions they stand for.

Aim of the present research is precisely to assess on line the interaction between the meaning of gesture and that of language. If the meaning of words can prime that of gestures the Information Packaging Hypothesis, with its emphasis on the cognitive planning of the content to be expressed by the two systems, will be verified. Otherwise, if the meaning of gesture can prime that of words, the Lexical Retrieval Hypothesis will be verified. In fact, if an effect of priming can be obtained between words and gestures outside any communicative context, this means that the language and the gesture systems are characterized by two independent meaning systems that can match or mismatch depending on the situation. Thus, the complex and articulated messages, which characterize the communicative settings, can be conceived of as the result of the deep interaction of the language and the gesture systems at the conceptual level in planning the meanings to be conveyed by the two modalities.

Moreover, studying deaf children, Goldin-Meadow, Butcher, Mylander, & Dodge (1994) have identified different types of iconic gestures corresponding to the different functional roles of words such as the referential function of nouns and the predicative and commenting function of verbs. Thus, it is also possible to suppose that these same functional roles of words can affect the priming effect in both words and gestures differently.

In order to verify these hypotheses, two experiments were carried out. The first was aimed at verifying whether words with four different functional roles could prime gestures having the same meaning as words. The second was aimed at verifying whether gestures could prime words with four different functional roles having the same meaning as gestures. The words’ functional roles considered were: simple reference to an object (e.g., eye glasses), reference to an object usually used to accomplish an action (e.g., glass), a simple action (e.g., to knock), and an action to be performed with a tool (e.g., to cut). Thus, the words in the two first functional roles were nouns, while the ones in the second two functional roles were verbs.

**Experiment 1**

The hypotheses of Experiment 1 were:

- If the language and the gesture systems share meaning at the cognitive level, then words should prime the recognition of the meaning of pantomimes.

- If the gesture system, when used without language, shares the functional roles of linguistic categories, such as those expressed by nouns and verbs, with that of language, then different types of reference (reference to a simple object, to an object usually used to accomplish an action, to a simple action and to an action to be performed with a tool) should interact with the priming of pantomimes.

In order to verify whether words can prime pantomimes, apt materials had to be checked and selected from the materials that were devised by the experimenters only on intuitive grounds. Thus, two different pre-tests were performed. The first pre-test was performed on gestures and the second on words.

**Gesture Assessment**

The experimenters identified 50 pantomimes representing both objects and actions according to the referential function of nouns and verbs. Four types of referential functions where devised: to simple objects, like “tie” (13 pantomimes), to objects strictly linked to an action, like “guitar” (12 pantomimes), to actions to be performed directly on objects, like “to knock” (13 pantomimes), and to actions requiring a tool to be performed, like “to write” (12 pantomimes). For each of the 50 pantomimes a short video-clip was prepared in which the half body of an actor appeared while gesturing with both arms and hands. The face of the actor was blurred so that the participants could focus only on the pantomime. The duration of each pantomime, i.e., of each video-clip, varied from a minimum of 2320 ms to a maximum of 4680 ms. Each video-clip started with the actor with his hands placed on a table in a fixed position, and beginning the movement after 280 ms (7 frames). The actor ended the pantomime with his arms and hands in the same position as at the beginning. The video-clips were projected on a 19” monitor with cathodic ray tube filling the entire surface.

Twenty students at the University of Trieste volunteered for their participation in this pre-test. They were presented with the video clips one at a time, while sitting on a chair 80 cm distant from the monitor. From this distance, the vertical dimension of the actor’s half-body was approximately 15 degrees of visual angle. Their task was to watch the video-clip and to name the pantomimes presented while one of the experimenters recorded the words used to name them. From the set of the 50 video-clips presented, 40 pantomimes (10 for each type) were retained to be used in the following experiments as all the participants labeled them with the same noun or verb.

In order to better assess the results of the following experiments, a baseline of the selected gestures recognition time was collected. Twenty more students at the University of Trieste volunteered for their participation in this assessment. They were presented with the 40 video-clips of pantomimes (10 for each type) that were retained from the naming task to use in the following experiments. The description of the apparatus for pantomime presentation and response collection is the same as that described in the procedure section of Experiment 1. This time participants were asked to name the pantomime on the video-clip as soon as they could recognize it. The mean pantomime
recognition time was 2339 ms (sd = 462 ms). Two ANOVAs were performed one with subjects and the other with material as random factors, the results of which will be presented together. The ANOVA main factor was the referential function of the pantomime at four levels (reference to a simple object, to an object usually used to accomplish an action, to a simple action and to an action to be performed with a tool). The analysis showed a significant difference between the four referential functions (Fpart (3,57) = 49.583, Mse = 1.305, p < 0.0001; Fmat (3,36) = 3.144, Mse = 0.591, p < 0.0369). Pantomimes representing simple objects were the slowest to recognize (mean = 2645 ms; sd = 291 ms), while those representing objects usually used to accomplish an action was the fastest to recognize (mean = 2024 ms; sd = 230 ms). The recognition time for the pantomimes of the two other referential functions, i.e., simple action and action to be performed with a tool was 2305 ms (sd = 265 ms), and 2380 ms (sd = 327 ms) respectively.

As the video-clips of the 40 pantomimes varied in length, a regression analysis between the pantomimes’ duration and their recognition time was performed in order to prevent any possible bias. The results did not show any correlation between these two factors (t = 1.63, p = 0.11).

Fig. 1: Examples of the four different types of pantomimes.

Word Assessment. In addition to the 40 words used to name the pantomimes by all the participants in the Gesture Assessments, 40 other words were added (10 for each referential function) that were unrelated to those used to name the gestures. Two independent groups of fifteen students each, who did not take part in any other part of the research, were presented with the list of the 80 words thus obtained. The first group was asked to rate each word on a 7-point Likert scale for concreteness and the other group for familiarity.

The mean concreteness value for the words naming the pantomimes was 5.33 (± 0.78) and for the unrelated words was 4.96 (± 0.82). The mean familiarity value for the words naming the pantomimes was 4.57 (± 1.14), and for the unrelated words 5.06 (± 0.91). The mean syllable length for the words naming the pantomimes was 3.1 (± 0.2) and for the unrelated words 3.3 (± 0.1).

In order to better assess the results of the following experiments, a baseline of the reading times of the 80 words (20 for each referential function) used in the following experiments, was collected. Twenty more students at the University of Trieste volunteered for their participation in this assessment. Each of them sat in front of a computer screen on which each word appeared one at a time. Participants were asked to read the word aloud. The reading times were collected from the onset of the word to the beginning of its reading. Two ANOVAs were performed on the words’ reading times, one with subject and the other with items as random factors. No significant difference between the four conditions of the factor referential function was found. Mean words’ reading times were 461 ms (sd = 49 ms), 465 ms (sd = 56 ms), 449 ms (sd = 45 ms), and 460 ms (sd = 40 ms) for reference to a simple object, to an object usually used to accomplish an action, to a simple action and to an action to be performed with a tool respectively.

Method

Participants Twenty students at the University of Trieste volunteered for their participation in the experiment receiving course credit for their participation. All of them were fluent Italian speakers (12 female; mean age = 23 years) and did not take part in any of the others studies.

Materials The materials were the 40 (10 for each referential function) video-clips of the pantomimes selected in the pretest to be used as targets and the 80 words (20 for each referential function, 10 out of which naming the pantomimes and 10 unrelated to them) already checked for familiarity, concreteness and length to be used as primes.

Procedure Each participant, sat in front of a computer screen, was asked to attend to the screen, to read the word and to name the following pantomime as soon as possible. On the screen, a central fixation cross (500 ms) preceded the word used as prime (500 ms) and a beep-sound (200 ms), used as trigger for the following analyses, signaled the starting of the pantomime video-clip (average duration 3627 ms). The inter-trial interval was 1000 ms.

The randomization and the presentation of the stimuli were controlled automatically by a routine written with Matlab Software (The MathWorks, Inc.) A headset microphone
plugged into a PC recorded both the beep sound and the voice of the participants. Each of the 40 video-clips was paired to 2 words, one naming the pantomime and the other unrelated. The pairs of word and video-clip were arranged in a balanced way so that participants were presented with all the video-clips in the two conditions (same vs. different meaning). A practice trial (four items) was run to familiarize participants with the equipment and materials before the experimental session.

Data analysis and results
The dependent variable analyzed was response time. Response times were calculated as the difference between the onset of the video-clip and the beginning of the response recorded by the emission of the voice. These computations were made manually using the Praat Open Source Software (www.praat.org). In a follow up recall test on 80 words, 40 of which belonging to the experimental materials, the average percent correct response was 82% (t (79) = 7.60; p < 0.0001).

Overall, the gesture recognition times for prime related gestures was 1973 ms, while for unrelated gestures was 2143 ms. Two ANOVAs were performed on the mean response times, one with participants and the other with materials as random factors (that will be presented together). In both the analyses the main factors were the relation between the word and the pantomime at two levels (same/different meaning) and the different referential functions at four levels (reference to a simple object, to an object usually used to accomplish an action, to a simple action and to an action to be performed with a tool).

![Fig.2: Average recognition time. SO means pantomimes representing simple objects; OA means pantomimes representing objects strictly linked to an action; SA means pantomimes representing actions to be performed directly on objects; AT means pantomimes representing actions requiring a tool to be performed. Error bars are 1 standard error.](image)

The analyses showed a significant effect of the main factor relation between word and pantomime ($F_{part}$ (1,19) = 21.052, $Mse = 2.994$, $p = 0.0002$; $F_{mat}$ (1,72) = 10.89, $Mse = 1.618$, $p = 0.0015$). On average, the recognition of the pantomime was 168 ms faster when the word had the same meaning than when it had a different meaning. Also the other main factor, referential function, was significant ($F_{part}$ (3,57) = 41.954, $Mse = 1.566$, $p < 0.0001$; $F_{mat}$ (3,72) =5.153, $Mse = 0.766$, $p = 0.003$). The referential function of simple objects was the slowest to recognize (mean = 2360 ms; sd = 726 ms), while that of objects usually used to accomplish an action was the fastest to recognize (mean = 1917 ms; sd = 540 ms). In the analyses on both participants and material Neumann-Keuls post hoc test (p level at .01) showed that the time required to recognize the referential function of simple objects (mean = 2360 ms; sd = 726 ms) was significantly slower than that of both objects usually used to accomplish an action (mean = 1917; sd = 540 ms) and simple actions (mean = 2023 ms; sd = 524 ms). In the analysis on participants there was also a difference between actions to be performed with a tool (mean = 2131 ms; sd = 698 ms), which took longer to be recognized than the 3 other referential functions.

As the participants had to name the recognized gestures, the percentage of nouns and verbs used was calculated. The referential function of simple objects was named with nouns in 78% of the responses, the referential function of objects usually used to accomplish an action was named with verbs in 67% of the responses while both the referential functions of simple actions and actions to be performed with a tool were named with verbs in 100% of the responses.

The eventual effect of the video-clips different lengths was ruled out performing an Analysis of Covariance with the presentation times of the video-clips as covariate. The eventual effect of the video-clips different lengths was ruled out performing an Analysis of Covariance with the presentation times of the video-clips as covariate (Relation factor: $F_{relation}$ (1,19) = 16.059, $Mse = 14.46$, $p = 0.0008$).

Discussion
As predicted by the Information Packaging Hypothesis, according to which the language and the gesture systems activate meaning at the cognitive level, the words sharing meaning with the pantomimes facilitated their recognition to a greater extent (1973 ms) than unrelated words did (2141 ms). The comparison with the mean baseline recognition times of pantomimes (2339 ms) clearly shows that the priming word had a positive effect on gesture recognition even when it was unrelated in meaning to the pantomime (198 ms) perhaps due to the overall activation of the meaning system. However, when both words and pantomimes shared meaning, the advantage was significantly greater (366 ms; $t$ (78) = 3.86; $p = 0.0002$). Moreover, the effect of the different types of reference (to a simple object, to an object usually used to accomplish an action, to a simple action and to an action to be performed with a tool) affected the recognition of pantomimes. In fact, while the referential function of simple objects was the slowest to recognize, that of objects usually used to accomplish an action was the fastest, with those of both simple actions and actions to be performed with a tool in between. The slower recognition time of pantomimes referring to objects can be explained with the contrasting static character of objects and the dynamic character of gestures, which are best suited to convey actions and events.
This is shown by the fastest recognition times of objects usually used to accomplish an action, which are simpler than the pantomimes of events as expressed by verbs. In fact, in naming the pantomimes of objects usually used to accomplish an action participants used verbs instead of nouns. Thus, it can be concluded that while words generally prompt the activation of the meaning system of gesture, when there is semantic priming as that produced by the same meaning, gesture recognition is significantly enhanced.

Regarding the second hypothesis, the lack of interaction between the factor relation and the factor functional role of the categories contrasts the Lexical Retrieval Hypothesis according to which gestures do not carry meaning by their own.

**Experiment 2**

In this experiment the reversed condition was studied as pantomimes were used as primes and words as targets. Accordingly, the hypotheses were the following:

- If the language and the gesture systems share meaning at the cognitive level, then pantomimes should prime the recognition of same meaning word. Moreover, if the Lexical Retrieval Hypothesis holds, the activation of words’ meaning should be faster than in the baseline condition as gestures help accessing the lexicon. Instead, if the Information Packaging Hypothesis holds, the activation of words’ meaning should not differ from the baseline.

- If it is the gesture system that shares the functional roles of linguistic items with that of language, when used without language, then no difference should be found in accessing the meanings of words differing in the type of reference. In fact, reference is the specific function of words.

**Method**

**Participants** Twenty students at the University of Trieste volunteered for their participation in the experiment receiving course credit for their participation. All of them were fluent Italian speakers (14 female; mean age = 23 years) and did not take part in any of the others studies in this research.

**Materials** The sets of gestures and words used was the same as in Experiment 1.

**Procedure** The apparatus and settings for presenting the materials and collecting the responses were the same as described in Experiment 1. Only the order of the presentation of the materials was reversed. Participants were presented with the pantomimes used as prime after a central fixation cross (500 ms). A beep-sound (200 ms), used as trigger for the analysis, signaled the presentation of the word (500 ms). The inter-trial interval was 1000 ms. The task of the participants was to pay attention to the gesture and to name the word as soon as possible. The time recorded was from the onset of the word to the beginning of naming.

**Data analysis and results**

Two ANOVAs were performed on the mean reading times, one with subjects and the other with materials as random factors, the results of which will be presented together. In both the analyses the main factors were the relation between the pantomimes and the words at two levels (same/different meaning) and the different referential functions of pantomimes at four levels (reference to a simple object, to an object usually used to accomplish an action, to a simple action and to an action to be performed with a tool). A significant effect of the main factor relation in meaning between the pantomimes and the words was found (F\text{part} (1,19) = 47.174, Mse = 0.064, p < 0.001; F\text{mat} (1,72) = 19.201, Mse = 0.032, p = 0.00004). On average, the naming time was 470 ms (sd = 26 ms) in the same meaning condition, and 509 ms (sd = 22 ms) in the different meaning condition. The other main factor, i.e., the different referential functions of pantomimes did not reach significance in any of the analyses.

![Fig.3: Average naming time. SO means nouns referring to simple objects; OA means nouns referring to objects strictly linked to an action; SA means verbs referring to actions to be performed directly on objects; AT means verbs referring to actions requiring a tool to be performed. Error bars are 1 standard error.](image)

**Discussion**

When gestures and words match in meaning, gestures do not prime naming words as words naming times (470 ms) do not differ from the baseline (459 ms; t = 1.499, p = .138), even if they are faster than in the mismatching condition. When gestures and words mismatch in meaning, there is an interference effect. In fact, words naming times takes longer than the baseline. These results, while supported by the Information Packaging Hypothesis, disprove the prediction allowed by the Lexical Retrieval Hypothesis. As expected,
the different referential functions of pantomimes do not affect word-naming time.

Conclusion
As suggested by recent research on the effects of gesture on cognitive processes, the meanings of the gesture system do not depend on the meanings of the linguistic system. Even if they jointly co-operate in shaping the complex meanings that characterize the communicative settings in everyday social interaction, nonetheless both the systems preserve their specificity with the gesture meaning system assuming linguistic-like features when replacing the linguistic one (Experiment 1). However, the meaning system of gesture is independent from the linguistic one (Experiment 2) as it is grounded in sensori-motor experience and gesture production may be an overt way of accessing embodied knowledge (Schwartz & Black, 1999). It is from this last that it draws the visuo-spatial knowledge necessary to integrate the abstract knowledge of language.

Acknowledgments
We would like to thank Isabella Poggi for insightful discussions on the relationship between language and gesture. This research was supported by the Scuola Superiore di Studi Umanistici at Alma Mater Studiorum-University of Bologna with a postdoctoral fellowship (Bando n. 1723) to the first author.

References