

# When a hit sounds like a kiss: An electrophysiological exploration of semantic processing in visual narrative



Mirella Manfredi<sup>a,b,\*</sup>, Neil Cohn<sup>c,d</sup>, Marta Kutas<sup>c</sup>

<sup>a</sup> Department of Psychology, University of Milano-Bicocca, Milan, Italy

<sup>b</sup> Social and Cognitive Neuroscience Laboratory, Center for Biological Science and Health, Mackenzie Presbyterian University, São Paulo, Brazil

<sup>c</sup> Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA

<sup>d</sup> Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, Netherlands

## ARTICLE INFO

### Article history:

Received 26 August 2016

Revised 2 February 2017

Accepted 7 February 2017

### Keywords:

Visual language

Comics

Visual narrative

Event-related potentials

N400

Late positivity

Onomatopoeic words

## ABSTRACT

Researchers have long questioned whether information presented through different sensory modalities involves distinct or shared semantic systems. We investigated uni-sensory cross-modal processing by recording event-related brain potentials to words replacing the climactic event in a visual narrative sequence (comics). We compared Onomatopoeic words, which phonetically imitate action sounds (*Pow!*), with Descriptive words, which describe an action (*Punch!*), that were (in)congruent within their sequence contexts. Across two experiments, larger N400s appeared to Anomalous Onomatopoeic or Descriptive critical panels than to their congruent counterparts, reflecting a difficulty in semantic access/retrieval. Also, Descriptive words evinced a greater late frontal positivity compared to Onomatopoeic words, suggesting that, though plausible, they may be less predictable/expected in visual narratives. Our results indicate that uni-sensory cross-modal integration of word/letter-symbol strings within visual narratives elicit ERP patterns typically observed for written sentence processing, thereby suggesting the engagement of similar domain-independent integration/interpretation mechanisms.

© 2017 Elsevier Inc. All rights reserved.

## 1. Introduction

Researchers have long questioned whether understanding the world is tied to the specific modality of input, e.g., visual or verbal information, or whether these modalities share a common semantic system. Neurophysiological research has examined this question by focusing on cross stimulus semantic processing, such as co-occurring speech and gesture (Habets, Kita, Shao, Özyürek, & Hagoort, 2011; Özyürek, Willems, Kita, & Hagoort, 2007), or on different stimuli within the same sensory (uni-sensory) modality, such as written words and pictures (e.g., Gates & Yoon, 2005; Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996). Other studies have crossed modalities by replacing a sentential word with a line drawing or picture depicting the word's referent (Ganis, Kutas, & Sereno, 1996; Nigam, Hoffman, & Simons, 1992). In much of this research, language is the dominant modality, supplemented by pictorial or gestural information, typically related to the semantic category of objects. Such work has implicated a

multi-modal (verbal, visual), distributed semantic processing system (Nigam et al., 1992; Özyürek et al., 2007), in which specific brain areas are selectively activated by particular types of information (Ganis et al., 1996; Holcomb & McPherson, 1994; Vandenberghe et al., 1996). In the present study, we reversed this visual-into-verbal embedding, inserting a written word (letter/symbol string) into a sequential image narrative (comic strip). In so doing, we could ask whether or not, and if so, how, the context of a visual narrative sequence would modulate the lexico-semantic processing of a written word.

The contextual processing of different types of information has been investigated by analyzing the N400, an electrophysiological event-related brain potential (ERP) indexing semantic analysis (Kutas & Federmeier, 2011). The N400 is typically observed in linguistic contexts, in which it is associated with access to perceptuo-semantic information about critical words in semantic priming paradigms (e.g., Bentin, McCarthy, & Wood, 1985), sentences, or discourse (e.g., Camblin, Ledoux, Boudewy, Gordon, & Swaab, 2007; Kutas & Hillyard, 1980, 1984; Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Zwitterlood, Hagoort, & Brown, 2003). The N400 component also has been observed in meaningful non-linguistic contexts, e.g., using line drawings (Ganis et al., 1996), faces (Olivares, Iglesias, & Bobes, 1999), isolated pictures (Bach,

\* Corresponding author at: Social and Cognitive Neuroscience Laboratory, Center for Health and Biological Sciences, Mackenzie Presbyterian University, Rua Piaui, 181, São Paulo 01241-001, Brazil.

E-mail address: [mirella.manfredi@gmail.com](mailto:mirella.manfredi@gmail.com) (M. Manfredi).

Gunter, Knoblich, Prinz, & Friederici, 2009; Proverbio & Riva, 2009), sequential images/video of visual events (Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008; Sitnikova, Kuperberg, & Holcomb, 2003) and visual narratives (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012; West & Holcomb, 2002). The consistent finding of an N400 both for images and words has led to the suggestion that linguistic and nonlinguistic information rely on similar semantic memory networks (Kutas & Federmeier, 2011).

Most of the N400 research has examined semantic processing within a single modality. However, researchers have begun to investigate cross stimulus semantic processing. Research on multi-sensory cross-modal processing has used stimuli presented in different sensory modalities (i.e., vision and sound). For example, speech and/or natural sounds combined with semantically inconsistent pictures or video frames have been found to elicit N400s (Plante, Petten, & Senkfor, 2000; Puce, Epling, Thompson, & Carrick, 2007; Cummings, Ceponiene, Dick, Saygin, & Townsend, 2008; Liu, Wang, Wu, & Meng, 2011). Similar results have been obtained for gestures combined with verbal information, where the (in)congruity of information across the two modalities modulates the amplitude of N400 effects (Cornejo et al., 2009; Proverbio, Calbi, Manfredi, & Zani, 2014a; Wu & Coulson, 2005, 2007a, 2007b). The congruity of gesture-music pairings also affects N400 amplitudes, at least in musicians (Proverbio, Calbi, Manfredi, & Zani, 2014b). Although present, masked priming paradigms show a much weaker and later N400-like effect in cross-modal repetition priming (verbal vs. visual) than within-modality repetition priming (Holcomb & Anderson, 1993; Holcomb, Reder, Misra, & Grainger, 2005; Sitnikova et al., 2008).

Other work has investigated unisensory cross-modal semantic processing of stimuli within the same sensory modality (i.e., vision), albeit from different systems of communication (i.e., text and images). For example, one means of investigating cross-modal but unisensory semantic processing is via substitution of an element from one modality (image) for an element in another modality (symbol string). For example, a non-linguistic visual stimulus can be inserted into a sentence (as in *I ♥ New York*); something akin to these appear in slogans, children's books, and pervasively in the use of emoticons or emoji within digital texting communications (Cohn, 2016). Electrophysiological studies have substituted a picture for a word in sentences to investigate the extent to which the two access a common semantic system (Ganis et al., 1996; Nigam et al., 1992). In particular, they were designed to determine whether N400 elicitation was specific to the linguistic system. For example, Ganis et al. (1996) reported that incongruous minus congruous sentence-final pictures and words were associated with different ERP scalp distributions: the N400 effect for words was more posterior than it was for pictures. Also, the N400 to pictures had a longer duration over frontal sites. The authors concluded that sentence-final written words and pictures are processed similarly, but not by identical brain areas.

The studies we have discussed thus far investigated the processing of objects (words and pictures) embedded in grammatical sentences. In the current study, by contrast, words referring and/or relating to events were inserted into visual narrative sequences. Recent work has demonstrated that visual narratives such as those found in comics are governed by structural constraints analogous to those found in written sentences (Cohn, Jackendoff, Holcomb, & Kuperberg, 2014; Cohn et al., 2012). For example, a “narrative grammar” organizes the semantics of event structures in sequential images much as syntax organizes meaning in sentences (Cohn, 2013b). Manipulations of this narrative grammar elicit electrophysiological responses similar to manipulations of linguistic syntax (Cohn & Kutas, 2015; Cohn et al., 2014); the N400 does not appear to be sensitive to this “grammar,” suggesting that

narrative structure is distinct from meaning in visual narratives (Cohn et al., 2012).

Visual narratives have conventional ways of inserting words into the grammar of sequential images, reflecting canonical multi-modal interactions between images and text (Cohn, 2013c, 2016; Forceville, Veale, & Feyaerts, 2010; McCloud, 1993). In particular, verbal information can replace the climactic events of a sequence depicted in a “Peak” panel (Cohn, 2013a, 2013b), typically with *onomatopoeia* (Cohn, 2016). Onomatopoeia phonetically imitate sounds or suggest the source of described sounds, and have long been recognized as a prototypical feature of comics (Bredin, 1996; Hill, 1943). As a substitution in a visual narrative, a written onomatopoeia (*Bang!*) can replace a panel depicting a gun firing, rather than being juxtaposed alongside the depicted action. This type of a substitution works on a semantic level due to the metonymic link between a gun firing and its sound, presumably via a shared semantic system (Cohn, 2016). Often, these onomatopoeic substitutions appear inside “actions stars,” a conventionalized star-shaped “flash” in comic strips, representing the culmination of an event, thereby leaving that information to be inferred (Cohn, 2013a; Cohn & Wittenberg, 2015). Because visual narratives conventionally substitute words for objects and/or events, such substitutions provide a natural way to explore cross-modal unisensory processing. In particular, we ask whether event comprehension can be accessed across different modalities. To this aim, we investigated the semantic processing of written words substituted for omitted events in visual narratives. The replacement words differed in their expectancy and were either semantically congruent or incongruent with the event they replaced (Experiment 1). In particular, we assessed whether different lexical items that occur in comics might elicit similar or different semantic processing. In addition, we asked whether this processing was modulated by the lexical forms in which the information appeared (discussed below). Given the results of Experiment 1, we conducted a more controlled comparison in Experiment 2 in which we also crossed the lexical type (i.e., the type of lexical information) and semantic congruity of written words in the visual narrative sequences.

In two experiments, we recorded ERPs to words within action stars, which replaced the primary climactic events of visual narrative sequences. In both experiments, we contrasted *onomatopoeic* words (*Pow!*) with *descriptive* words (*Impact!*) that overtly described the omitted events rather than mimicked their sound. Both descriptive and onomatopoeic “sound effects” appear in comics (Caticola & Guidi, 2015; Guynes, 2014), though corpus research indicates that onomatopoeia occur with greater frequency at least in U.S. comics (Pratha, Avunjian, & Cohn, 2016). Because both onomatopoeic and descriptive “sound effects” ostensibly index the same information, we investigated if different types of lexical information carrying the same meaning affected the cross-modal comprehension of implied events.

In Experiment 1, we contrasted the Onomatopoeic (*Pow!*) and Descriptive action star panels (*Impact!*), with Anomalous panels that used an onomatopoeic word inconsistent with the context (*Smooch!*), and “Grawlix” panels (\$#@&!?) as a baseline condition, which used symbolic strings traditionally implying swear words in comics (Walker, 1980), but ostensibly have no specific semantic representation. In Experiment 2, we focused on the processing of lexico-semantic (in)congruity versus an onomatopoeic-semantic (in)congruity. Therefore, we contrasted congruous Onomatopoeic (*Pow!*) and Descriptive panels (*Impact!*) with contextually Anomalous Onomatopoeic (*Smooch!*) and Descriptive panels (*Kiss!*).

We expected to observe modulation of N400 amplitudes in response to the different action stars across the four sequence types. Specifically, when the Onomatopoeic word was congruent

with its sequence context, we expected N400 amplitude attenuation; in this regard, we did not expect Onomatopoeic panels to differ from Descriptive panels, as both were contextually congruent. Conversely, we expected an equally large N400 both to Anomalous Onomatopoeia (Experiment 1) and Anomalous Descriptive panels (Experiment 2), suggesting difficulty retrieving verbal meaning incongruent with the ongoing visual context; this assumes that the different types of lexical information presented in a visual narrative engaged similar semantic processes (i.e., similar N400s to Anomalous Descriptive and Anomalous Onomatopoeic critical panels in Experiment 2). Finally, the Grawlix panel (Experiment 1) was neither incoherent with the context nor associated with a specific meaning. Thus, we expected a strong N400 attenuation (i.e., no N400) in Experiment 1.

## 2. Experiment 1

### 2.1. Materials and methods

#### 2.1.1. Participants

Twenty-eight right-handed undergraduate students (12 males) with self-described “comic reading experience” were recruited from the University of California at San Diego. Participants were native English speakers between 18 and 27 years of age ( $M = 20.9$ ,  $SE = 1.92$ ), had normal or corrected-to-normal vision, and reported no major neurological or general health problems or psychoactive medication. Each participant provided written informed consent prior to the experiment and received course credit for participating.

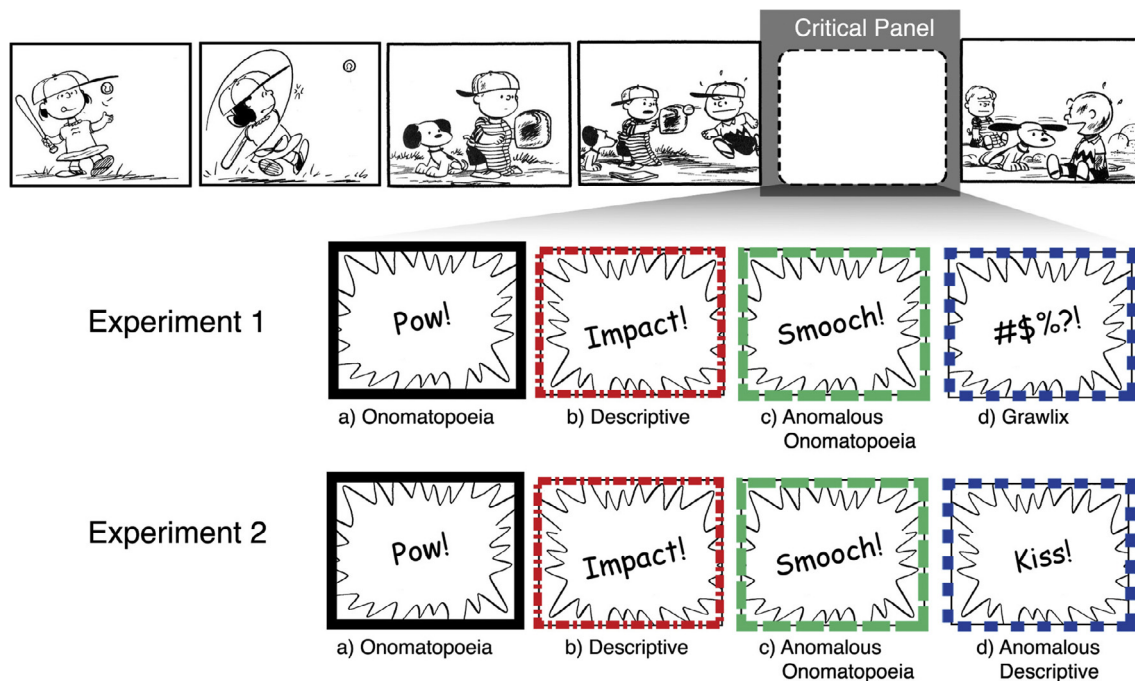
Participants’ comic reading “fluency” was assessed using the Visual Language Fluency Index (VLFI) questionnaire that asked participants to indicate the frequency with which they read various types of comics, read books for pleasure, watched movies, and drew comics, both currently and while growing up. Participants had a mean fluency of 17.82 ( $SE = 6.36$ ), a high average. An idealized average VLFI score would be 12, with low being below 7 and high being above 20 (Cohn et al., 2012). All participants knew *Peanuts*.

#### 2.1.2. Stimuli

We used black and white panels from the *Complete Peanuts* volumes 1–6 (1950–1962) by Charles Schulz (Fantagraphics Books, 2004–2006) to design 100 novel 4- to 6-panel long sequences, originally created for prior studies (Cohn & Kutas, 2015; Cohn & Wittenberg, 2015; Cohn et al., 2012). To eliminate the influence of written language in non-critical panels on comprehension, we used panels either without text or with text deleted. All panels were adjusted to a single uniform size. All sequences had a coherent narrative structure, as defined by Visual Narrative Grammar theory (Cohn, 2013b), and confirmed by behavioral ratings (Cohn et al., 2012).

We manipulated these base sequences by replacing the climactic image (Peak) of each strip with action star panels containing one of four different types of words (as in Fig. 1): *Onomatopoeic* panels (1a) contained an onomatopoeic word coherent with the context (*Pow!*), *Descriptive* panels (1b) included a word describing the hidden action (*Impact!*), *Anomalous* panels (1c) used an onomatopoeic word incoherent with the context (*Smooch!*), and *Grawlix* panels (1d) used strings of symbols typically used in comics to represent swearwords (*#\$@&!?*). The action star panels appeared in the second to the sixth panel positions, with equal numbers at each position.

The words used in all conditions (action star panels in Onomatopoeic, Descriptive, Anomalous and Grawlix) were balanced in length, varying between 3 and 13 characters. The number of the characters was multiplied for their frequency of appearance across the words in each condition (i.e., if three-character words were present nine times in the Onomatopoeic condition,  $9 \times 3 = 27$ ) in order quantitatively measure the frequency of appearance for each word length. The average values (number of characters  $\times$  their frequency) were not significantly different across the conditions, ( $F_{2,18} = 0.33$ ;  $p > 0.5$ ) (Onomatopoeic = 53.4,  $SD = 20.7$ ; Descriptive = 49,  $SD = 15.31$ ; Grawlix = 59.8,  $SD = 15.2$ ). In addition, the Onomatopoeic and Descriptive action star words were balanced in mean orthographic neighborhood density, according to the CELEX database ( $p > 0.5$ ;



**Fig. 1.** Example of visual sequences used as experimental stimuli. The climactic panel was replaced by Onomatopoeia, Descriptive, Anomalous Onomatopoeia and Grawlix action star panel in Experiment 1 (first row) and by Onomatopoeia, Descriptive, Anomalous Onomatopoeia and Anomalous Descriptive action star panel in Experiment 2 (second row).

Onomatopoeic = 4.6, SD = 4.56; Descriptive = 6.35, SD = 5.23). We also balanced the average number of word repetitions (i.e., how many times a word appears in a condition) in the Onomatopoeic (3.13, SD = 1.28) and Descriptive (3.40, SD = 1.72) action stars ( $p > 0.5$ ). Finally, the average frequency of the orthographic form for Onomatopoeic (4.13, SD = 7) and Descriptive (56.44, SD = 70.07) did differ according to the CELEX database ( $p < 0.01$ ).

Altogether, the stimulus set included 100 experimental sequences (25 per condition) and 100 filler sequences. A total of four lists (each consisting of 100 strips in random order) were created, with the four conditions counterbalanced using a Latin Square Design such that participants viewed each sequence only once in a list.

### 2.1.3. Procedure

Participants sat in front of a monitor in a sound-proof, electrically-shielded recording chamber. Before each strip, a fixation cross appeared for a duration of 1400 ms. Experimental and filler strips were presented panel-by-panel in the center of the monitor screen. Panels stayed on screen for 1350 ms, separated by an ISI of 300 ms (e.g., Cohn & Kutas, 2015). When the strip concluded, a question mark appeared on the screen and participants indicated whether the strip was easy or hard to understand by pressing one of the two hand-held buttons. Response hand was counterbalanced across participants and lists.

Participants were instructed not to blink or move during the experimental session. A post-test questionnaire asked participants to reflect on whether they were aware of any specific patterns or stimulus characteristics. The experiment had six sections separated by breaks, which altogether lasted roughly 1 h. Experimental trials were preceded by a short practice to familiarize participants with the procedures.

### 2.1.4. Electroencephalographic recording parameters

The electroencephalogram (EEG) was recorded from 26 electrodes arranged geodesically in an Electro-cap, each referenced online to an electrode over the left mastoid (Fig. 2). Blinks and eye movements were monitored from electrodes placed on the outer canthi and under each eye, also referenced to the left mastoid process. Electrode impedances were kept below 5 K $\Omega$ . The EEG was amplified with Grass amplifiers with a pass band of 0.01–100 Hz and was continuously digitized at a sampling rate of 250 samples/s.

### 2.1.5. Statistical analysis of ERPs

Trials were visually inspected for each subject. Trials contaminated by blinks, muscle tension (EMG), channel drift, and/or amplifier blocking were discarded using individualized rejection criteria. Approximately 9% of critical panel epochs were rejected due to such artifacts, with losses distributed approximately evenly across the four conditions. Each participant's EEG was time-locked to critical action star panels, and ERPs were computed for epochs extending from 500 ms before stimulus onset to 1500 ms after stimulus onset.

Two ERP component regions of interest were identified: an N400 and a Late Positivity (LP). The mean amplitude voltage of the N400 response was measured at frontal (MIPf, LMFf, RMPf) central (LMCe, MiCe, RMCe) and parieto-occipital (LDPa, MiPa, RDPa, LLOc, RLOc) electrode sites in the 300–500 ms time window. The mean amplitude of the LP was considered in the 600–800 ms time-window and measured at frontal (LMPf, MIPf, RMPf) central (LMCe, MiCe, RMCe) and parietal (LDPa, MiPa, RDPa) electrode sites.

Mean amplitude of each component was analyzed using repeated measures ANOVAs with Stimulus-type (4 levels: Onomatopoeic, Descriptive, Anomalous, Grawlix) and ROI (levels depending on the components). Multiple comparisons of means

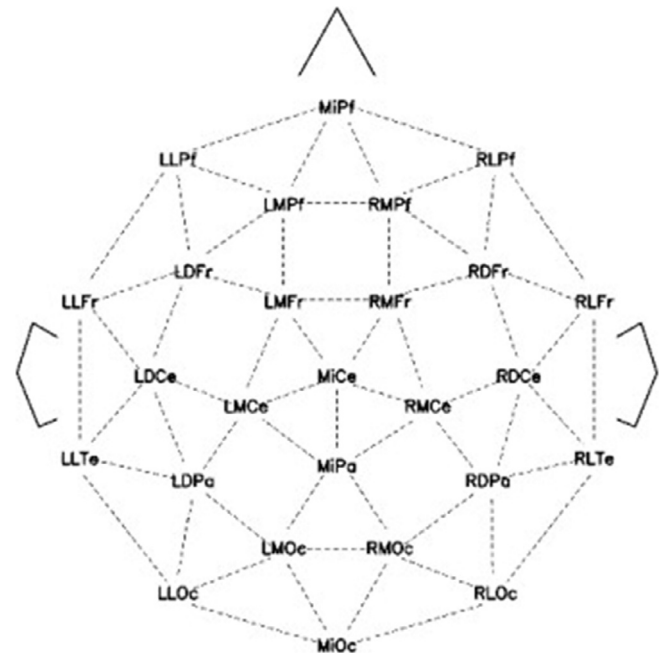


Fig. 2. Schematic showing the 26 channel array of scalp electrodes from which the EEG was recorded.

were performed by means of the post-hoc Fisher's tests. The Huynh–Feldt adjustment to the degrees of freedom was applied to correct for violations of sphericity associated with repeated measures. Thus, the corrected *degrees of freedom* and the Huynh–Feldt epsilon value are reported.

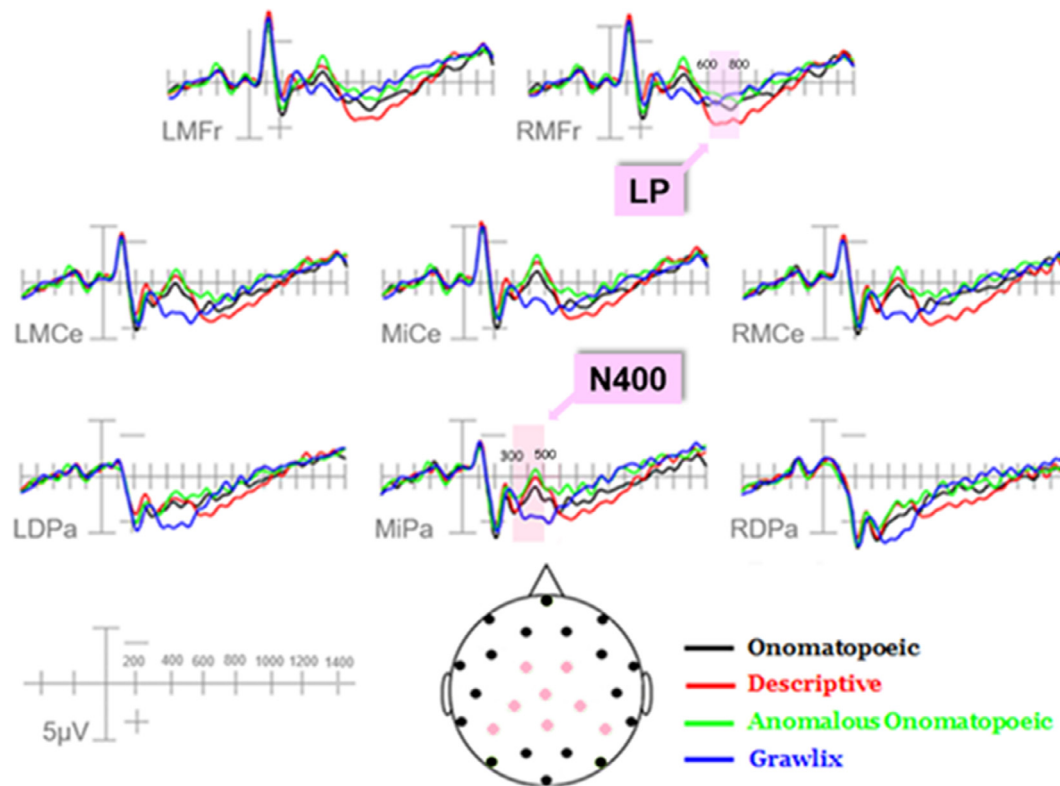
## 2.2. Results

### 2.2.1. Behavioral results

Overall, comprehensibility ratings differed between all sequence types ( $F_{3,81} = 19.96$ ,  $p < 0.01$ ). Sequences with Anomalous action stars (60%, SE = 0.02), were considered significantly less coherent ( $p < 0.01$ ) than those with Onomatopoeia panels (77%, SE = 0.02), Descriptive panels (79%, SE = 0.02) and Grawlix panels (71%, SE = 0.03). Sequences with Grawlix panels (71%, SE = 0.03) were considered less coherent than the Descriptive (79%, SE = 0.02) and Onomatopoeic sequences (77%, SE = 0.02), but significantly more coherent than Anomalous (60%, SE = 0.02). No differences were found between the Onomatopoeic and the Descriptive sequences ( $p > 0.05$ ).

### 2.2.2. Electrophysiological results

**2.2.2.1. N400 (300–500 ms).** Mean amplitude of the N400 (300–500 ms) region of the ERP showed a main effect of Stimulus-type ( $F_{2,76} = 5.85$ ,  $p < 0.05$ ,  $\epsilon = 0.94$ ), revealing a greater amplitude negativity to Anomalous than Onomatopoeic ( $p < 0.05$ ; Anomalous = 1.13  $\mu\text{V}$ , SE = 0.53; Onomatopoeic = 2.06  $\mu\text{V}$ , SE = 0.64) and Grawlix critical panels ( $p < 0.05$ ; Grawlix = 2.78  $\mu\text{V}$ , SE = 0.74). No differences were found between the Anomalous and the Descriptive critical panels ( $p = \text{n.s.}$ ; Anomalous = 1.13  $\mu\text{V}$ , SE = 0.53; Descriptive = 1.40  $\mu\text{V}$ , SE = 0.73). The N400 response was also significantly greater to Descriptive and Anomalous critical panels compared to Grawlixes ( $p < 0.05$ ; Anomalous = 1.13  $\mu\text{V}$ , SE = 0.53; Descriptive = 1.40  $\mu\text{V}$ , SE = 0.73; Grawlix = 2.78  $\mu\text{V}$ , SE = 0.74). No statistical differences appeared between the Onomatopoeic and the Descriptive critical panels or between the Onomatopoeic and the Grawlix critical panels (Fig. 3).



**Fig. 3.** Grand-average ERP waveforms recorded at frontal, central, parietal and occipital sites in response to Onomatopoeic (black), Descriptive (red), Anomalous (green) and Grawlix (blue) critical panels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The further significance of ROI ( $F_{2,65} = 24.51$ ;  $p < 0.01$ ;  $\epsilon = 0.24$ ) reflected a larger N400 at frontal than central and parieto-occipital sites and greater on the left than the right side over the parieto-occipital scalp ( $p < 0.01$ ).

Moreover, the Stimulus-type  $\times$  ROI interaction ( $F_{5,149} = 5.66$ ;  $p < 0.01$ ;  $\epsilon = 0.18$ ) revealed that Descriptive critical panels were more negative compared to Onomatopoeic ones over the medial prefrontal and occipito-lateral areas. Anomalous critical panels were significantly more negative than Onomatopoeia at all sites (except LMfr, RMfr, RDpa). The N400 response to Onomatopoeic critical panels was greater than Grawlixes over the centro-parietal sites. In addition, the N400 component was greater to Anomalous critical panel than Descriptives only at parieto-occipital sites. Finally, a greater amplitude of the N400 response appeared to Descriptive and Anomalous critical panels compared to Grawlixes at all sites (except MIPf).

**2.2.2.2. LP (600–800 ms).** A late positivity (LP) was significantly affected by Stimulus-type ( $F_{3,81} = 4.92$ ,  $p < 0.01$ ) and was greater in response to Descriptive critical panel compared to all other sequence types. A significant main effect of ROI ( $F_{8,216} = 3.59$ ,  $p < 0.01$ ) showed that the LP was significantly more positive at MiPf ( $p < 0.01$ ;  $4.55 \mu\text{V}$ ,  $\text{SE} = 1.22$ ) compared to the other electrode sites (Fig. 3).

### 2.3. Discussion

In Experiment 1, we asked whether event meaning can be accessed across different modalities. Previous studies have investigated cross-modal semantic processing by substituting a sentential word with a line/picture (Ganis et al., 1996; Nigam et al., 1992). In the present study, we used the reverse methodology to investigate whether the context of a visual narrative sequence would modu-

late the semantic analysis of a(n inserted) written word or symbol that substituted for an omitted event.

All critical panels but the Grawlixes elicited a large N400, greater to Anomalous onomatopoeia and Descriptive words and smaller to the congruent Onomatopoeic word. In addition, there was a greater fronto-central late positivity (LP) to Descriptive words, consistent with continued processing of the word presumably in relation to the visual narrative context.

These results indicate that the analysis of a visual narrative sequence can modulate the semantic analysis of individual words. Such findings are consistent with research showing that the N400 can be modulated by meaningful nonlinguistic contexts, such as sequences of visual images (Sitnikova et al., 2008, 2003) and visual narratives (Cohn et al., 2012; West & Holcomb, 2002). Our results complement similar observations of semantic processing across two modalities (Ganis et al., 1996; Nigam et al., 1992).

We take these findings to imply that the access/analysis of words that are semantically anomalous with respect to a visual narrative context is more difficult than of semantically congruous onomatopoeic.

In addition to the as expected larger N400 to semantically incongruous critical words, we also observed a larger N400 to the Descriptive than Onomatopoeic words. Unlike the Anomalous Onomatopoeia, the Descriptive words were semantically coherent within their visual narrative contexts, as reflected in the comprehensibility ratings.

Several studies have demonstrated that N400 amplitude is highly correlated with an offline measure of the eliciting word's cloze probability (Kutas & Federmeier, 2011; Jordan and Thomas, 2002; Kutas & Hillyard, 1984; Rayner and Well, 1996; Schuberth, Spoehr, & Lane, 1981; DeLong, Urbach, & Kutas, 2005). In particular, words with higher cloze probabilities in their contexts lead to a robust, graded, facilitative influence of expectancy, as reflected in reduced N400 amplitudes (Federmeier, Wlotko, De Ochoa-

Dewald, & Kutas, 2007). In our study the expectancy of the critical word types likewise might have influenced their N400 amplitudes: smallest N400 amplitudes to onomatopoeic words (with high cloze probability), largest to anomalous words (with low cloze probability) and Descriptive words, which were informative and interpretable but more unusual (unexpected and less accessible) than onomatopoeic words in the context of comics (Pratha et al., 2016).

In contrast to all other stimulus types, Grawlixes did not elicit any N400, suggesting it was probably not being processed semantically. The electrophysiological response to Grawlixes is more consistent with their processing as physical rather than semantic incongruities. After the experiment, we asked participants what meanings they associated with the grawlixes. Most participants responded that they associated grawlixes with swearwords in general, but not with any specific meanings.

Therefore, although Grawlixes are conventionally associated with swearing in general (Walker, 1980), they may not have evoked any specific conceptual or semantic content (i.e., standing in for a specific swear word) given the ongoing narrative context. The observed response is most similar to the positivity typically seen in response to illegal non-words (Martin, Kaine, & Kirby, 2006).

Beyond the N400, we observed a fronto-central late positivity that was most pronounced for the critical Descriptive panels compared to all other critical panels. This response might reflect processing of words that are contextually consistent and plausible, but with a low probability of occurrence in comics. This view is consistent with reports of frontal positivities, similar to those observed here elicited by congruent but lexically unpredicted, low probability words in sentences (Coulson & Van Petten, 2007; DeLong, Groppe, Urbach, & Kutas, 2012; DeLong, Urbach, Groppe, & Kutas, 2011; Federmeier et al., 2007; Kutas, 1993; Moreno, Federmeier, & Kutas, 2002; Van Petten & Luka, 2012; DeLong et al., 2014). This LP has been hypothesized to reflect the consequence of preactivating, but not receiving, a highly expected (i.e., high cloze probability) continuation (DeLong, 2014), or not receiving a specific lexical expectation, independent of its semantic similarity or dissimilarity to the expected word (Thornhill & Van Petten, 2012).

On this construal, the LP elicited by Descriptive words reflects a disconfirmed prediction for the lexical category of the critical word. This possibility is consistent with corpus data showing that far fewer descriptive action words than onomatopoeia appear in comics (Pratha et al., 2016).

In sum, our results confirm our hypothesis that event meaning can be accessed across different domains/modalities. In fact, we observed a modulation of the semantic processing of written words that substituted for an omitted event in a visual narrative.

While the N400 effect to an onomatopoeic incongruity resembled that for descriptive words, we could not make a direct comparison within this group of participants. We thus conducted Experiment 2 to afford a direct within subject comparison of lexico-semantic (in)congruity versus an onomatopoeic-semantic (in)congruity effects. Specifically, we contrasted congruous Onomatopoeic (*Pow!*) and Descriptive panels (*Impact!*) with contextually Anomalous Onomatopoeic (*Smooch!*) and Descriptive panels (*Kiss!*). In this way, we aimed to see what, if any, aspects of the responses in Experiment 1 were due to the onomatopoeic nature of the stimuli.

## 3. Experiment 2

### 3.1. Materials and methods

#### 3.1.1. Participants

Thirty-two right-handed (Oldfield, 1971) undergraduate students (12 males, 20 females, mean age: 20.1) were recruited from

the University of California at San Diego. Participants had a high mean VLF score of 18.39 (SD = 6.4), and inclusion criteria, recruiting practice, and subject characteristics were identical to Experiment 1. The data of four participants were excluded from the ERP statistical analyses because of EEG artifacts.

#### 3.1.2. Stimuli

The same stimulus types were used in Experiment 2 as in Experiment 1. Onomatopoeic, Descriptive, and Anomalous Onomatopoeic sequence types remained the same; additionally, we replaced Grawlix panels for action stars containing Anomalous Descriptive words, using panels from Descriptive sequences in semantically incongruous contexts. This resulted in an experimental design that crossed lexico-semantic type (Onomatopoeia, Descriptive) with semantic congruity (Coherent, Anomalous) (Fig. 1).

The words used in all conditions (action star panels in Onomatopoeic, Descriptive, Anomalous Onomatopoeic and Anomalous Descriptive) were matched in length, varying between 3 and 13 characters. As in Experiment 1, the number of the characters was multiplied by their frequency of appearance across the words in each condition in order to quantitatively assess the frequency of appearance for each word length. The values did not differ across the conditions. Moreover, the Onomatopoeic and Descriptive words were matched on orthographic neighborhood density, according to the CELEX database. We also equated the average number of word repetitions in the Onomatopoeic and Descriptive action stars. Finally, the average frequency of the orthographic form for Onomatopoeic and Descriptive did differ according to the CELEX database. As in Experiment 1, our 100 experimental sequences (25 per condition) accompanied 100 filler sequences. We created four lists (each consisting of 100 strips in random order) counterbalancing conditions using a Latin Square Design such that participants viewed each sequence only once in a list.

#### 3.1.3. Procedure, recording and statistical analysis of ERPs

The experimental procedure and recording parameters were identical to Experiment 1. The mean amplitude voltage of the N400 response was measured at frontal (LMFr, RMFr) central (LMCe, MICE, RMCE) and parietal (LDPa, MIPa, RDPa) sites in the 300–500 ms time window. The mean amplitude of the LP was analyzed in the 600–800 ms time-window and measured at frontal (LMPf, MiPf, RMPf, LMFr, RMFr) central (LMCe, MICE, RMCE) and parietal (LDPa, RDPa) electrode sites.

In Experiment 2, we analyzed mean amplitude using a repeated measures ANOVA with Stimulus-type (onomatopoeia vs. descriptive), congruence (congruous vs. incongruous) and ROI (frontal, central, posterior).

In addition, in order to observe potential differences in the time course of the semantic processing of the two types of information, we analyzed the onset latencies of the onomatopoeic and descriptive congruity effects, in each case anomalous minus congruent ERP. To this aim, we compared the latency of the N400 difference ERPs. We determined the onset of the mean latency of the N400 difference waves by measuring the area under the curve and setting the onset as the latency at which 10% of the total area was reached.

We expected that the results of this analysis provided information about differences in the speed of the semantic processing of the two types of information.

Mean latency of the difference ERPs was analyzed using repeated measures ANOVAs with Stimulus-type (2 levels: Onomatopoeic difference wave and Descriptive difference wave) and ROI (central (LMCe, MICE, RMCE), parietal (LDPa, MIPa, RDPa) and occipital (LMoc MIOc RMOc)). Multiple comparisons of means were performed via post hoc Fisher's tests.

Finally, we contrasted the onomatopoeic N400 congruity effect in Experiment 2 with that in Experiment 1. Mean amplitude and mean latency of the N400 congruity effects were analyzed using repeated measures ANOVAs with Experiment as between-subjects factor (Onomatopoeic N400 congruity effect of Experiment 1 and Onomatopoeic N400 congruity effect of experiment 2) and ROI as within-subjects factor (central (LMCe, MICe, RMCe), parietal (LDPa, MIPa, RDPa) and occipital (LMoc, MIOc, RMOc)).

Multiple comparisons of means were performed via post hoc Fisher's tests.

### 3.2. Results

#### 3.2.1. Behavioral results

Sequences with Anomalous Descriptive action stars (46%, SE = 0.03), were considered significantly less coherent ( $F(3,90) = 58.8$ ,  $p < 0.01$ ) than the Onomatopoeic (77%, SE = 0.02), Descriptive (79%, SE = 0.02) and Anomalous Onomatopoeic sequences (54%, SE = 0.03). The Anomalous Onomatopoeic sequences were rated as significantly less coherent than the Onomatopoeic and Descriptive strips ( $p < 0.01$ ). No differences were found between the Onomatopoeic and the Descriptive strips.

#### 3.2.2. Electrophysiological results

**3.2.2.1. N400 (300–500ms).** At 300–500 ms a significant main effect of Congruity ( $F(1,27) = 9.35$ ,  $p < 0.05$ ) showed a greater N400 to incongruent ( $-1.04$ ; SE = 0.53) than congruent words ( $-0.16$ ; SE = 0.49). No differences were observed between the onomatopoeic and descriptive types. The interaction between stimulus-type and congruity also was not significant ( $p > 0.05$ ; Anomalous Onomatopoeic =  $-1.18 \mu\text{V}$ , SE = 0.57; Onomatopoeic =  $-0.42 \mu\text{V}$ , SE = 0.52; Anomalous Descriptive =  $-0.89 \mu\text{V}$ , SE = 0.53; Descriptive =  $0.09 \mu\text{V}$ , SE = 0.52).

(Fig. 4). There was a significant interaction between congruence and ROI ( $F(7,189) = 3.67$ ,  $p < 0.05$ ), with greater N400 amplitudes to incongruent than congruent stimuli especially at the centro-frontal regions. In addition, we found no significant difference between the mean latencies of the Anomalous Onomatopoeic minus Onomatopoeic versus Anomalous Descriptive minus Descriptive waveforms at the centro-parietal-occipital areas ( $F(1,29) = 0.61$ ;  $p = 0.80$ ) (Fig. 5).

Finally, the ANOVA performed on the mean latency of the onomatopoeic N400 congruity effect (Anomalous Onomatopoeic – Onomatopoeic difference ERPs) in the two experiments, revealed shorter latency of the onomatopoeic N400 congruity effect in Experiment 1 (335 ms, SE = 5.78) than Experiment 2 (363 ms, SE = 5.58) ( $F(1,56) = 12.452$ ,  $p < 0.01$ ). No differences were found between the mean amplitude of the N400 congruity effect of the two experiments ( $F(1,56) = 3.5$ ,  $p = 0.06$ ) (Fig. 6).

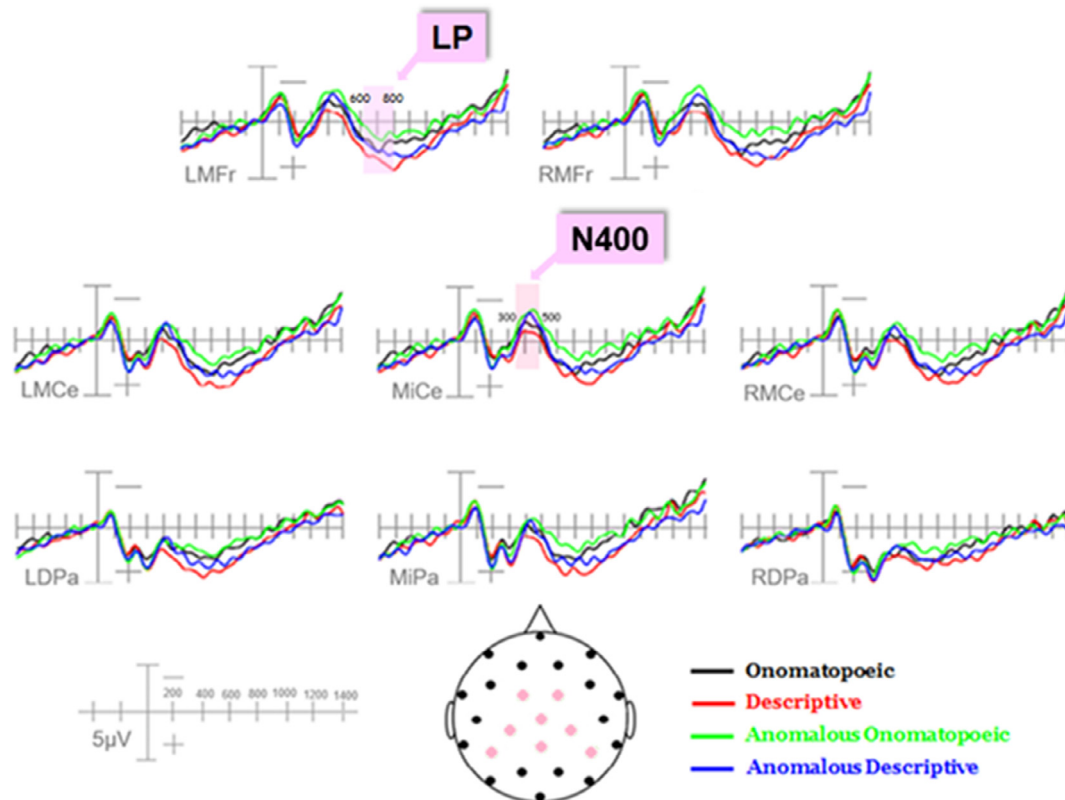
**3.2.2.2. LP (600–800 ms).** A significant main effect of Congruity ( $F(1,27) = 5.04$ ,  $p < 0.05$ ) showed a greater LP to congruent (1.99; SE = 0.43) than incongruent stimuli (1.15; SE = 0.45).

A main effect of Stimulus type ( $F(1,27) = 7.45$ ,  $p < 0.01$ ) suggested that the Late Positivity was greater to Descriptive and Anomalous Descriptive critical panels (2.07  $\mu\text{V}$ , SE = 0.44) compared to the others (Onomatopoeic and Anomalous Onomatopoeic: 1.07  $\mu\text{V}$ , SE = 0.44) (Fig. 4).

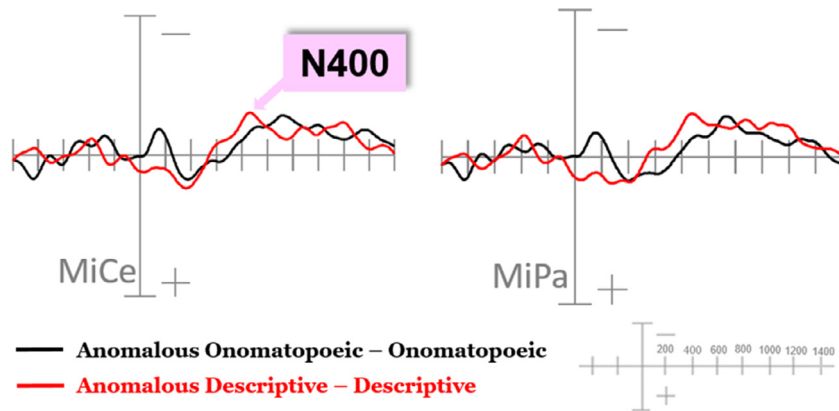
In addition, a significant main effect of ROI ( $F(10,270) = 5.07$ ,  $p < 0.01$ ) showed that the LP component was significantly more positive at MiPf compared to the other electrode sites.

### 3.3. Discussion

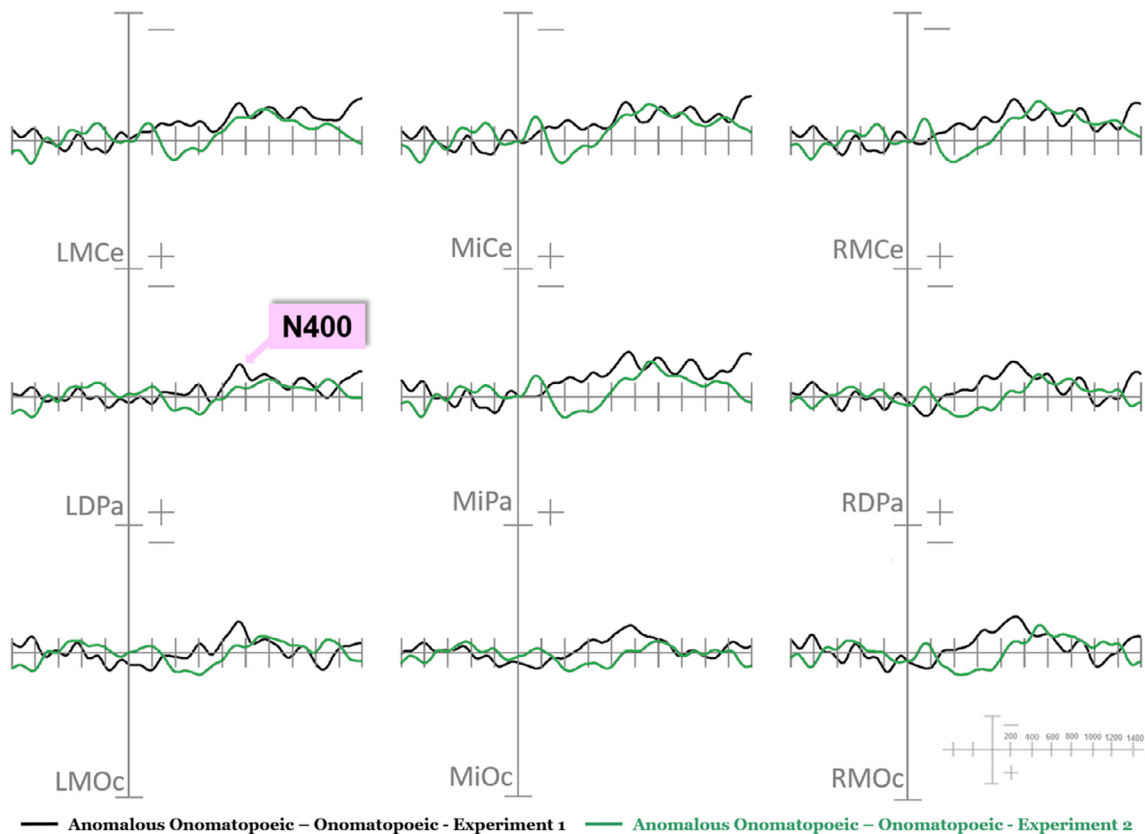
In Experiment 2, we further examined how lexical information affects the semantic processing and integration of images within a



**Fig. 4.** Grand-average ERP waveforms recorded at frontal, central and parietal sites in response to Onomatopoeic (black), Descriptive (red), Anomalous Onomatopoeic (green) and Anomalous Descriptive (blue) critical panels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** Grand-average difference ERPs recorded at central and parietal sites in response to Anomalous Onomatopoeic minus Onomatopoeic critical panel (black) and to Anomalous Descriptive minus Descriptive critical panel (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** Grand-average difference ERPs recorded at central, parietal and occipital sites to Anomalous Onomatopoeic minus Onomatopoeic critical panel of Experiment 1 (black) and to Anomalous Onomatopoeic minus Onomatopoeic critical panel of Experiment 2 (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

visual narrative. To this aim, we crossed the lexical type and semantic congruity of written words in visual narrative sequences.

We found a larger N400 to incongruous than congruous letter strings, regardless of lexical category. We also observed a greater frontal positivity to Descriptive words, regardless of congruity. Overall, these data suggest that the lexical category of words – descriptive or onomatopoeic – did not modulate the nature of its semantic analysis within a visual narrative, but may have triggered additional processing of a different nature.

As in Experiment 1, we observed modulation of N400 amplitude with congruity in response to different types of letter/symbol

strings embedded in visual narratives: incongruous strings elicited larger N400s than their congruous counterparts, whether they were descriptive words or onomatopoeic. Indeed, the congruity effects for these different types of letter strings were statistically indistinguishable, though onset is earlier for descriptive words (see Fig. 5). Taken together, the results suggest semantic processing of these two classes of stimuli (Onomatopoeic versus Descriptive) in a visual narrative is indistinguishable in the N400 time region.

In addition, we compared the mean latency of the onomatopoeic N400 congruity effect of the two experiments. The



results revealed shorter latency of the onomatopoeic N400 congruity effect in Experiment 1 compared to Experiment 2 (Fig. 6). In Experiment 1 only one type of semantic anomaly was presented (Anomalous Onomatopoeic), while in the Experiment 2 there were two types (Anomalous Onomatopoeic and Anomalous Descriptive). Thus, it is possible that the semantic incongruity in Experiment 1 was easier to recognize compared to that in Experiment 2. If so, it may explain the faster onomatopoeic N400 congruity effect in Experiment 1 compared to Experiment 2 as well as the more delayed latency to recognize anomalous onomatopoeic words in Experiment 2 compared to Experiment 1.

Also as in Experiment 1, we observed a modulation of the LP in response to different lexical categories of critical words. In particular, we observed a small fronto-central LP to both the critical Descriptive panels compared to the Onomatopoeic panels. As for Experiment 1, we hypothesized that this might be a response to a low probability lexical item (i.e., onomatopoeia vs. descriptive) within the context of a comic strip (Thornhill & Van Petten, 2012; Van Petten & Luka, 2012; De Long, 2014). Such a view aligns with corpus data showing that far fewer descriptive action words than onomatopoeia appear in American comics (Pratha et al., 2016).

In this experiment, this LP response is greater to congruent than incongruent stimuli. If replicable, it might suggest that this class of words although plausible was not predictable in the context of a visual narrative. Our results suggest that frontal late positivities may appear to low probability stimuli based on lexical type, not just specific lexical items, as in sentential contexts (Thornhill & Van Petten, 2012; Van Petten & Luka, 2012; De Long, 2014).

Nevertheless, overall, our results suggest that mechanisms similar to those involved in processing disconfirmed lexical predictions within sentences also might be engaged in the processing of words/strings within a visual narrative sequences. These results indicate that such a frontal positivity is not domain specific, as we now show such an effect associated with low lexical probability in the context of visual narratives rather than in the context of sentences. As there is no sentence context, syntactic structure cannot be relied on to make a lexical prediction. In the present case, the failed lexical prediction occurs in the context of sequential images, where all potential lexical items fulfill their roles in the narrative grammar (as a climactic “Peak”), even if some may violate the local semantics (i.e., the anomalies). The difference in probability here stems from the different rates at which certain types of lexical items appear in the context of comics at all: descriptive sound effects are less prevalent in comics than onomatopoeic ones (Pratha et al., 2016). To the extent that our observed frontal positivity engages the same mechanism as during sentence processing, our results imply that such an ERP effect occurs across domains and to different types of failed lexical predictions.

In conclusion, this study indicates that the lexical type and semantic congruity of written words, which substituted for an omitted event, affects the cross-modal processing of implied events at different stages of analysis. This suggests that lexical type and semantic congruity of words may be dissociable aspects of visual narrative comprehension.

#### 4. General discussion

In this study, we examined cross-modal semantic processing by inserting written words relating to depicted events, into action stars, within visual narrative sequences (comics). To this aim, we conducted two experiments that looked at whether event comprehension processes can be accessed across different domains and how the nature of the lexical information affects the semantic processing of information depicted in a visual narrative.

In Experiment 1, we observed that the semantic processing of words embedded in comics varied with their contextual congruence. In Experiment 2, we found qualitatively similar semantic congruity effects, namely a greater N400 to Anomalous words than semantically congruent words, regardless of lexical category.

Overall, both experiments showed similar N400 congruity effects for words and letter strings that represented descriptions/sounds of actions, from which hidden actions can be inferred. In fact, the ERPs in the two experiments did not show any reliable differences in amplitude, latency and scalp distribution of the N400 congruity effect in response to the different information embedded in the visual narrative. This is in line with reports of similar N400 congruity effects both within and across stimulus domains. In fact, previous work points to similar N400 effects to semantic violations within a wide array of meaningful stimuli, including visual and auditory words, visual narratives (West & Holcomb, 2002), short videos of events (Sitnikova et al., 2003), faces (Olivares et al., 1999), environmental sounds (Van Petten & Rheinfeldert, 1995), and actions (Bach et al., 2009; Proverbio & Riva, 2009).

Similar N400 effects have also been observed when information is presented across different sensory modalities, for example, during the combination of speech and/or natural sounds with inconsistent pictures or video frames (Plante et al., 2000; Puce et al., 2007; Cummings et al., 2008; Liu et al., 2011), gestures with inconsistent verbal information (Cornejo et al., 2009; Proverbio et al., 2014a; Wu & Coulson, 2005, 2007a, 2007b), gestures and music (Proverbio et al., 2014b). Others have investigated cross-modal semantic processing by substituting an element from one modality for an element in another modality: e.g., substitution of a picture of an object in place of a word in sentences to examine the extent to which pictures and words may access a common semantic system (Ganis et al., 1996; Nigam et al., 1992). The results suggest that similar processes were implied because there was an N400 in both cases, albeit with somewhat different scalp topography differences between words and images.

Although the N400 effects elicited by different information types seem to be functionally similar, studies have noted some differences (like scalp distribution) both within (Olivares et al., 1999; Van Petten & Rheinfeldert, 1995) and between stimulus modalities (Ganis et al., 1996; Nigam et al., 1992). These findings indicate that the meanings of different stimulus types may be processed by different brain areas, in line with the view that the semantic system is a distributed cortical network accessible from multiple modalities (e.g., Kutas & Federmeier, 2000).

In our work herein, there were no reliable statistical interactions between stimulus type and ROI. In sum, there were no reliable differences in N400 scalp distributions or timing in response to the different stimulus types that replaced an event in a visual narrative. That is, the N400 apparently was not sensitive to the different types of meaningful information embedded in the visual sequences of events. These results may implicate identical neural generators for descriptive words that replace a hidden action and for words and letter strings that represent the sound of the same action in the context of a visual narrative.

Our results raise several questions regarding crossmodal and multimodal semantic processing. In this study, we examined cross-modal processing by substituting words into the structure of a visual narrative sequence, replacing a visual event. While such a construction does appear naturally within some comics (Cohn, 2014), a similar comparison could be made when the sound effects and visual events are depicted concurrently. Similarly, we could verify whether onomatopoeic words embedded in unusual positions in comics (i.e., descriptions, captions) might elicit a LP response, similar to that observed to descriptive words appearing within an action star. This could help clarify if this a specific response to descriptive words embedded in an action star or more

general. Such work would begin to extend work on multimodal processing, as in co-speech gesture (Proverbio, et al., 2014a), applied to relationships between images and written words. In addition, we used the same sensory modality (graphics/vision), albeit examining crossmodal processing (text and images). However, it may be worth asking whether similar responses might be elicited when stimuli are presented in a *different* sensory modality, such as images and auditory words. This could indicate whether the semantic processing in visual narrative can be affected by sensory stimulus modality.

In general, these questions could clarify whether or not and if so how the semantic processing could be modulated by multiple stimulus types and features linked to meaningful visual information in a visual narrative. Specifically, we might ask which stimulus types and/or features are able to affect the semantic processing both within and across modality in a visual narrative.

In this study, we investigated the cross-modal processing of written language embedded in visual narratives (e.g., comics). As in studies using the reverse methods—embedding images into sentences—we found that the context of a visual narrative sequence modulated the semantic processing of words. The present investigation thus provides additional evidence that brain responses typically observed within written sentences also can arise in unisensory contexts that cross modalities.

## Acknowledgements

This research was funded by NIH grant #5R01HD022614, and support for Dr. Cohn was provided by an NIH funded (T32) post-doctoral training grant the Institute for Neural Computation at UC San Diego. Fantagraphics Books generously donated *The Complete Peanuts*. Mirella Manfredi was supported by a doctoral researcher grant of the University of Milano-Bicocca.

## References

- Bach, P., Gunter, T. C., Knoblich, G., Prinz, W., & Friederici, A. D. (2009). N400-like negativities in action perception reflect the activation of two components of an action representation. *Social Neuroscience*, 4, 212–232.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography & Clinical Neurophysiology*, 60, 353–355.
- Bredin, H. (1996). Onomatopoeia as a figure and a linguistic principle. *New Literary History*, 27(3), 555–569.
- Camblin, C. C., Ledoux, K., Boudewy, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research*, 1146, 172–184.
- Catricalà, M., & Guidi, A. (2015). Onomatopoeias: A new perspective around space, image schemas and phoneme clusters. *Cognitive Processing*, 16(1), 175–178.
- Cohn, N. (2013a). *The visual language of comics: Introduction to the structure and cognition of sequential images*. London, UK: Bloomsbury.
- Cohn, N. (2013b). Visual narrative structure. *Cognitive Science*, 37(3), 413–452.
- Cohn, N. (2013c). Beyond speech balloons and thought bubbles: The integration of text and image. *Semiotica*, 197, 35–63.
- Cohn, N. (2014). You're a good structure, Charlie Brown: The distribution of narrative categories in comic strips. *Cognitive Science*, 38(7), 1317–1359.
- Cohn, N. (2016). A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition*, 146, 304–323.
- Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2014). The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia*, 64, 63–70.
- Cohn, N., & Kutas, M. (2015). Getting a cue before getting a clue: Event-related potentials to inference in visual narrative comprehension. *Neuropsychologia*, 77, 267–278.
- Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*, 65(1), 1–38.
- Cohn, N., & Wittenberg, E. (2015). Action starring narratives and events: Structure and inference in visual narrative comprehension. *Journal of Cognitive Psychology*, 27(7), 812–828.
- Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., & Núñez, R. E. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition*, 70(1), 42–52.
- Coulson, S., & Van Petten, C. (2007). A special role for the right hemisphere in metaphor comprehension? ERP evidence from hemifield presentation. *Brain Research*, 1146, 128–145.
- Cummings, A., Ceponiene, R., Dick, F., Saygin, A. P., & Townsend, J. (2008). A developmental ERP study of verbal and non-verbal semantic processing. *Brain Research*, 1208, 137–149.
- DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, 121(3), 226–239.
- DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, 48, 1203–1207.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121.
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75–84.
- Forceville, C., Veale, T., & Feyaerts, K. (2010). Balloons: The visuals of balloons in comics. In J. Goggin & D. Hassler-Forest (Eds.), *The rise and reason of comics and graphic literature: Critical essays on the form* (pp. 56–73). Jefferson: McFarland & Company Inc..
- Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8(2), 89–106.
- Gates, L., & Yoon, M. G. (2005). Distinct and shared cortical regions of the human brain activated by pictorial depictions versus verbal descriptions: An fMRI study. *NeuroImage*, 24(2), 473–486.
- Guynes, S. A. (2014). Four-color sound: A Peircean semiotics of comic book onomatopoeia. *The Public Journal of Semiotics*, 6(1), 58–72.
- Habets, B., Kita, S., Shao, Z., Ozyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience*, 23(8), 1845–1854.
- Hill, G. E. (1943). The vocabulary of comic strips. *Journal of Educational Psychology*, 34(2), 77.
- Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, 8, 379–412.
- Holcomb, P. J., & McPherson, B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, 24(2), 259–276.
- Holcomb, P. J., Reder, L., Misra, M., & Grainger, J. (2005). The effects of prime visibility on ERP measures of masked priming. *Cognitive Brain Research*, 24(1), 155–172.
- Jordan, T. R., & Thomas, S. M. (2002). In search of perceptual influences of sentence context on word recognition. *Journal of Experimental Psychology Learning Memory & Cognition*, 28, 34–45.
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single word versus sentence context effects. *Language and Cognitive Processes*, 8(4), 533–572.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463–470.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–208.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Liu, B., Wang, Z., Wu, G., & Meng, X. (2011). Cognitive integration of asynchronous natural or non-natural auditory and visual information in videos of real-world events: An event-related potential study. *Neuroscience*, 180, 181–190.
- Martin, F. H., Kaine, A., & Kirby, M. (2006). Event-related brain potentials elicited during word recognition by adult good and poor phonological decoders. *Brain and Language*, 96, 1–13.
- McCloud, S. (1993). *Understanding comics: The invisible art*. New York, NY: Harper Collins.
- Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, 80, 188–207.
- Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, 4(1), 15–22.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Olivares, E. I., Iglesias, J., & Bobes, M. A. (1999). Searching for face-specific long latency ERPs: A topographic study of effects associated with mismatching features. *Cognitive Brain Research*, 7, 343–356.
- Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, 19, 605–616.
- Plante, E., Petten, C. V., & Senfor, J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia*, 38(13), 1669–1684.
- Pratha, N. K., Avunjian, N., & Cohn, N. (2016). Pow, punch, pika, and chu: The structure of sound effects in genres of American comics and Japanese manga. *Multimodal Communication*, 5(2), 93–109.

- Proverbio, A. M., Calbi, M., Manfredi, M., & Zani, A. (2014a). Comprehending body language and mimics: An ERP and neuroimaging study on Italian actors and viewers. *PLoS ONE*, 9(3), e91294.
- Proverbio, A. M., Calbi, M., Manfredi, M., & Zani, A. (2014b). Audio-visuomotor processing in the Musician's brain: An ERP study on professional violinists and clarinetists. *Scientific Reports*, 4, 5866.
- Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, 459, 142–146.
- Puce, A., Epling, J. A., Thompson, J. C., & Carrick, O. K. (2007). Neural responses elicited to face motion and vocalization pairings. *Neuropsychologia*, 45(1), 93–106.
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review*, 3, 504–509.
- Schubert, R. E., Spoehr, K. T., & Lane, M. D. (1981). Effects of stimulus and contextual information on the lexical decision process. *Memory & Cognition*, 9, 68–77.
- Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, 20, 1–21.
- Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, 40(1), 160–164.
- Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83(3), 382–392.
- Van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671.
- Van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17(3), 701–718.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 83(2), 176–190.
- Van Petten, C., & Rheinfeldert, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, 33(4), 485–508.
- Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, 383, 254–256.
- Walker, M. (1980). *The lexicon of comicana*. Port Chester, NY: Comicana Inc..
- West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, 13, 363–375.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6), 654–667.
- Wu, Y. C., & Coulson, S. (2007a). How iconic gestures enhance communication: An ERP study. *Brain & Language*, 101, 234–245.
- Wu, Y. C., & Coulson, S. (2007b). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*, 14, 57–63.