Narrative conjunction’s junction function: A theoretical model of “additive” inference in visual narratives

Neil Cohn (neilcohn@visuallanguagelab.com)
Center for Research in Language, University of California, San Diego
9500 Gilman Dr. Dept. 0526, San Diego, La Jolla, CA 92093-0526

Abstract

Visual narratives often depict images of individual characters without showing the larger scene, meaning that this whole spatial environment must be inferred from these component parts. However, few theoretical models of narrative or discourse have attempted to explain the generation of such “additive” inference. This paper explores the complex interactions between narrative structure and meaning within these types of discourse phenomena, situated within the model of Visual Narrative Grammar based on Jackendoff’s (2002) Parallel Architecture of linguistic structure. Narrative “Conjunction” repeats a single narrative category within a broader constituent, allowing for expansion of a sequence beyond the canonical narrative arc. These conjoined units then correspond to semantic structures in a variety of ways, allowing an “additive” inference of actions, scenes, characters, and/or semantic associative networks. This simple yet powerful architecture enables us to account for a large variety of phenomena in visual narratives and other discourse contexts, while providing a structure that can be tested in empirical research.

Keywords: Narrative; discourse; inference; Kuleshov effect; situation model; comics; film; visual language.

Introduction

Theories of film editing have long noticed that disparate images of a scene or individual are understood as “adding up” a larger conceptualization. Experiments conducted by filmmaker Lev Kuleshov in the 1920s (Kuleshov, 1974) combined film shots of moving body parts of different women (hands, feet, eyes, heads), yet film viewers interpreted these shots as a single woman in motion. In another experiment, people responded to different shots of a scene as if they formed a coherent spatial environment when each of the shots was actually filmed in different locations. These experiments nicely showed that people inferentially construct a coherent representation, in a way that is different from the linear bridging inferences commonly discussed in research on discourse and visual narrative (Magliano & Zacks, 2011; McCloud, 1993).

This issue becomes even more apparent in static visual narratives, such as those in comics. Consider Figure 1. These sequences only differ in that information from the single panel in 1a is dispersed into two panels in 1b. There is no difference in meaning—just how the panels selectively create a “window” on the different characters. Because of this fact, it stands to reason that the two panels in 1b “add up to” the single panel in 1a. In other words, we must infer that these characters belong to a single spatial environment, thereby creating a “virtual” single panel like that in 1a. Furthermore, in 1b knowledge of *both characters* within this inferred environment must connect to the final panel.

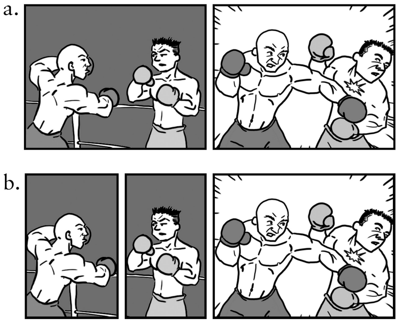

Figure 1: Panels providing a “window” on a visual scene.

All of this belies a linear focus on inference in the understanding of this sequence. First, the inference is constructed out of the combination of two panels’ content, not simply from what is left out “between” them. Second, the *additive* value of those panels must progress to the third panel, not simply the linear juxtaposition between the penultimate and final panels.

In contrast to linear approaches, Visual Narrative Grammar (VNG) assigns narrative categories to discourse units like panels, and then organizes them into larger constituents, analogous to the way that words are organized into larger phrase structures in syntax (Cohn, 2013b). This constituent structure can directly address the issue of “additive inference” at a single narrative state.

VNG is based on Jackendoff’s (2002) Parallel Architecture for language, meaning narrative structure and semantics are kept separate, yet interface in specific and predictable ways. This differs from previous “grammatical” approaches to narrative structure such as story grammars from the late 1970s (e.g., Mandler & Johnson, 1977), and grammars of film (e.g., Carroll, 1980), where the relationship between structure and meaning remain ambiguous. In VNG, narrative categories are independent from semantics, yet they interface in prototypical ways. This is analogous to how grammatical categories like nouns and verbs might have prototypical correspondences to objects and events, though these semantic qualities do not solely define them.

The canonical narrative arc in VNG starts with an Establisher, which passively introduces an interaction or situation. An Initial then begins the events of the interaction (as in the boxer’s preparation in Figure 1a), which climaxes in a Peak (the boxer's punch). Finally a Release dissipates the tension of the Peak, depicting a coda or response to an action, as in an ending where the boxer might be knocked out in the final panel.

Following Jackendoff's Parallel Architecutre, VNG uses four main components, as depicted for Figure 1a in Figure 2. ***Graphic structure*** is the physical structure of lines and shapes, which map these physical features to basic meanings (Cohn, 2013a). Conceptual information is broken up into two parts: The ***event structure*** stores the meaning of the situations and events that take place in and between images. Discrete events typically have a preparation, head, and coda, while continuous processes end in a termination (Jackendoff, 2007). ***Spatial structure*** conveys then a geometric type of meaning (Jackendoff, 2002) such as how characters relate to a larger environmental space. Finally, the ***narrative structure*** organizes meaning into a coherent sequence. Here, the narrative structure is fairly simple: the preparatory action in the first panel maps to an Initial, while the completed action maps to a Peak. These are both canonical mappings of semantics to narrative. Together, graphic structure and narrative structure provide the way that meaning is presented (its "textbase") while the additive sum of spatial and event structures constitute the meaning itself (the "situation model") (e.g., van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998).

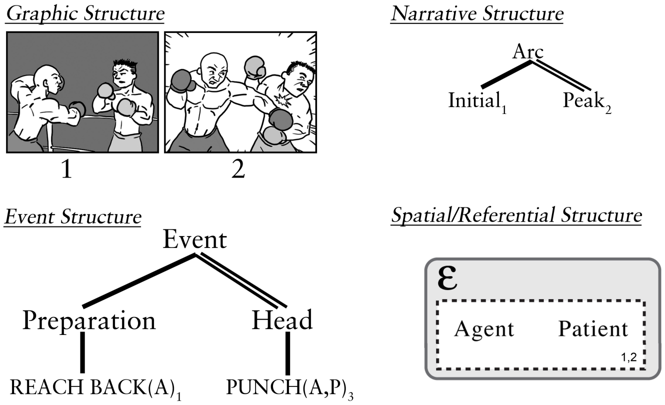

Figure 2: Parallel architecture for Figure 1a.

Environmental-Conjunction

Now let's consider what happens when the first panel is split apart, as in Figure 1b, now represented in Figure 3. In this case, the basic event structures remain the same: There is still a reaching back and punching of agent to patient. The overall spatial structure also still involves the same two characters. What has changed is how that spatial structure is divided by the graphic structure. Now, panels 1 and 2 each show a single character.

This alteration in spatial structure changes the narrative structure. According to VNG, the overarching narrative category remains the same—it is still an Initial—only it is divided into a node that contains subordinate Initials, forming a ***Conjunction*** node. Here, the higher-level Initial still maps to the overall environment, just like the single first panel in Figure 2. However, this larger environment is now inferred (notated with epsilon), and the individual panels map to parts of that spatial structure to highlight individual characters. This inference is not based on linear bridging inferences between panels, but rather the two panels together unite to infer a "virtual" environment.

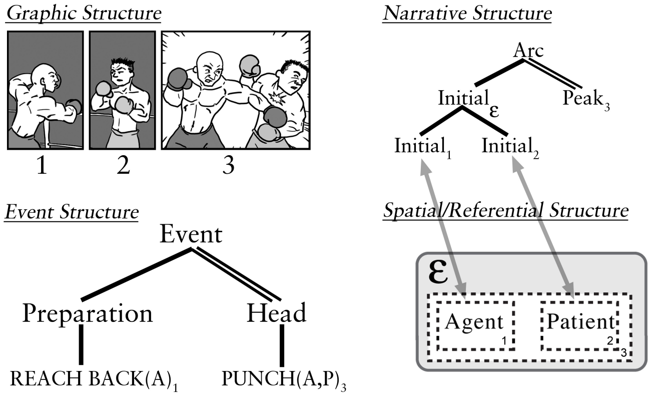

Figure 3: Parallel architecture for Figure 1b.

This basic narrative schema is a "Conjunction" phase (Cohn, 2013b), whereby a single node can contain any number of daughters of the same category:

> *Conjunction Rule*
> Given a surface structure where panels do not clearly belong to separate progressing narrative states, assign them both the same narrative category and conjoin them into a superordinate constituent reflecting their shared narrative role: *Constituent X* $\rightarrow_{Conj}(X_1, X_2, \ldots X_n)$

Figure 3 depicts ***Environmental-Conjunction*** *(E-Conjunction),* a particular type of conjunction where two characters combine to form a larger spatial environment. E-Conjunction here operates to unify the two panels in (1b) into a virtual structure equated to the single environment of (1a). The whole environment is not provided, and must be constructed in the mind in absence of being drawn.

In this approach, E-Conjunction reflects the narrative-semantics interface—the linking of narrative ordering to semantic information—rather than a purely grammatical operation (like a phrase structure rule). This might be stated as a "correspondence rule" between Narrative Structure and Conceptual Structure (NS-CS Rule):

NS-CS Rule 1: E-Conjunction

1. Given a surface structure that uses the Conjunction Rule where each panel features *different entities of a broader environment* (1,2,…n), map each narrative category to their corresponding entity in referential structure.
2. Then interface the whole constituent to a *broader semantic environment* consisting of the entities depicted in the panels of that constituent (ε).

This rule allows for us to map the entities in conjoined narrative categories into a broader environment in conceptual and spatial structure. This correspondence is diagrammed in Figure 3. Here, boxes with dotted lines correspond to actual panels (identified by numbers), while the Greek subscripts correspond to E-Conjunction mappings between the structures, i.e., the "Mental Environment" ("ε") for the scene. That is, they designate the spatial structure built by the concatenation of multiple entities.

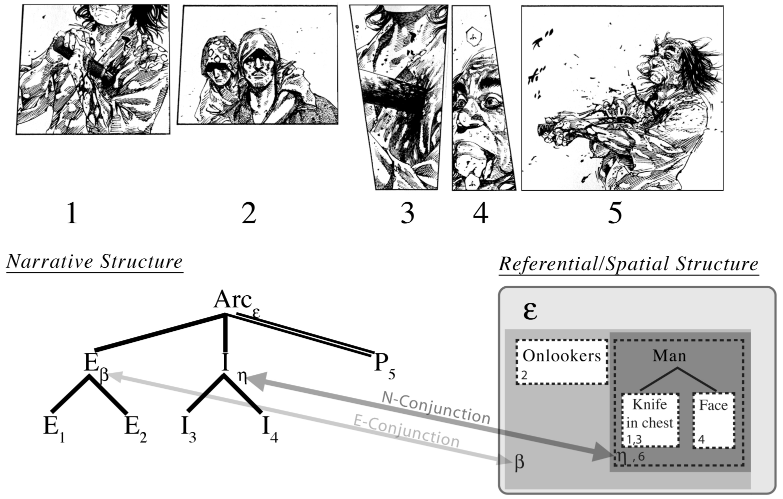

Figure 4. Entity-Conjunction in the manga *Vagabond* (Takehiko 2004)

Entity-Conjunction

E-Conjunction is not the only type of interface between conjoined panels and meaning. Consider the somewhat gruesome example in Figure 4 from Inoue Takehiko's *Vagabond*. Here, a single character pulls a knife out of his chest while his friends (panel 2) look on. After setting up the characters in the first two panels, panels 3 and 4 show Initials that start the action of the man pulling the knife out of his chest. Though they are conjoined, these panels do not use E-Conjunction, because they show a part-whole relationship to construct the notion of a *single entity* rather than a whole scene.

Entity-Conjunction *(N-Conjunction)* thus uses panels showing parts of a character to build a singular entity. We can stipulate a correspondence rule to reflect this difference:

NS-CS Rule 2: N-Conjunction

1. Given a surface structure that uses the Conjunction Rule where each panel features *different aspects of a single entity*, map each narrative category to their corresponding parts of an entity in semantic structure.
2. Then interface the whole constituent to a broader *semantic entity* consisting of the parts contained in the panels of that constituent (η).

The narrative Conjunction Rule changes only in how it maps narrative to conceptual structure: the interface to semantics connects to entities instead of environments. N-Conjunction is the type of Kuleshov effect described at the outset where viewers saw disparate body parts and understood them to add up to a single woman.

In Figure 4, the full "Man" entity is both constructed out of the two Initials, and also given in full in the Peak—hence a dotted line around a grey box in the referential/spatial structure.

Action-Conjunction

Beyond the construction of referential information, another type of mapping to Conjunction involves events, as in Figure 5. The repetition in this sequence sustains an Initial of conjuring fire across several panels before the Peak, where the light extinguishes. The repetition here does not show parts of an environment or parts of a single character. Rather, this ***Action-Conjunction*** *(A-Conjunction)*, repeats a narrative category to show the iterations of a single action. This interface connects to event structures describing the actions as opposed to referential or spatial structures describing the entities involved in the action.

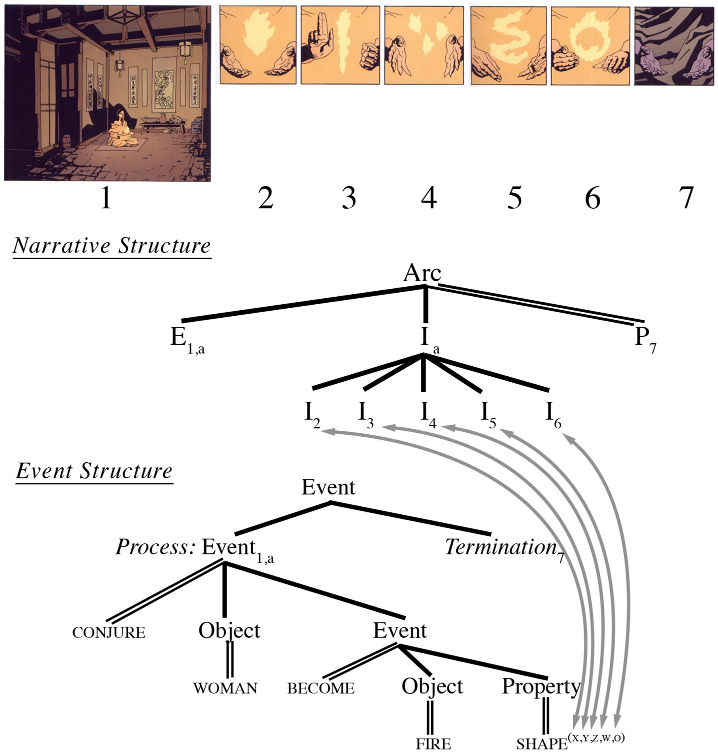

Figure 5. Action-Conjunction in the comic *B.P.R.D.* (Mignola, Sook, et al. 2003)

We can state this version of Conjunction in rule form as:

NS-CS Rule 3: A-Conjunction

1. Given a surface structure that uses the Conjunction Rule where each panel features *different aspects of a single event or action*, map each narrative category to their corresponding parts of an event in semantic structure.
2. Then interface the whole constituent to a broader *event structure* consisting of the parts contained in the panels of that constituent.

Again, this correspondence rule uses the general Conjunction Rule, but interfaces the panels to various aspects of events. In Figure 5, the event of 'conjuring fire' maps each shape onto a different Initial panel, though the full event maps to the whole upstairs Initial. That is, the whole constituent is about conjuring fire, which manifests iteratively in different representations.

Semantic networks

Another type of Conjunction draws together related or unrelated panels to form a larger meaning. Saraceni (2000, 2001) noted that panels may share broader aspects of a semantic network, without conveying an explicit narrative. Consider Figure 6, which shows Schroeder in a "training montage" in the Initials where he prepares like an athlete to play the piano in the Peak. With no coherent narrative progression, these panels are bound only through a semantic field expressing the concept of "exercise/training." These conjoined panels have no discernable connections to a scene, individual, or actions, but rather provide disconnected glimpses of a broader idea—a semantic network—which otherwise has no inherent spatial or causal connections:

NS-CS Rule 4: S-Conjunction

1. Given a surface structure that uses the Conjunction Rule where each panel features *different aspects of a semantic network or seemingly unconnected panels*, map each panel to various parts of a semantic network.
2. Then, when possible, interface the whole constituent to a *superordinate conceptual structure* consisting of the parts contained in the panels of that constituent.

This rule captures panel connections that may not have a specified structure, though may be connected through semantic associations alone. In Figure 6, all of the Initial panels could be rearranged within this phase without impacting the felicity of the sequence (they could also be deleted without much effect). This is because these relative concepts are also unordered in conceptual structure—indicated by the curly brackets around the items in the semantic network. Because of this, each panel does not necessarily play an explicit role, but together they convey the broader meaning of the sequence.

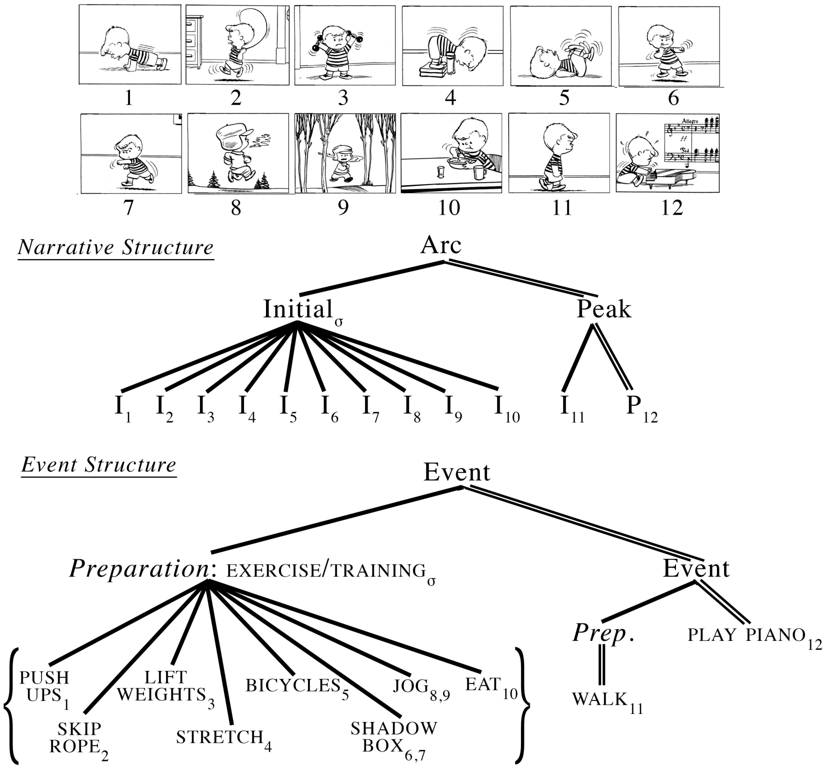

Figure 6. S-Conjunction showing conjoined panels sharing a common semantic network (Schulz 2004 [1953])

Conjunction and attention structure

NS-CS interfaces also connect to the basic representations of information in individual panels. Panels act as "attention units" that frame information in ways that can be categorized (Cohn, 2011; Cohn, Taylor-Weiner, & Grossman, 2012). ***Macros*** depict full scenes, ***Monos*** depict individual entities, and ***Micros*** depict less than a single entity. ***Polymorphic representation*** can also alter those framings, by repeating figures doing an action within a single panel (such as repeating arms to show movement, rather than having multiple limbs).

Each of these categories are ways that individual panels frame information, and essentially, the different types Conjunction are constructing "virtual" versions of these categories. For example E-Conjunction depicts the component parts of a scene (often Monos) while the full scene is constructed in spatial structure alone—a "virtual" Macro. Similarly, N-Conjunction uses various panels depicting less than a single character (usually Micros), and constructs that character—a virtual Mono. A-Conjunction unites iterations or repetitions of a single event or action—just as all that information can be conveyed in a single panel using polymorphic morphology. Finally, S-Conjunction depicts disparate information bound through only a common semantic network or superordinate category. This

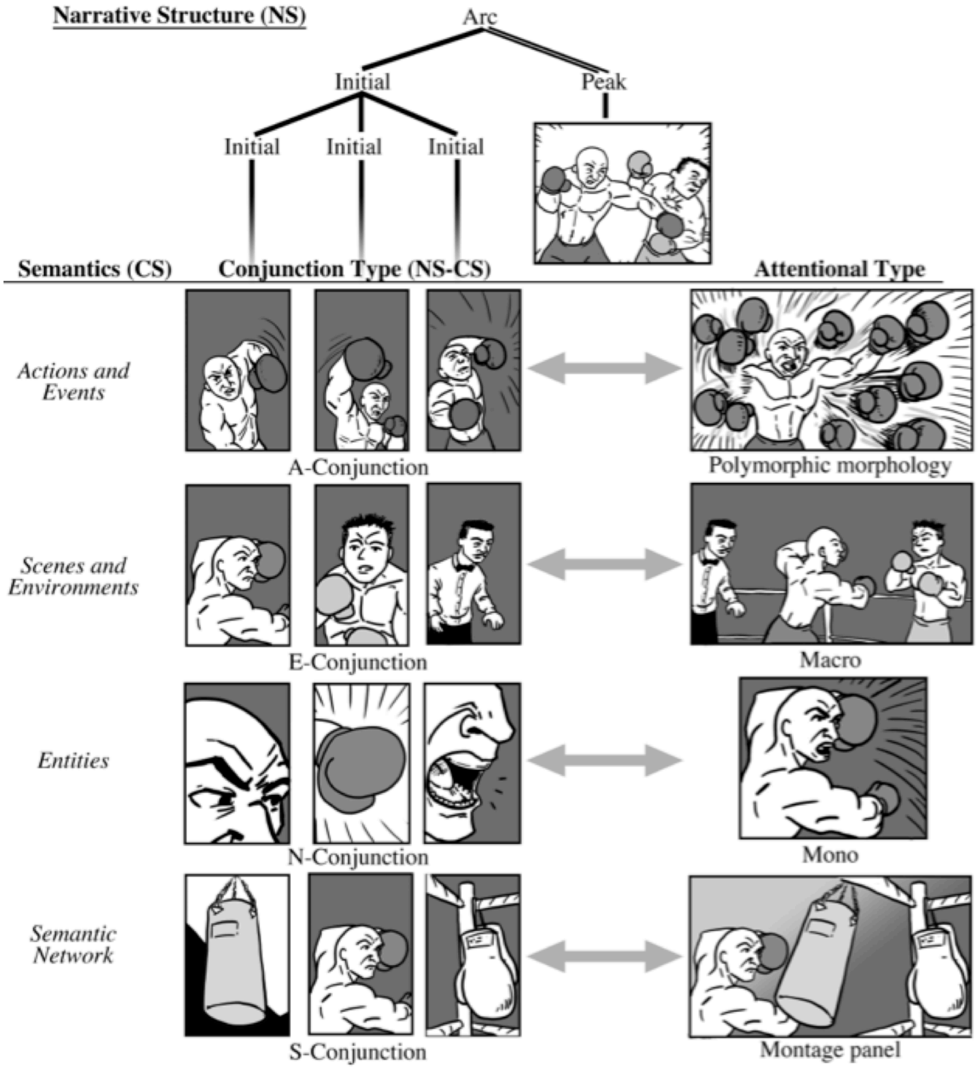

Figure 7: Various types of meaning expressed either through an NS-CS interface (such as in Conjunction) or through a single panel using a particular framing.

information could be conveyed in a single panel as a "montage"—a blended or layered representation.

As depicted in Figure 7, all of these interfaces provide options for the ways in which information is framed—either through individual panels or across sequences of panels. This highlights that there are numerous options for showing the same semantic information. If a creator of visual narratives wanted to covey a whole scene, they then have a choice: Do they want to show the scene as a whole in a Macro? Do they want to highlight portions of a scene, and leave their reader to infer the scene as a "virtual Macro"? Both options convey similar conceptual information by highlighting (or muting) aspects of that meaning through the framing provided either in or across panels.

Comprehension and diversity

Though empirical research has yet to explore the cognition of Conjunction explicitly, some work has suggested that the division of characters into individual panels using E-Conjunction elicits more predictions about subsequent events than single Macro panels (Kaiser & Li, 2013). In addition, altering the position of panels within a Conjunction phase has little effect on subsequent panels, but only if those panels have no distinguishable difference in their semantic roles. For example, the order of conjoined Establishers depicting two characters at passive states had no discernable effect on viewing times to subsequent panels (Cohn & Paczynski, 2013). However, subsequent Peak panels were viewed faster when preceding conjoined Initials were presented in an agent-patient order than in a patient-agent order. Nevertheless, these alterations did not impact felicity ratings of the sequences. These findings suggest that the ordering of semantic components involved in Conjunction phases can impact later parts of the sequence, though the felicity as a whole is not affected.

A recent study also showed that individuals who are inexperienced with watching films, coming from a remote village in Turkey, had difficulty understanding that film shots of individual characters were meant to be understood simultaneously in a common environment (Schwan & Ildirar, 2010). In other words, the comprehension of E-Conjunction may require proficiency in the grammar of the visual language. Additional evidence has also suggested that the proportional use of E-Conjunction may differ crossculturally. Corpus analyses have revealed that Japanese manga use substantially more proportions of Monos and Micros than American comics, which use equal if not greater amounts of Macros than Monos (Cohn, 2011; Cohn et al., 2012). This higher proportion of Monos in Japanese manga suggests that their narrative grammar uses more E-Conjunction than the system used in American comics, and thus readers of manga may be more habituated to such conventions in drawn visual narratives.

Altogether, this research suggests that E-Conjunction as a facet of narrative grammars may differ across cultures, and that its comprehension may be tied to familiarity with these specific conventions. Further research analyzing both corpus data and psychological measures will be needed to explore these issues regarding E-Conjunction and other types of Conjunction in more depth. However, the outlines of this model provide the necessary structure to make predictions about such processing with regards to inference and structure.

Conclusion

Conjunction allows narrative constituents to be composed of several panels of the same category. However, this simple narrative structure allows for significant complexity via the interfaces made by these panels to a conceptual structure. These interfaces demand inferences in the construction of even larger meaning beyond the represented panels. In addition, this complex interface between structure and meaning helps explain how, as in Figure 7, a series of panels can play the same functional role while conveying different semantic information. In these cases, the narrative structure all uses Conjunction while the meaning changes based on the interface to semantics. All of this structure extends beyond the linear connections between panels, and provides further support for the necessity of a model that separates narrative structure and meaning, and organizes that structure into hierarchic constituents.

Graphic References

Mignola, M., R. Sook, et al. 2003. *Mike Mignola's B.P.R.D.: Hollow Earth & Other Stories.* Milwaukie, OR: Dark Horse Comics.

Schulz, C. M. 2004. *The Complete Peanuts: 1953-1954.* Edited by G. Groth. Seattle, WA: Fantagraphics Books.

Takehiko, I. 2004. *Vagabond*, Vol. 19. Japan: Kodansha

References

Carroll, J. M. (1980). *Toward a Structural Psychology of Cinema*. The Hague: Mouton

Cohn, N. (2011). A different kind of cultural frame: An analysis of panels in American comics and Japanese manga. *Image [&] Narrative, 12*(1), 120-134.

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. doi: 10.1111/cogs.12016

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognitive Psychology, 67*(3), 73-97. doi: http://dx.doi.org/10.1016/j.cogpsych.2013.07.002

Cohn, N., Taylor-Weiner, A., & Grossman, S. (2012). Framing Attention in Japanese and American Comics: Cross-cultural Differences in Attentional Structure. *Frontiers in Psychology - Cultural Psychology, 3*, 1-12. doi: 10.3389/fpsyg.2012.00349

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackendoff, R. (2007). *Language, Consciousness, Culture: Essays on Mental Structure (Jean Nicod Lectures)*. Cambridge, MA: MIT Press.

Kaiser, E., & Li, D. C.-H. (2013, March 21-23, 2013). *Visuospatial grouping influences expectations about upcoming discourse.* Paper presented at the 26th Annual CUNY Conference on Human Sentence Processing, Columbia, SC.

Kuleshov, L. (1974). *Kuleshov on Film: Writings of Lev Kuleshov* (R. Levaco, Trans.). Berkeley: University of California Press.

Magliano, J. P., & Zacks, J. M. (2011). The Impact of Continuity Editing in Narrative Film on Event Segmentation. *Cognitive Science, 35*(8), 1489-1517.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*, 111-151.

McCloud, S. (1993). *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.

Saraceni, M. (2000). *Language Beyond Language: Comics as Verbo-Visual Texts.* (Doctoral Dissertation), University of Nottingham, Nottingham.

Saraceni, M. (2001). Relatedness: Aspects of textual connectivity in comics. In J. Baetens (Ed.), *The Graphic Novel* (pp. 167-179). Leuven: Leuven University Press.

Schwan, S., & Ildirar, S. (2010). Watching Film for the First Time: How Adult Viewers Interpret Perceptual Discontinuities in Film. *Psychological Science, 21*(7), 970-976. doi: 10.1177/0956797610372632

van Dijk, T., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.