Brain & Language 169 (2017) 28–38

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

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

Mirella Manfredi [a,b,*], Neil Cohn [c,d], Marta Kutas [c]

[a] Department of Psychology, University of Milano-Bicocca, Milan, Italy
[b] Social and Cognitive Neuroscience Laboratory, Center for Biological Science and Health, Mackenzie Presbyterian University, São Paulo, Brazil
[c] Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA
[d] 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
Available online 24 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 Onomatopoetic words, suggesting that, though plausible, they may be less predictable/expected in visual narratives. Our results indicate that uni-sensory cross-model 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.

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, Ozyurek, & 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, Zwitserlood, 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 (M. Manfredi).

http://dx.doi.org/10.1016/j.bandl.2017.02.001
0093-934X/

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 multisensory 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 multimodal 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 *onomatopoetic* words (*Pow!*) with *descriptive* words (*Impact!*) that overtly described the omitted events rather than mimicked their sound. Both descriptive and onomatopoetic "sound effects" appear in comics (Catricalà & 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 onomatopoetic 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 onomatopoetic-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