Narrative conjunction's junction function: The interface of narrative grammar and semantics in sequential images

Neil Cohn *

Center for Research in Language, University of California, San Diego, 9500 Gilman Dr. Dept. 0526, La Jolla, CA 92093-0526, United States

Abstract

While simple visual narratives may depict characters engaged in events across sequential images, additional complexity appears when modulating the framing of that information within an image or film shot. For example, when two images each show a character at the same narrative state, a viewer infers that they belong to a broader spatial environment. This paper argues that these framings involve a type of "conjunction," whereby a constituent conjoins images sharing a common narrative role in a sequence. Situated within the parallel architecture of Visual Narrative Grammar, which posits a division between narrative structure and semantics, this narrative conjunction schema interfaces with semantics in a variety of ways. Conjunction can thus map to the inference of a spatial environment or an individual character, the repetition or parts of actions, or disparate elements of semantic associative networks. Altogether, this approach provides a theoretical architecture that allows for numerous levels of abstraction and complexity across several phenomena in visual narratives.
© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Sequential images have been a basic system of human expression dating back at least as far as cave paintings, and in contemporary society appear in comics and films (McCloud, 1993). While simple visual narratives may depict characters engaged in events across sequential images, additional complexity appears when modulating the framing of that information within an image or film shot. Consider Fig. 1a, where the first panel shows a boxer reaching back in preparation, while the second panel shows him striking his opponent. In Fig. 1b, the first panel shows only the puncher, while the second panel shows only the opponent, before coming together in the same final panel.

These sequences differ in that 1a uses a single panel to show the same information as appears in two panels in 1b. Both sequences convey similar referential entities (boxers) and their events (punching)—but differ in how the panels selectively create a "window" on the characters. Because of this, these two panels in 1b must "add up" to the single panel in 1a. This implies a *hierarchic* relationship between both panels 1 and 2 with that of panel 3, since this relation is equivalent to the single image in Fig. 1a. This relationship is here posited as a type of **conjunction**, whereby the first two panels share a common role in the sequence, i.e., analogous to the syntactic sense of "conjunction" (e.g., Culicover and Jackendoff, 2005) not the semantic/discourse sense (e.g., Martin, 1983). In addition, in Fig. 1b both characters are

* Tel.: +1 858 822 0736; fax: +1 858 822 5097.
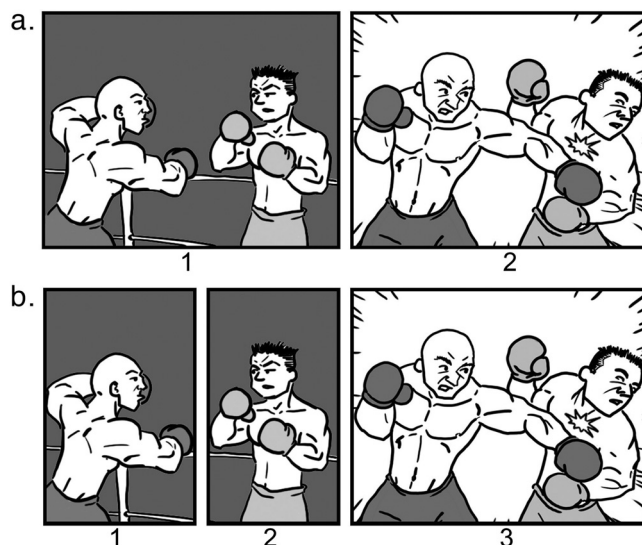*E-mail address:* neilcohn@visuallanguagelab.com.

Fig. 1. Variation in how panels provide a ''window'' on a visual narrative scene.

inferred as belonging to the same spatial environment, thereby creating a ''virtual'' single panel like that in 1a. Thus, while the same basic information occurs across both examples (one boxer reaching back to punch another), the sequence with conjunction appears to warrant additional semantic inference in order to be understood.

This work proposes a theoretical architecture for explaining these relationships in visual narratives in order to make predictions and testable hypotheses about how they operate in comprehension. This analysis focuses on the *representational* level—the basic patterns and structures that underlie understanding—rather than the level of *processing*—how those representations are operated upon in comprehension (Jackendoff, 2002, 2007; Marr, 1982). While the broader research program aims toward describing online processing, establishing theoretical representations can provide a basis for empirical research in the same way that constructs from linguistics have long informed experimentation in psycholinguistics. Such work is already underway for the broader paradigm in which this work is embedded (Cohn, 2014b; Cohn et al., 2014, 2012a; Cohn and Paczynski, 2013).

2. Visual narrative grammar

The current discussion will expand on the theory of **Visual Narrative Grammar** (VNG), which has argued that sequential images are organized using a narrative structure analogous to the way that words are organized by a syntactic structure in sentences (Cohn, 2013b). However, because images typically convey information above the level of a single word, this structure organizes semantic information closer to a discourse level of semantics. Thus, the analogy between narrative and syntax operates with regard to the *abstract* structural and functional principles of their architectures, not their surface features: Both syntax and narrative function to organize semantic information using units (panels, words) that take on categorical roles embedded into larger constituents (phases, phrases), which thereby enables hierarchic embedding, distance dependencies, and the resolution of structural ambiguities, among others. This hierarchic quality of constituent structure enables VNG to directly address the groupings posed by Fig. 1. However, we must first define the basic principles of this theory.

VNG draws on Jackendoff's (2002) model of a Parallel Architecture which argues that language involves an equal interaction between *phonology*, *conceptual structure*, *spatial structure*, and *syntax*. Because these components exist in parallel, none are privileged, and each structure operates with its own constraints while connecting to each other through ''interface rules.'' The whole of their interactions results in linguistic utterances. Such a separation of structures is commensurate with the psycholinguistic literature showing differences in processing between syntax and semantics (e.g., Marslen-Wilson and Tyler, 1980; Osterhout and Nicol, 1999; Van Petten and Kutas, 1991). In turn, VNG also argues that visual narratives involve the interaction of several components, again keeping structure (narrative) distinct from meaning (semantics). Such an architecture is supported by empirical work showing a separation between, and different neural responses evoked by, narrative structure and semantics in the processing of visual sequences (Cohn et al., 2014, 2012a).

This overall orientation thus differs from several models of narrative and sequential images where structure and meaning are either conflated or left ambiguous. For example, this contrasts with previous ''grammatical'' approaches such