Available online at www.sciencedirect.com

ScienceDirect

Journal of Pragmatics 88 (2015) 105–132

journal of PRAGMATICS

www.elsevier.com/locate/pragma

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

Received 13 December 2014; received in revised form 1 September 2015; accepted 4 September 2015

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.

Keywords: Visual narrative; Inference; Situation model; Comics; Film; Visual language

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.

http://dx.doi.org/10.1016/j.pragma.2015.09.001
0378-2166/

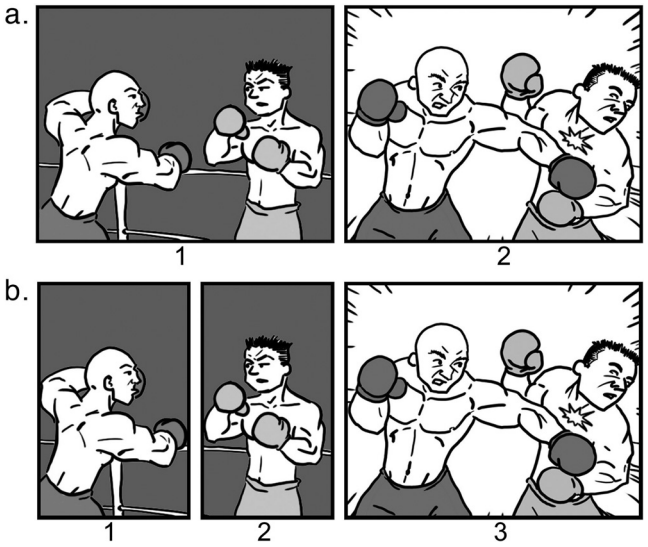

Fig. 1. Variation in how panels provide a ''window'' on a visual narrative scene.

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

as story grammars (e.g., Mandler and Johnson, 1977; Rumelhart, 1975; Stein and Nezworski, 1978; Thorndyke, 1977) and generative grammars of film (e.g., Carroll, 1980; Colin, 1995b). The conflation of structure or meaning in these models may stem from ambiguity intrinsic to their source of inspiration, Chomksyan phrase structure grammars (e.g., Chomsky, 1965), and/or from methods of experimentation, such as memory tasks, which retain semantics but not structure (van Dijk and Kintsch, 1983).

This demarcation between components also distinguishes VNG from theories focusing solely on the meaningful relationships between images in sequence. Some approaches emphasize the *linear* semantic relationships between images or film shots (Eisenstein, 1942; McCloud, 1993; Saraceni, 2000). As applied in psychological research (Magliano et al., 2001; Magliano and Zacks, 2011), these theories have drawn from models of discourse (Zwaan and Radvansky, 1998) to show that discontinuity between images alter the *situation model* of a narrative—the overall meaning constructed in the mind throughout understanding a text (van Dijk and Kintsch, 1983). Other approaches have posited either pairwise relations (van Leeuwen, 1991) or hierarchic structure derived from the semantic/discourse connections between images (Bateman and Schmidt, 2012; Bateman and Wildfeuer, 2014a, 2014b), again importing constructs from linguistic discourse models (Asher and Lascarides, 2003; Halliday and Hasan, 1976; Martin, 1983), though without utilizing psychological experimentation.

VNG shares features with many of these approaches (see Cohn, 2013b), and attempts to integrate their insights into a broader model, especially psychological findings (e.g., Magliano et al., 2001; Magliano and Zacks, 2011; Mandler and Johnson, 1977). However, the clear demarcation between narrative grammar and semantics in VNG is both consistent with empirical research (Cohn et al., 2014, 2012a) and allows for balancing seemingly-contradictory observations made in different approaches (e.g., patterned categorical sequencing entrenched in memory versus semantic relations computed spontaneously). In addition, this separation will allow for a primary argument herein: that a single structural pattern (conjunction) maps to numerous types of semantic information.

2.1. Constructs of VNG

Below, we will discuss the components of the parallel architecture for visual narratives more fully, but first we will define the constructs of Visual Narrative Grammar and their mapping to conceptual (semantic) structures. VNG uses several basic narrative categories with prototypical correspondences to semantics:

Establisher (E) – sets up an interaction without acting upon it, often as a passive state.
Initial (I) – initiates the tension of the narrative arc, prototypically a preparatory action and/or a source of a path.
Prolongation (L) – marks a medial state or extension, often the trajectory of a path or a ''pause'' with a passive state.
Peak (P) – marks the height of narrative tension and point of maximal event structure, prototypically a completed action and/or goal of a path, but also often an interrupted action.
Release (R) – releases the tension of the interaction, prototypically the coda or aftermath of an action.

These descriptions outline prototypical mappings between the semantic information cued by the ''morphology'' of visual images' content and the structural narrative categories. However, other semantic information can correspond to narrative categories in non-prototypical ways (Cohn, 2013b, 2014b). Altogether these narrative categories are organized into phases, which use a canonical sequence pattern of:

Canonical narrative schema
$[_{\text{Phase X}}$ (Establisher) − (Initial − (Prolongation)) − Peak − (Release)]

This schema states that a ''Phase'' (a constituent) consists of these narrative categories in this order. The parentheses indicate which categories are non-obligatory. Because a sequence is motivated by the events of a Peak, it is the only non-obligatory element (though it can also be omitted in certain tightly constrained, inference-generating situations). This schema is not a ''rule'' in the sense of traditional phrase structure grammars (e.g., Carroll, 1980: for film; Chomsky, 1965: for syntax; Mandler and Johnson, 1977: for narrative). Rather, this canonical narrative sequence is a ''construction'' stored in memory as an abstract schematic pattern, akin to syntactic patterns stored in the lexicon of language (Culicover and Jackendoff, 2005; Goldberg, 1995; Jackendoff, 2002). Because VNG is a construction grammar, it therefore allows both abstract schematic patterns (described throughout), as well as systematic idiomatic patterns that may or may not be comprised of these abstract schemas (for examples, see Cohn, 2013a). This model thus differs from approaches positing only spontaneously computed relations between images (Bateman and Wildfeuer, 2014a, 2014b; Magliano and Zacks, 2011; McCloud, 1993; Saraceni, 2001), where no sequencing patterns would thus be entrenched in memory. However, this approach may find precedents in taxonomies outlining patterned sequential image relations, particularly as applied to film (for review, see Bateman, 2007; e.g., Branigan, 1992; Metz, 1974).

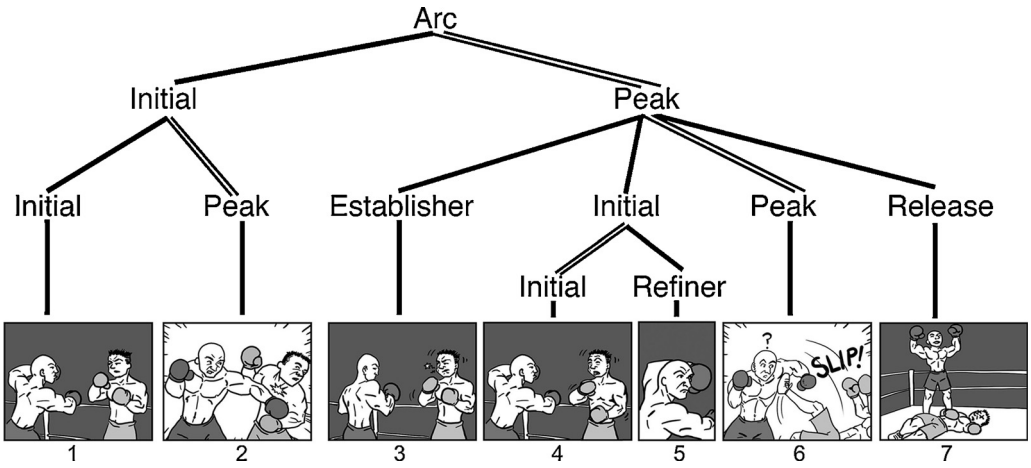

Fig. 2. A narrative sequence with constituent structure.

VNG identifies narrative categories through the interaction between bottom-up semantic cues (as described above) and a top-down context in the broader sequence (specified by the canonical schema). For example, preparatory actions depicted in a panel would prototypically map to an Initial in narrative structure, which must be confirmed by its context, canonically following an Establisher in the narrative schema. If the panel had non-prototypical semantic information, it still could potentially be identified as an Initial if its semantically congruous information followed an Establisher and preceded a Peak, where the context would determine the category. Again, this relationship is structurally analogous to grammatical categories for words in sentences, though they convey semantic information at different levels. While in isolation the phonological string ''hit'' conjures various semantic meanings and presumed syntactic roles, its grammatical category also relies on its phrasal context: as a noun (*Give me a hit of scotch*), a verb (*He hit the wall*), or an adjective (*It was a hit record*). This combination of bottom-up content and top-down context is also important because, like words in a sentence, some images can play multiple roles in a sequence (Cohn, 2014b).

In addition, these categories do not just characterize the narrative roles of individual panels, but each category can expand into its own phase. This can best be illustrated with an example. Fig. 2 adds more panels to Fig. 1a. The first panel depicts one boxer reaching back in a preparatory action, prototypical of an Initial. The second panel shows this boxer completing the action of punching, prototypical of a Peak. The sequence then resets in panel 3, an Establisher, here ''reintroducing'' the new situation after the first punch, where the boxers now stand adjacent to each other in a (relatively) passive state. Panel 4 then shows another preparatory action (an Initial). Next, a zoom on the punching boxer repeats the information in this Initial, posited as spatial modifier called a ***Refiner*** (discussed below). The penultimate panel, a Peak, does not have a completed action like the previous one, but rather shows an unexpected interruption. Finally, the last panel dissipates the tension of this Peak as a Release showing the opponent knocked out.

Complexity can be introduced to a narrative sequence in several ways. First, each category can be expanded to constitute its own constituent. Any grouping of panels can play a role in a larger structure, and a constituent with no role in a larger structure is an ''Arc''—the maximal node. In Fig. 2, the first two panels form an Initial that sets up the final five panels, which are the Peak of the overall narrative. The double-barred lines from all the Peaks indicate ''headedness''—their content motivates the category of their higher constituent and their local sequence often ''hangs off'' of them. Thus, narrative categories apply to both individual panels and whole constituents, a recursive structure.

Categories can also expand through modifiers. The second Initial becomes a constituent by using a Refiner, which repeats information from a previous panel to provide a more focused viewpoint (Cohn, 2013a). Here, the refined viewpoint depicts only the punching boxer. Because this panel ''modifies'' the previous one, the larger viewpoint panel becomes the ''head'' (double bar lines) with the Refiner as its modifier:

Refiner schema

$[_{\text{Phase X}}$ (Refiner) − X − (Refiner)]

Thus, any category (X) can be expanded with modifiers (Refiners) on either side. Refiners are identified as relative to a head—without a prior panel, the Refiner here would become the Initial. This is structurally analogous to certain adjectives (a syntactic modifier), which can become nouns, like when *red* in *I'll take the red* means *red wine* (Cohn, 2013a). A sequence can also add complexity by repeating a category several times, often by breaking a panel into its component parts, as in Fig. 1b. We will explore this phenomenon throughout the rest of this paper.

Finally, it is worth mentioning experimental evidence has supported that these hierarchic relations in sequential images do not arise from the spontaneously computed semantic relationships between images alone, be it semantic discontinuity as signals for constituent breaks (Magliano and Zacks, 2011) or hierarchy derived solely from meaningful relations between images (Bateman and Wildfeuer, 2014a, 2014b). First, the neural responses appearing to violations of this narrative grammar (Cohn et al., 2014) are similar to those typically shown in violations of syntax in sentences, not semantics (Friederici, 2002; Hagoort, 2003). Second, the neural mechanisms for semantic processing are not sensitive to the constructs of this narrative grammar in the absence of semantic associations between images (Cohn et al., 2012a). Third, disruptions placed within constituents versus between constituents yield differences in neural responses *prior* to the first image of the second constituent (Cohn et al., 2014). Because comprehenders could not yet reach a subsequent panel to define this semantic relation, this means that they were using cues within images to make predictions about the upcoming constituent structure. This thereby provides strong evidence against models defined solely by the relations between images' content. Nevertheless, VNG hypothesizes that this relational content interfaces with narrative in predictable ways. For example, breaks between narrative constituents often (but not always) align with changes in characters or spatial location (Cohn, 2013b; Magliano and Zacks, 2011), despite not relying on them alone (Cohn et al., 2014). Thus, this *semantic* information interacts in parallel with the *structural* information in VNG to lead to a broader understanding of the sequence.

2.2. The parallel architecture

Having defined the basic constructs of Visual Narrative Grammar, let's now address the structural difference between Fig. 1a and b. We now expand the parallel architecture for visual narratives to four components: *graphic structure*, *conceptual/event structure*, *spatial structure*, and *narrative structure*. Reanalysis of Fig. 1a can illustrate these canonical mappings between structures, depicted in Fig. 3a.

As a visual-graphic modality, drawn narratives must manipulate aspects of lines and shapes to convey meaning. ***Graphic structure*** governs the constraints on the physical structure of lines and shapes, in the same way that phonological structure governs articulated sound (Cohn, 2013a; Willats, 1997). Graphics-meaning mappings can result in a ''morphological'' structure (Cohn, 2013a), again comparable to the sound-meaning mappings in spoken language (Jackendoff, 2002). For example, a combination of physical lines within panel 1 are understood as a boxer reaching back his arm, a preparatory action. We could further notate Fig. 3 to include these elements (ex. each character marked with indices to Agent and Patient, the Agent's arm in panel 1 mapped to REACH BACK, etc.), but are excluded for clarity. I will not elaborate on the underlying structure of these components for simplicity, and instead default to showing the representation of the sequence for ''graphic structure.'' Nevertheless, any cues about meaning that may be relevant for narrative structure are ultimately extracted from the graphic/morphological structure of the physical representations.

Conceptual structure comprises the understanding of meaning, such as elements like states, events, objects, places, paths, etc. (Jackendoff, 1990, 2002; Zwaan and Radvansky, 1998). Within this, ***event structures*** specify the meaning of the situations and events that take place in and between images. Simple events have been hypothesized as comprising two types. Discrete events have a preparatory action (*reaching back*), a completion in the event's ''head'' (*punching*), and a coda (*withdrawing arm*), while continuous processes (*running, dancing, walking*) end in a termination (*stopping*) (Jackendoff, 2007). Like in narrative, events themselves are hierarchic and recursive (Jackendoff, 2007; Zacks and

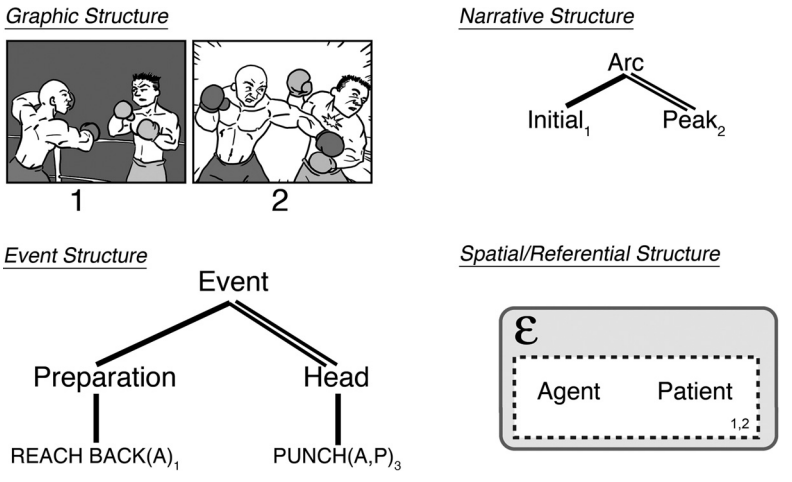

Fig. 3. Parallel architecture diagram for the structures of Fig. 1a.

Tversky, 2001; Zacks et al., 2001). For example, the process of walking involves repeating the discrete parts of lifting a leg (preparation), extending (head) and shifting weight (coda). In this view, the well-known ''scripts'' for events (Schank and Abelson, 1977) are essentially ''lexicalized'' versions of event structure stored in a ''vocabulary'' of event knowledge (Sacerdoti, 1977), the same way that constructions of syntax (or narrative) are ''rules'' stored as part of the lexicon of language (Goldberg, 1995; Jackendoff, 2002).

Fig. 3 focuses on the primary actions motivated by the agent of the action (the puncher), since research has suggested that agents motivate the comprehension of event structures (Cohn and Paczynski, 2013).[1] Panel 1 shows a preparatory action of the agent reaching back, which then completes as the ''head'' event of a punch in panel 2. These conceptual/event structures are the semantic elements hypothesized above to map to structural narrative categories (e.g., graphic structures depict preparatory actions in event structure, which map to Initials in narrative structure).

Beyond propositional information about entities and their actions, meaning also involves a ***spatial structure***. While conceptual/event structure is a type of propositional algebraic semantics, spatial structure is a geometric type of meaning (Jackendoff, 1987, 2002; Zwaan, 2004). For example, the characters in Fig. 3 are 2D images giving the illusion of 3D shapes, and these elements belong to a larger spatial environment. Panels thus serve as ''attention units'' to highlight different aspects of a scene, and we can characterize panels based on how much information they contain (Cohn, 2007, 2013a):

Macros – depict multiple interacting elements.
Monos – show only one active entity.
Micros – show less than one entity (usually with a close up).

These categories characterize the interfaces between graphic structure (the physical panel), spatial structure (spatial meaning), and narrative structure (sequential presentation) (Cohn, 2014a). In Fig. 3 both panels are Macros, depicting the complete spatial structure with multiple characters (indicated by subscripts for panels 1 and 2 in a dotted box). The ''full environment'' will be notated with an epsilon (ε). It is worth noting that, while they are somewhat similar, these categories do not necessarily equate to filmic shot types like long, full, medium, close, and close up shots (Bordwell and Thompson, 1997). Filmic shots modify the graphic structure itself, determining how to present visual information (*How should characters be depicted?*), while attentional framing categories specify how much information is relevant to convey meaning (*Should a single character or multiple characters be depicted?*). Thus, a Mono with a single character could be depicted many ways: a whole body (full shot), half body (medium shot), a bust (close shot), etc.

Finally, the ***narrative structure*** (i.e., VNG) organizes the meaning (event and spatial structures) into a coherent sequence, as discussed above. Here, the narrative structure is fairly simple: the preparatory action in panel 1 maps to an Initial, while the completed action maps to a Peak in panel 2. These are both prototypical mappings of conceptual/event structure to narrative.

The comprehension of the sequence is posited as involving the interaction between these components. The graphic and narrative structures modulate the *presentation* of the sequence, while the event/conceptual and spatial structures comprise its *meaning*. The sum total of the event/conceptual and spatial structures comprises the information which is incorporated in memory into a *situation model* (van Dijk and Kintsch, 1983; Zwaan and Radvansky, 1998)—the constructed conception of the meaning of the discourse. Meanwhile, the graphic and narrative structures function to package and convey that meaning (its *textbase*).

Now let's consider the changes that occur when splitting up panel 1, as in Fig. 1b, now represented in Fig. 4. The basic event structures remain the same: The agent still reaches back and punches the 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 are each Monos, showing an individual character, while panel 3 remains a Macro depicting the whole scene (indicated by the panel numbers indexing each structure throughout).

This alteration in spatial structure also changes the narrative structure. The overarching narrative category remains the same—it is still an Initial—only it divides into subordinate Initials, forming a ***conjunction*** phase. In this case, the higher-level node remains mapped to the overall environment, just like the single first panel in Fig. 3. However, now this larger environment is inferred, and the individual panels map to *parts* of the spatial structure that highlight each character (again, subscripts marking the interface points between structures). This inference does not ''fill in the gaps'' for the juxtaposed relations *between* panels (e.g., McCloud, 1993), but rather the two panels inferentially build a ''virtual'' environment out of their parts and map this mental model to the narrative constituent (notated with ''e''). Thus, while the basic semantic parts of the sequence remain the same (objects, events), dividing the first image should create additional narrative (conjunction) and semantic (inference) demands.

[1] In the full model, all entities involved in an action have their own tree structures (i.e., here both agent and patient would have their own event structures). These aspects of event structure will be elaborated in a later publication. For simplicity, I omit this more complex representation, and trust the diagrams here can be understood without excessive formal elaboration.

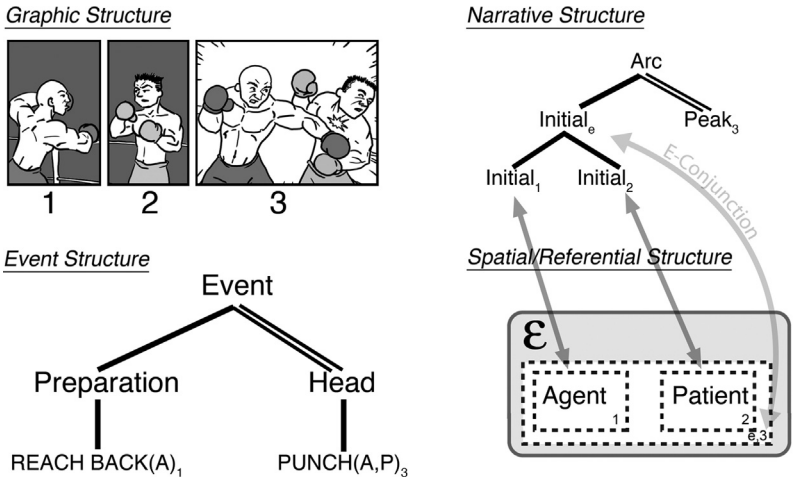

Fig. 4. Parallel architecture for the structures of Fig. 1b.

VNG identifies this altered narrative pattern as a ''conjunction'' phase (Cohn, 2013b), whereby a single constituent contains any number of daughters of the same category, in line with conjunction phrases in syntax (Culicover and Jackendoff, 2005):

Conjunction schema

$[_{\text{Phase X}}\ X_1 - X_2 - \ldots X_n]$

Again, because VNG is a construction grammar, this basic schema is posited as stored in memory as a pattern, like the canonical arc, and can apply to any category within the narrative grammar. Note also that, as formalized, this is a purely narrative pattern specifying only a repetition of narrative units, with no specific semantic information. The described mapping of different conjoined characters to a spatial environment is only one mapping between conjunction (a narrative function) and meaning (semantics). The remainder of this paper will argue that this conjunction schema can map to various types of semantic information: environments, entities, actions, and semantic networks. However, the abstract quality of this narrative schema allows the potential for other mappings to semantic information not outlined here.

3. Narrative and semantic mappings in conjunction

3.1. Logic of the methodology

We now turn to exploring conjunction in visual narratives, characterized by different types of mappings to semantics. Conjunction is hypothesized as having three basic traits: (1) Conjunction unites panels into a contiguous constituent. (2) The panels that constitute this constituent have the same narrative category. (3) These panels generate a superordinate conceptualization comprised of the component parts of the conjoined units. Hypotheses #1 and #2 relate to the properties of the conjunction schema, while trait #3 relates to the mapping of that schema to semantics.

In order to provide evidence for these hypotheses, we turn to diagnostic tests which have been used in linguistics for decades (e.g., Cheng and Corver, 2013) that manipulate the structure of sequences, such as through movement, deletion, and substitution of units. Insofar as VNG posits a construction grammar, ''ungrammaticality'' does not arise from errors in a rule-driven derivation (Chomsky, 1965), but rather from deviations from a patterned schema and/or mismatches between the constraints of interacting parallel structures (Culicover and Jackendoff, 2005). In using diagnostics, readers will thus be asked to assess the coherence of sequences based on their own intuitions. However, because expertise modulates the comprehension of visual narratives (Cohn, 2013a; Cohn et al., 2012a; Nakazawa, 2005, 2015), readers with varying fluency—both generally and possibly for specific cultural visual narrative systems (Cohn, 2013a)—may be more or less sensitive to the preferences constraining the coherence of visual sequences (Buckland, 2000). Nevertheless, this does not mean that *all* sequences of images may be ''acceptable'' or ''meaningful'' (Bateman and Wildfeuer, 2014a; McCloud, 1993; Saraceni, 2001), and the incoherence of various manipulations to sequential images has been supported empirically (e.g., Cohn et al., 2012a; Sitnikova et al., 2008; West and Holcomb, 2002).

Here, we focus on three diagnostics described previously for visual narratives (Cohn, 2013b, 2014a): *movement*, *deletion*, or *substitution*. While diagnostics can test the constructs found within a given sequence, such manipulations

should also reveal these structural constructs in the first place. This analysis is therefore informed by the underlying logic of such diagnostics, described below. Ultimately though, the theoretical constructs resulting from these methods are intended to frame empirical examination that can validate and clarify these testable claims, and such projects are already underway.

3.1.1. Movement

Because conjunctions should form a contiguous constituent (Hypothesis #1), panels within a conjunction phase should not be alterable with those outside the conjunction phase. Such rearrangement would violate the constituent boundaries, and thereby change aspects of the sequence *other than* the proposed conjunction (sometimes, but not always, altering or violating the event structure). However, because all panels in a conjunction phase should belong to the same narrative category (Hypothesis #2), rearranging panels within a conjunction phase should have little effect on their status as a grouping. Fig. 5b rearranges the conjoined panels in Fig. 5a, resulting in little change in the structure. However, moving a panel outside the conjunction phase (Fig. 5c) results in an awkward sequence.

Such diagnostics echo experimental findings that, while rearranged panels within constituents were less coherent than their original sequences (Cohn, 2014b), participants were more sensitive to rearrangements that crossed between constituents than those remaining within constituents (Hagmann and Cohn, under review). While the expectations of within-constituent movements are different for conjunction than a standard narrative constituent (because conjunction itself places no constraints on order within the constituent), the expectations of cross-constituent movements should remain the same, since both involve exiting into a separate narrative state.

In other studies examining conjunction specifically, switching the order of conjoined Establishers showing characters in passive states had no discernable effect on self-paced viewing times, but some differences appeared when switching the order of conjoined Initials using an agent-patient order compared with a patient-agent order (Cohn and Paczynski, 2013). Nevertheless, no rearrangements influenced coherence ratings of these sequences. Thus, though the particular content of conjoined panels may modulate their processing, this is independent of manipulations to position and does not affect the overall comprehensibility of a sequence.

Finally, a limitation of the movement diagnostic should be acknowledged. Because some panels' content can play multiple roles in a sequence (Cohn, 2014b), movement of a unit may simply create an alternative, congruent role in a sequence. However, as described above, this is not true of all panel rearrangements (cf., Jahn, 1997), and differentiation of these cases again relies on both the content of panels and their context. Thus, as with all diagnostics, movement should not be relied upon on its own, but rather belongs within a broader suite of tests.

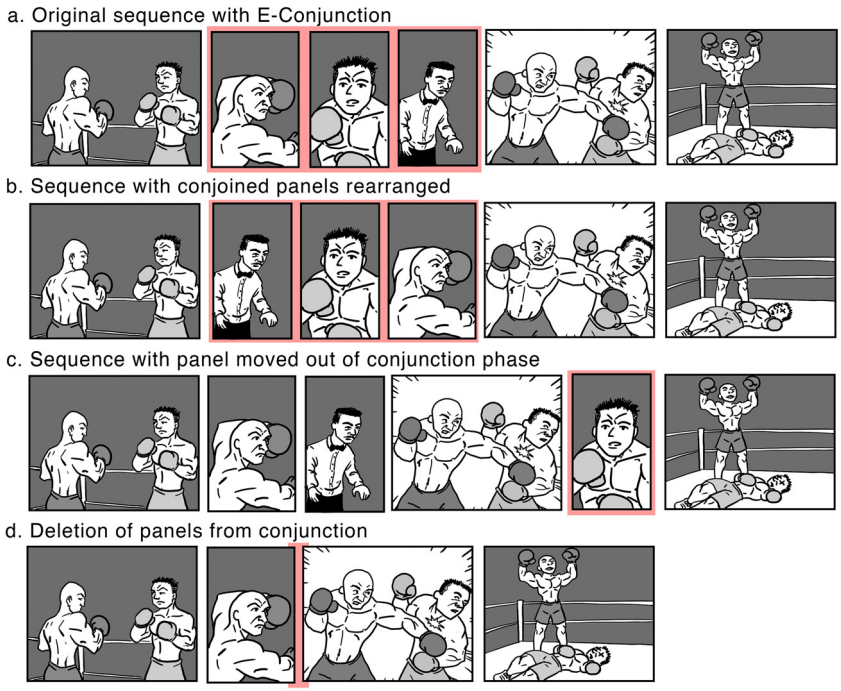

Fig. 5. Application of movement and deletion tests to visual sequences (highlighted).

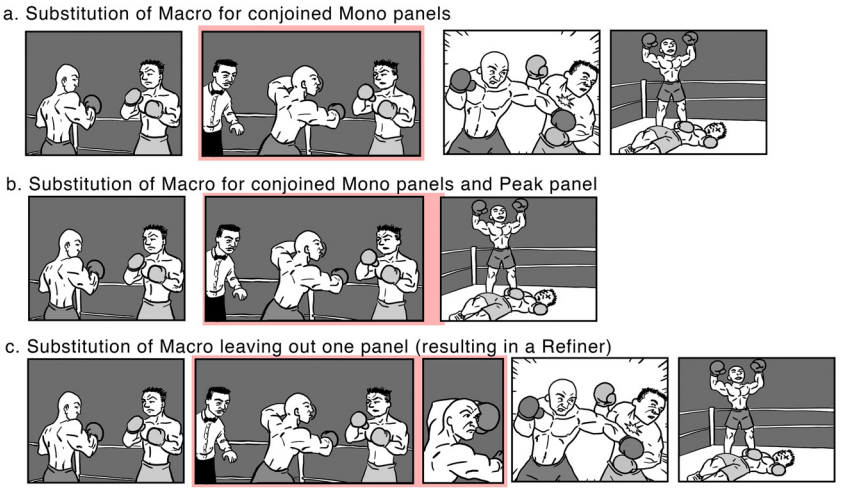

Fig. 6. Substitution test applied to a sequence with E-Conjunction (highlighted in red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.1.2. Deletion

Because all panels within a conjunction phase maintain the same narrative state (Hypothesis #2), omission of panels should retain the overall felicity of the constituent. This occurs in Fig. 5d, where deletion of two panels from the conjunction phase leaves a single Initial without altering its structural role. This is different than for Refiners, where the Refiner repeats a portion of information from its head. If a Refiner were deleted, the head would remain the same category, but if the head were deleted, the Refiner would *become* that dominant category. This asymmetry should not be the case with conjunction, since all panels share the same status.

Deletion tests can also provide information about constituent structures (Hypothesis #1). If multiple panels are deleted, we would expect a coherent sequence as long as they stayed within the conjunction phase, but we should not be able to omit panels across the constituent boundary, therefore affecting narrative states outside the conjunction.

3.1.3. Substitution

Both movement and deletion tests provide diagnostics for investigating Hypotheses #1 and #2. However, they cannot address Hypothesis #3: that the component parts of conjunction phases ''add up'' to a superordinate conception consisting of those parts (cf., Asher and Lascarides, 2003; Bateman and Wildfeuer, 2014a, 2014b). We can examine this hypothesis through a substitution test, where a single unit replaces all the units involved in the conjunction. This test should work because, as discussed previously, different types of conjunction create ''virtual'' panels with a wider attentional framing. In fact, a substitution test is provided in Fig. 1: The single Macro in Fig. 1a ''substitutes'' equivalently for the two Monos in Fig. 1b (and vice versa). It is worth noting that a similar idea to substitution was posited by Metz's (1974:152) ''commutation test'' for filmic shots (roughly similar to replacing two Monos for a Macro), though substitution tests can illustrate more aspects of structure than this precedent, as described below.

Dividing a single Macro panel into multiple Monos should create conjunction. Thus, as in Fig. 6a, substitution of a Macro for conjoined Monos (Fig. 5a) with the equivalent information supports that those Monos are involved in conjunction. However, this substitution should not be able to cross constituent boundaries—if a non-conjoined panel is included in the substitution, it should be less coherent, as in Fig. 6b, where the Macro replaces the Initial panels, but also deletes the Peak.

In addition, if a panel from the conjunction phase is *not* included in the substitution, it leaves information repeated in the substituted panel—by definition a Refiner. In Fig. 6c, the Refiner is a Mono that modifies the larger-scope Macro, both conveying a similar narrative state. Thus, if the substitution test leaves a remaining Refiner, the substitution did not fully ''absorb'' all conjoined panels.[2] This provides another test for determining the boundaries of the conjunction phase (Hypothesis #1). As will be discussed, because conjunction can use different semantic interfaces, substitution tests require particular types of panels in order to be successful (see Section 4).

[2] A diagnostic for Refiners should also be noted: the content of a Refiner should be capable of being highlighted in the ''head'' using an ''inset'' panel (a panel within another panel). Essentially, insets that frame information within a dominant panel are Refiners that occur without repeating information in a separate panel (Cohn, 2014a).

Finally, Hypothesis #3—whereby the units in a conjunction generate a broader conception consisting of those parts—implies a process of inference. In the examples so far, this has been argued as the inference of a broader environment that consists of those characters (discussed below). A substitution diagnostic test alone cannot inform about whether such inference indeed occurs in online comprehension. However, such a manipulation can form the basis of experimental designs that could test whether comprehenders make such inferences, and indeed such manipulations have already been used to examine the effects of ''framing'' on visual narrative understanding (e.g., Kaiser and Li, 2013).

3.2. Environmental-Conjunction

We now turn to describing various patterned mappings of semantics to conjunction. The examples so far have all discussed conjunction where multiple characters create an inferred spatial environment. In Fig. 1, ***Environmental-Conjunction*** *(E-Conjunction)* unifies the two panels in 1b into a virtual structure equated to the single environment of 1a. One might construe this construction of a spatial environment as a type of bridging inference, which provides the unstated meaning necessary to connect one part of a discourse to another (Haviland and Clark, 1974; McNamara and Magliano, 2009). Yet, the inference here does not necessarily ''connect'' various parts of a narrative, but rather ''adds up'' component parts into a sum greater than what is depicted (Kintsch, 1998). This is also somewhat different from the referential information involved in anaphoric coherence, since no ''anaphor'' is connected to an antecedent (or corresponding referential knowledge in a situation model) that appears prior in the text (Graesser et al., 1994; Magliano and Graesser, 1991). Rather, this is a ''part-whole'' relationship where the component parts are provided, and the whole is constructed out of them.

Also, this inference only constructs a common, holistic spatial structure, and does not necessarily suggest an iconic mapping to any relative placement of elements within that environment. Specific relations between elements may (or may not) arise from the ''morpho-semantic'' cues in the images (e.g., Huff and Schwan, 2012). For example, the two boxers in Fig. 1b are positioned in a way implying that they face each other, even when separated into two panels, and thus they should have a particular spatial relationship within an inferred environment (Colin, 1995a). Yet, if the boxers faced the viewer, they would imply no overt positional relationship. E-Conjunction would operate over both sequences in the same way, though the postural differences would specify their varying relations within that space.

We have already applied our diagnostic tests to E-Conjunction in Figs. 5 and 6, and thus will not provide more here. E-Conjunction can operate on any narrative category. In Fig. 7b, three ninjas throw claws-on-chains in conjoined Initials (i.e., source of the claws' paths), which are deflected or dodged by samurai in the conjoined Peaks (i.e., completed actions, goal of claws' paths). In Fig. 7b, Establishers show a pitcher and catcher in a baseball game, setting up these characters across the field. Both of these examples use Mono panels to direct attention to each character, while a common environment is constructed across panels. Functionally, E-Conjunction provides a way to accentuate different people taking actions (as in Fig. 1b) or to capture detailed views of figures that are separated by a distance (Fig. 7a and b). Note also that, at the highest level of structure, these sequences retain a canonical narrative sequence (7a: *I-P-R*, 7b: *E-I-P-R*), and the conjunction phase merely elaborates on narrative categories within that structure.

In a preceding model to VNG (Cohn, 2003), this unifying function was achieved by an ''environmental phrase'' that captured the spatial inference as a specific singular structure, akin to a phrase structure grammar (e.g., Carroll, 1980: for film; Chomsky, 1965: for syntax; Mandler and Johnson, 1977: for narrative). Similar insights arise in Bateman and Wildfeuer's (2014a, 2014b) application of discourse models (Asher and Lascarides, 2003) which treats this inference as an ''update'' in the dynamic semantics of the sequence. While these approaches share the intuitions that a part-whole inference is necessary, they miss the observation that such spatial inferences operate over panels playing a similar functional role in the overall structure (i.e., narrative categories), as illustrated by diagnostic tests. With a parallel architecture, the separation between structure (narrative) and meaning (inference) allows both observations to occur independently, and also permits mappings between domains that do not involve the construction of a spatial environment, as described below.

In VNG, E-Conjunction reflects this narrative-semantics interface (NS-CS) rather than a purely grammatical or semantic operation. With our narrative conjunction schema already in place, different ''correspondence rules'' can specify the mapping between Narrative Structure and the Conceptual Structure (NS-CS Rule). This correspondence for E-Conjunction is as follows:

NS-CS Rule 1: E-Conjunction

Narrative Structure		Conceptual Structure
$[_{\text{Phase }X}\varepsilon_{/e}\ X_1 - X_2 - \ldots X_n]$	$\Leftrightarrow_{\text{default}}$	$[_{\text{PLACE}}\varepsilon_{/e}\ \{\text{ENTITY}_1, \text{ENTITY}_2, \ldots \text{ENTITY}_n\}]$

$\Leftrightarrow_{\text{default}}$ Spatial Structure

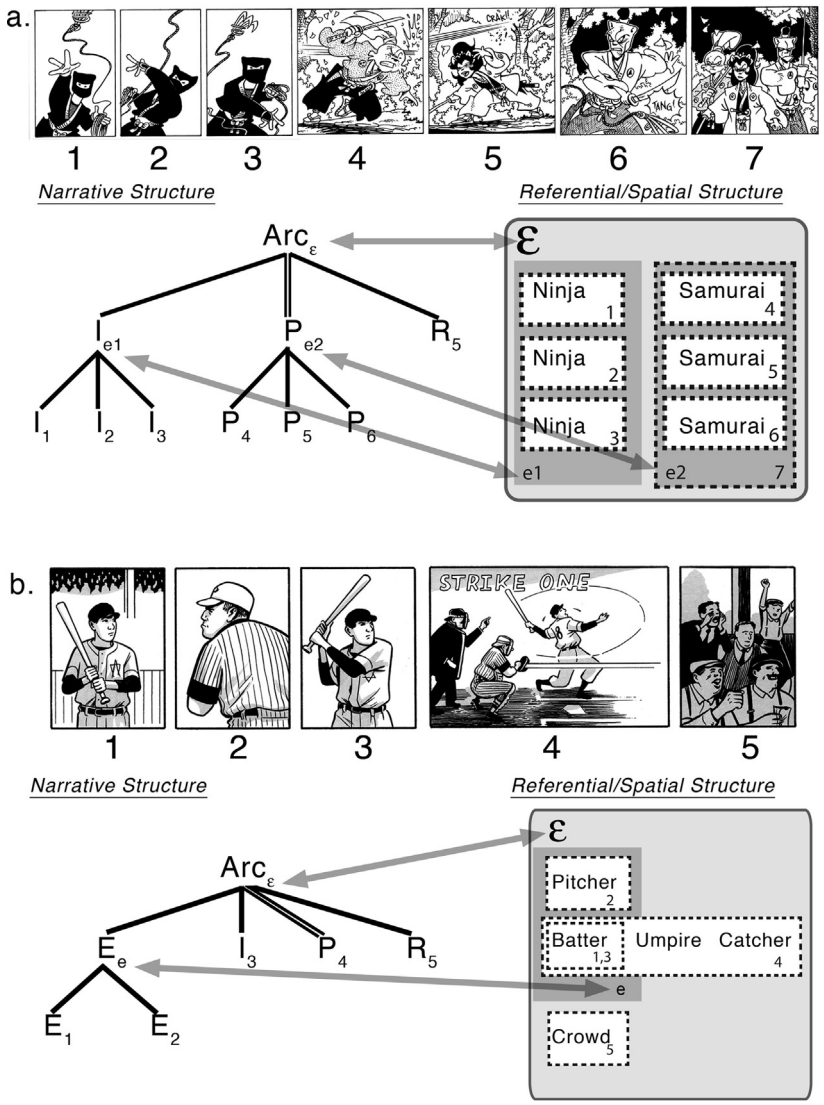

Fig. 7. Visual narratives where Environmental-Conjunction is used as (a) Initial and Peak panels (Sakai, 2008) and (b) Establishers (Sturm, 2003).

The notation here follows that of Culicover and Jackendoff (2005) for correspondence rules within the parallel architecture. The ⇔ indicates the licensing of a correspondence between structures, and the subscript *default* indicates that it is the ''preferred'' or ''unmarked'' mapping. The curly brackets in conceptual structure {. . .} designate flexibility for the order of the enclosed elements (as would be expected from the constituent-internal movement test). Subscript numbers index the interfaces between structures.

Overall, this correspondence rule allows for the entities in conjoined narrative categories within a phase to create a broader environment (PLACE) in conceptual/spatial structure. Thus, the term ''Environmental-Conjunction'' essentially means ''a *conjunction* in narrative structure involving component entities that create an *environment* in conceptual structure.'' This correspondence is diagrammed in Figs. 4 and 7: boxes with dotted lines correspond to actual panels (identified by numbers), while gray boxes correspond to E-Conjunction mappings between the structures. They designate spatial structure built by the concatenation of multiple entities (''e''). As can be seen with the overlap of the boxes, panels can window elements of a scene in several ways, all of which add up to the whole ''mental environment'' (''ε'') in the spatial structure for the scene. Note also that the ''environment'' indexed by conjunction need not be spatially contiguous—a shared ''environment'' can also be conceptualized from entities at a ''physical'' distance (such as talking on the phone to each other), in line with notions of mental models broadly (Johnson-Laird, 1983).

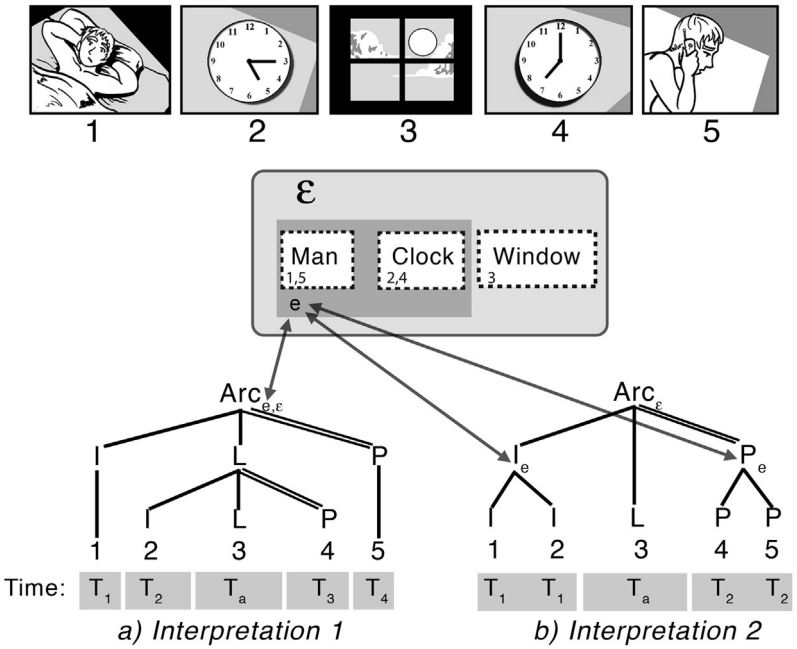

Fig. 8. Ambiguous correspondences between narrative and spatial structures.

It is worth noting that the part-constructing-whole inference proposed in E-Conjunction for spatial environments is not limited to the grammatical construct of conjunction. Such additive spatial inferences could operate across a whole sequence. Both sequences in Fig. 7 require a global part-whole inference for the entire spatial environment. In Fig. 7a, E-Conjunction is used to show ninjas throw claws (conjoined Initials), only to have samurai deflect and dodge those claws (conjoined Peaks). Individually, each phase constructs one set of characters in the scene in opposition to each other (ninjas versus samurai),[3] yet nowhere in this sequence do all characters appear in the same panel (though they do elsewhere in the broader discourse). This means that the broader space containing both ninjas and samurai is constructed globally. The same thing occurs in Fig. 7b: the Establisher phase uses E-Conjunction to unite the pitcher and catcher, while an umpire and catcher appear with the batter in panel 4, and the crowd appears in panel 5. In no panel or phase do all of these entities appear together, but we still understand that they belong to the same overall environment.

Because this part-constructing-whole inference applies both locally (within E-Conjunction) and globally (across a whole sequence) it operates similarly to other types of inference in discourse (Graesser et al., 1994; Kintsch, 1998; McKoon and Ratcliff, 1992; McNamara and Magliano, 2009; van Dijk and Kintsch, 1983). Furthermore, because this inference operates outside of conjunction, it further supports that it is not a facet of the narrative structure, but rather a semantic understanding that maps to this local structural relationship. That is, part-whole inferences operate throughout a visual narrative to integrate conceptual information across a sequence into the broader mental model for this spatial understanding (Rinck, 2005; Zwaan and Radvansky, 1998). Such spatial inferences then may be *localized* with conjunction to a specific constituent within the narrative structure.

These theoretical constructs can aid us in illustrating structural ambiguities that might arise in a sequence. Cohn (2013b) argued that the sequence in Fig. 8 involves at least two possible interpretations because of the ambiguous spatial orientation of the man relative to the clock in panels 1/2, and 4/5. The first two panels are Initials, which both undergo some change of state in the final Peak panels. The central panel is fully ambiguous, but is here marked as a Prolongation ("L"). The ambiguity is whether or not the man and the clock are at the same state in time, which, without any indication otherwise, are assumed to exist in the same spatial environment. Yet, because they are never shown together graphically, this spatial relation is inferred (e).

[3] Interestingly, this structure seems to maintain a serial-order dependency between conjunction phases. We know that the first samurai deflects the claw from the first ninja, the second samurai deflects the claw from the second ninja, and the third the third. This coordination maintains even if the ninja or samurai panels were reordered in their respective conjunction phases: first would still be paired with first, etc., no matter who they are. Thus, this dependency is not a facet of the semantics of each ninja relating with each samurai specifically, but rather it seems motivated by the structure itself. These constraints will be explored in future work.

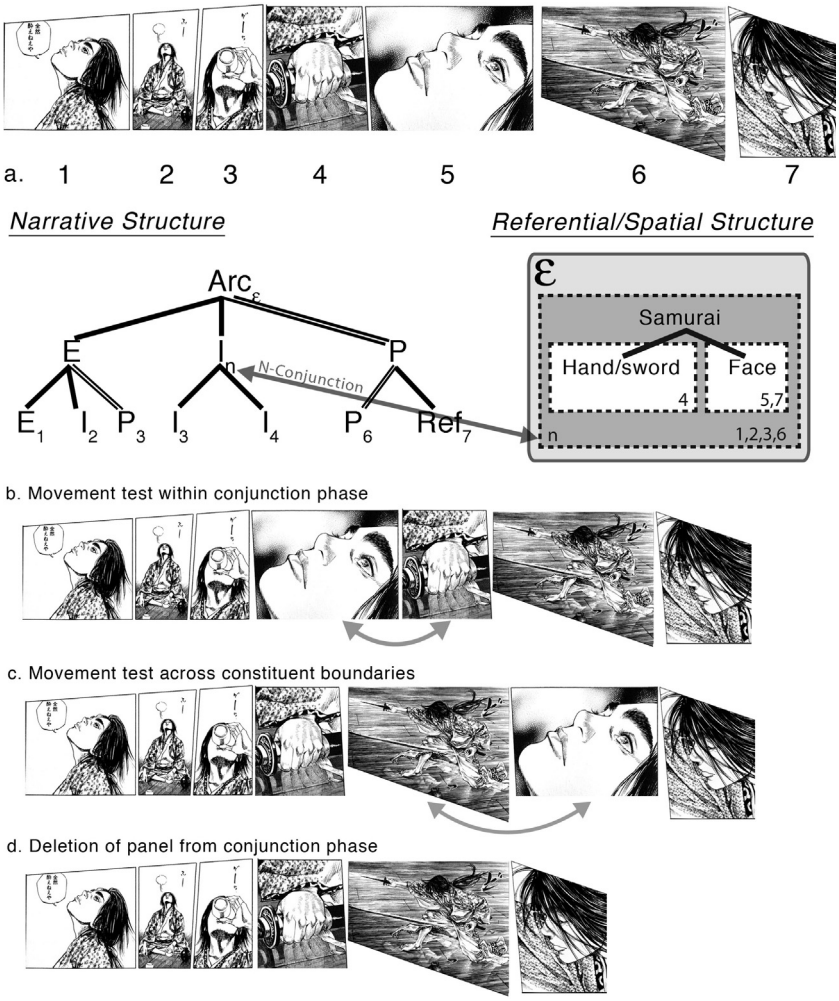

Fig. 9. (a) Entity-Conjunction builds the conception of a single character (in the Initial phase) in Inoue Takehiko's (2003) *Vagabond*, along with diagnostic tests examining its structure (b–d).

In the first interpretation (Fig. 8a), all the panels convey different temporal states: the man occupies his own narrative arc, and the progression of clocks belongs within a center-embedded Prolongation phase. Here, a global part-constructing-whole inference combines the man and clocks into a common environment only at the Arc level. The second interpretation (Fig. 8b) allows each grouping of man and clock to occur at the *same time*, thereby forming a common environment (e) at each conjoined phase.[4] Thus, E-Conjunction, along with the distinction between global and local inference, allows us to differentiate various possibilities in structurally ambiguous surface representations.

3.3. Entity-Conjunction

A similar process to E-Conjunction occurs with a different semantic scope. Rather than conjoining disparate entities within a scene, conjunction can also apply to disparate parts of an individual entity. Take for example the sequence in Fig. 9. Here, a samurai sits in contemplation. He begins by drinking tea (in an Establisher phase), then grabbing his sword while looking at the ceiling (an Initial phase), and finally drawing it rapidly (a Peak) with a Refiner zooming on his face at the end. This sequence is fairly straightforward except for panels 4 and 5—two Initials—which show him grasp the sword at his side in preparation to move. These panels only show his hand and face, but not both in the same image, and yet we know

[4] As stated, the windows are fully ambiguous here and could potentially be grouped in any number of constituents. I take the simplest analysis here and leave it isolated.

that both panels depict the same person. Though they are conjoined, these two Micro panels (panels 4 and 5) do not use E-Conjunction, because they depict the same character. Rather, these panels use a part-whole relationship to construct the notion of a single entity rather than a whole scene.

Where E-Conjunction repeated categories in order to integrate multiple entities into a broader environment, ***Entity-Conjunction*** *(N-Conjunction)* unites panels showing parts of a character into a singular entity. N-Conjunction thus uses the same part-constructing-whole inference as E-Conjunction, but differs in the level of the semantics being expressed: individual entities (characters or objects) instead of whole spatial scenes. Because of this slight difference, our correspondence rule only requires minimal change:

NS-CS Rule 2: N-Conjunction

Narrative Structure		Conceptual Structure
$[_{\text{Phase } Xn}\ X_1 - X_2 - \ldots X_n]$	$\Leftrightarrow_{\text{default}}$	$[_{\text{ENTITY}n}\ \{\text{PART}_1, \text{PART}_2, \ldots \text{PART}_n\}]$
	$\Leftrightarrow_{\text{default}}$	Spatial Structure

Here, the narrative conjunction schema stays the same as in E-Conjunction, while the only change is the mapping between form and meaning: the interface connects to entities instead of environments in conceptual structure. The parts of the entity map to each panel, with the ''virtual'' whole entity map to the phase node (notated ''n''). We can see this correspondence to meaning in Fig. 9. The full ''Samurai'' entity is constructed out of the two Initials, but also given in the rest of the panels of the sequence—hence a dotted line around a gray box in the referential/spatial structure.

To further confirm that this sequence does indeed use conjunction—despite the different semantics than E-Conjunction—diagnostic tests are provided in Fig. 9b–d. First, moving the position of the conjoined Initial panels has little effect on the sequence (Fig. 9b). However, moving the face from out of the conjunction phase into the Peak phase makes the sequence less coherent (Fig. 9c). This violation occurs because the panel crossed a constituent boundary, leading also to a disruption of the contiguity of the head-modifier relationship between Peak and Refiner. Finally, omission of a panel from the conjunction (Fig. 9d) has little effect on the sequence: The remaining Initial plays its same role, only it does so without conjoining to another panel. Though not depicted here, one can imagine a substitution test where these two conjoined panels would form a single panel of the character looking up while grabbing his sword.

3.4. Action-Conjunction

Beyond scenes (E-Conjunction) and referential entities (N-Conjunction), event information can also map to conjunction phases, as in Fig. 10. This sequence depicts a woman in a room (Establisher) where several panels show her conjuring fire (Initials) before the light extinguishes (Peak). This repetition does not show various characters in an environment (it shows one character) or the parts of a single character. Rather, this repetition is an ***Action-Conjunction*** *(A-Conjunction)*, where the interface connects to *event* structures as opposed to *referential* structures describing the entities involved in the action:

NS-CS Rule 3: A-Conjunction

Narrative Structure		Conceptual Structure
$[_{\text{Phase } Xa}\ X_1 - X_2 - \ldots X_n]$	$\Leftrightarrow_{\text{default}}$	$[_{\text{EVENT}a}\ \{\text{EVENT}_1, \text{EVENT}_2, \ldots \text{EVENT}_n\}]$
$[_{\text{Phase } Xa}\ X_1 - X_2 - \ldots X_n]$	$\Leftrightarrow_{\text{default}}$	$[_{\text{PATH}a}\ \text{VIA}([_{\text{PLACE}}\{\text{PART}_1, \text{PART}_2, \ldots \text{PART}_n\}])]$

Again, this correspondence uses the general conjunction schema, but interfaces the panels to various aspects of events (with the notation ''a'' mapping to the phase node). Semantically, these subordinate events are often the subparts of a process—unbounded actions with no internal discrete parts (Jackendoff, 1991, 2007; Pustejovsky, 1991)—though this may vary. For example, in Fig. 10, the event of ''conjuring fire'' maps each shape onto a different Initial panel, though the full event maps to the upstairs Initial phase. That is, the Initial phase is about conjuring fire, and is manifested iteratively in different panels. While Fig. 10 depicts this relationship by mapping variations of Properties to the conjunction phase, this is notational shorthand; in actuality, each iteration repeats this whole event structure, as in the correspondence rule.

Again, diagnostic tests can support this as conjunction. Fig. 11a shows that rearranging multiple panels within the conjunction phase has little effect on their understanding (unless we knew, for example, that the order mattered as some

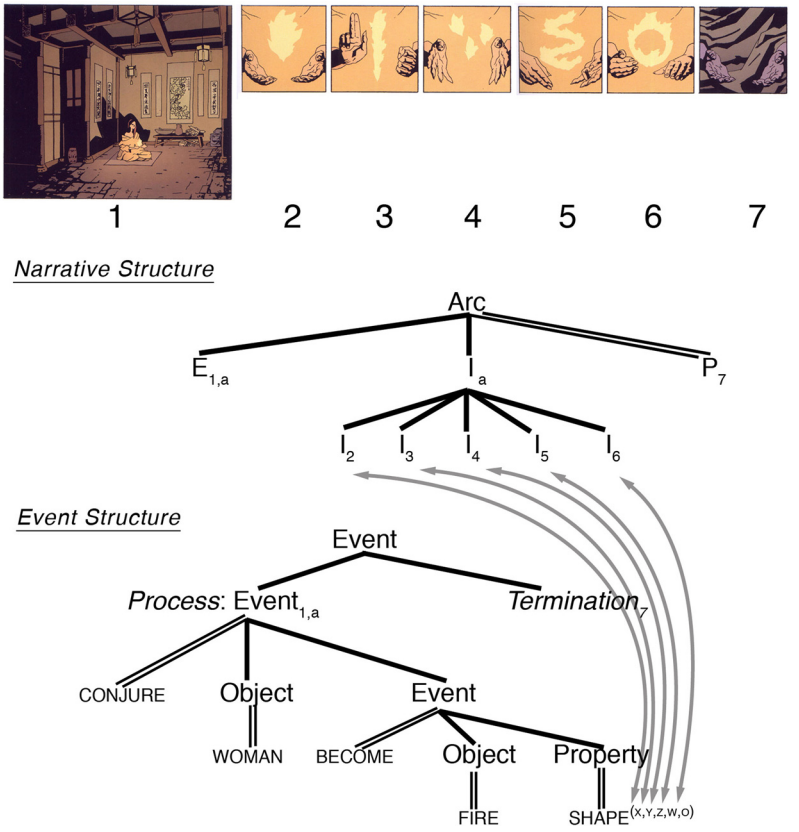

Fig. 10. Action-Conjunction in the parallel architecture for iterative event structures (Mignola et al., 2003).

stepwise sequence). However, in Fig. 11b, moving a single frame out of the conjunction phase changes the meaning significantly—she now conjures fire, stops, and then resumes conjuring fire. A deletion test in Fig. 11c shows that omission of multiple panels from the conjunction phase has little effect on the sequence.

Not all A-Conjunction shows elaboration or variation. Some may just show the repetition of a single event, as in Fig. 12a, where a juggler tosses pins continuously until one hits him in the head, leading him to throw it away angrily. Unlike Fig. 10, this sequence does not depict variation of an action, but repeats the same action until the motion is interrupted. However, both are semantic processes that progress until a termination. Also, the first panel here uses ‘‘polymorphic morphology’’ (Cohn, 2007, 2013a), where the component parts of an image are reduplicated to depict several actions all within one image (Kennedy, 1982). The juggler’s arms repeat in various positions to show multiple throws—not multiple arms! This first panel shows the full scope of the actions, which then becomes individuated in the other Initial panels. Thus, here repeating panels do not just show iterated variation, but depict a single event several times. Indeed, because the first panel uses this polymorphic morphology to repeat the action over and over, this single panel could substitute for all the other panels in this constituent.

Finally, A-Conjunction can also show the component parts of a single action. In Fig. 12b, a character falls from the sky to make a hard crash-landing. The center panels depict the trajectory of his fall across three separate Prolongation panels. The NS-CS correspondence rule notates this trajectory using the function ‘‘VIA’’ for a medial path segment (Jackendoff, 1996, 2010), though it would also map to part of an image schematic path in spatial structure (Jackendoff, 2010; Talmy, 2000). This A-Conjunction shows the component parts of this path, not iterated or repeated versions of the action. Again, any of these panels could be omitted (deletion test) and all of these could be replaced by a single Prolongation (substitution test) depicting the entire fall from the sky in one image. However, a movement test might not work here, because the semantics of this path mandates a particular order. The stepwise order of the juggling in Fig. 12a has a similar constraint, but would not if their temporal order was less distinct. Thus, because A-Conjunction operates over event information, the semantics of the structure may constrain its order, though this may not occur in all cases (imagine the character in Fig. 12b bouncing off of several buildings to show the manner of the path—each bounce panel could freely be rearranged). Such constraints arise from the semantics of the particular elements being conjoined (Cohn and Paczynski, 2013).

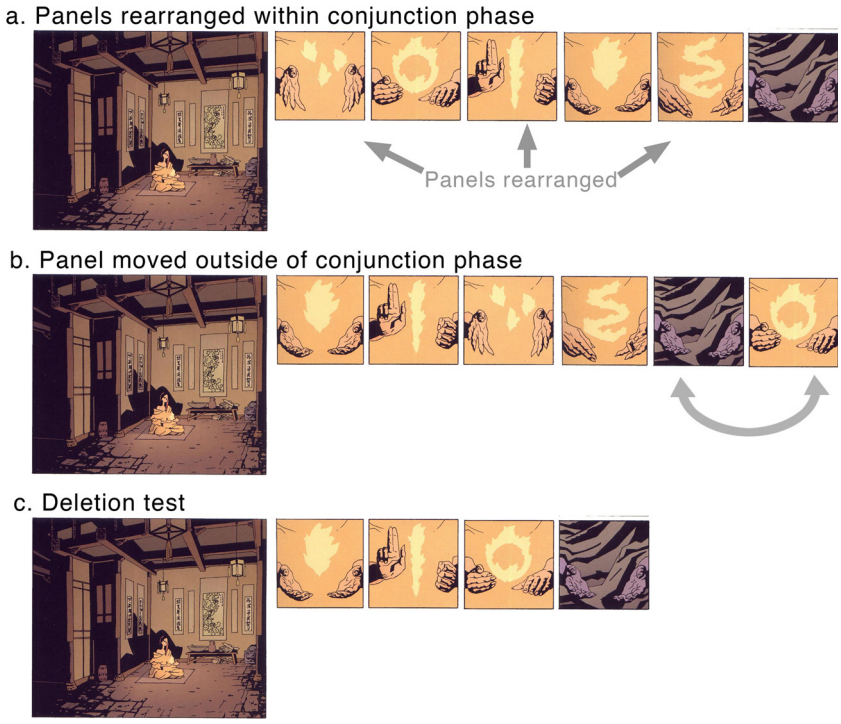

Fig. 11. Diagnostic tests for A-Conjunction applied to the sequence in Fig. 10.

In all of these examples, panels with a common narrative role conjoin to depict multiple aspects of events, be it iteration, repetition, or componential parts. All of these varying semantic traits are types of A-Conjunction. The part-whole relationships expressed in E- or N-Conjunction would require the inference of the superordinate conceptualization (scenes, individuals), but A-Conjunction does not require this same type of inference (except perhaps with paths). Rather, because A-Conjunction involves event information, it mostly uses the same types of bridging inferences that must operate

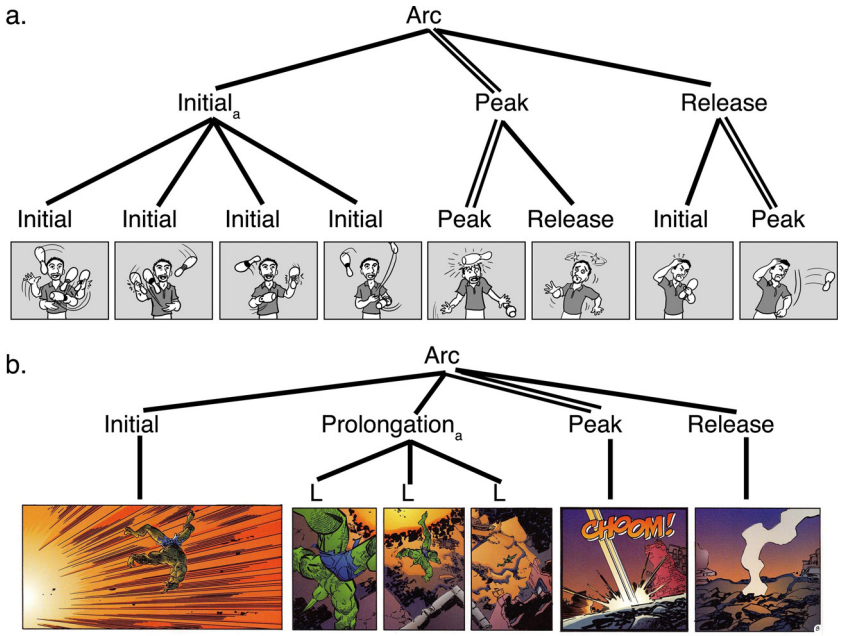

Fig. 12. A-Conjunction (subscript ‘‘a’’) through (a) repetition of a single action and (b) depicting the component parts of a single action/trajectory (Larsen, 2000:8).

between and across ''normal'' panel relationships. That is, conjunction itself does not motivate a special type of inference. Rather, different *semantics* operate on similar *structural* relationships for various conjunctions.

3.5. Semantic network-Conjunction

Conjunction also may map to disparate semantic information, drawing together related or unrelated panels to form a larger meaning. For example, Saraceni (2000, 2001) observed that a sequence of panels may share broader aspects of a semantic network, though they do not convey an explicit narrative. Similar observations have been made for film sequences (e.g., Branigan, 1992; Metz, 1974). Consider Fig. 13, which shows Schroeder in a ''training montage'' as if he were an athlete, only here preparing to play the piano. His actual playing occurs in the final panel (Peak), set up first by the penultimate Initial panel. Together, these panels form a Peak phase, which is set up by an Initial of all the prior panels. While some of the panels in this Initial constituent have logical connections, as a whole they relate only through a semantic field expressing the concept of ''exercise/training.'' There is no overt *narrative* order to these panels. In a sense, these individual panels inherit the narrative role of their superordinate phase.

The panels in this conjunction phase have no discernable connections to a scene, an individual, or actions, but rather are disconnected glimpses of a broader idea. Here, the semantic features of the individual panels link together in a broader semantic network, which otherwise has no inherent spatial or causal connections:

NS-CS Rule 4: S-Conjunction

Narrative Structure		Conceptual Structure
$[_{\text{Phase Xs}}\ X_1 - X_2 - \ldots X_n]$	$\Leftrightarrow_{\text{default}}$	$[_{\text{CATEGORYs}}\ \{W_1, Y_2, \ldots Z_n\}]$

This correspondence rule thus captures dispersed aspects of meaning that may not have a specified internal structure, though may be connected through semantic associations. In this case, the superordinate category maps to the phase node, notated with ''s.'' Each panel may not signal an explicit role through their bottom-up content, but together they convey the broader meaning of the sequence and may inherit those narrative roles through the top-down sequence context.

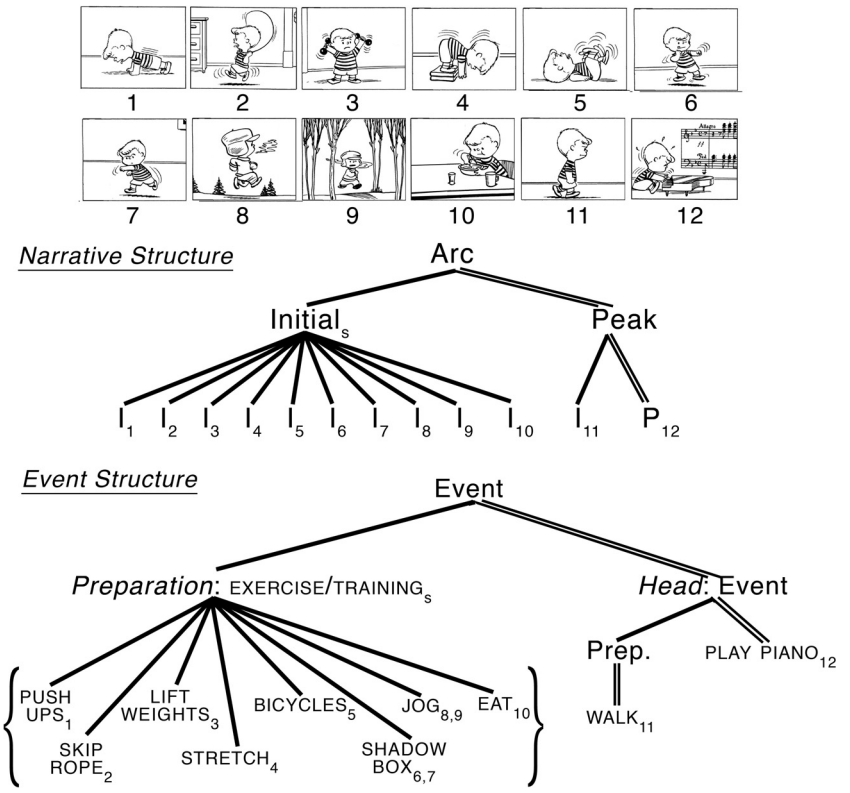

Fig. 13. S-Conjunction showing conjoined panels sharing a common semantic network (Schulz, 2004 [1953]).

Diagnostic tests can again illustrate this. Fig. 14a applies the movement test by rearranging multiple panels within the conjunction phase, resulting in little change in comprehensibility. This holds despite S-Conjunction using very different semantic information than the previously discussed types. In contrast, the meaning changes by moving a panel from the conjunction phase into the subsequent constituent, as in Fig. 14b. Here, the sequence now seems as though Schroeder is walking in order to eat. The final panel of him playing piano then appears isolated from the exercise and the eating. In Fig. 14c, the deletion test omits half of the sequence, while retaining a coherent ''training montage,'' because all of the subordinate panels in the conjunction phase retain the same status.

Finally, Fig. 14d replaces the conjunction phase for a single panel incorporating all of these actions. Here, the ''training montage sequence'' becomes a ''montage image'' with layered information creating a collage of actions and events. Montage images like this are different from the aforementioned ''polymorphic'' morphology because they do not explicitly repeat characters to show a dedicated action. Rather, montage panels visually blend disparate information, whether or not it repeats characters or other objects (depending on the component parts of the montage). In this case, the montage panel conveys the same information in a single panel as the entire previous sequence, providing evidence that this is indeed a conjunction phase.

S-Conjunction appears to involve a somewhat different inference than the previously discussed types. Global coherence across these panels uses a common semantic network or semantic associative field (Brown and Yule, 1983; Halliday and Hasan, 1985; van Dijk, 1977). Yet, in some sense, this remains a highly abstract version of the part-constructing-whole inference found in E-Conjunction and N-Conjunction, though not for a spatial/referential relationship between elements (entities, scenes). Rather, here the relationship is between an abstract superordinate category (whole) and various semantically associated subordinate elements (parts) that may comprise that category (e.g., Rosch et al., 1976).

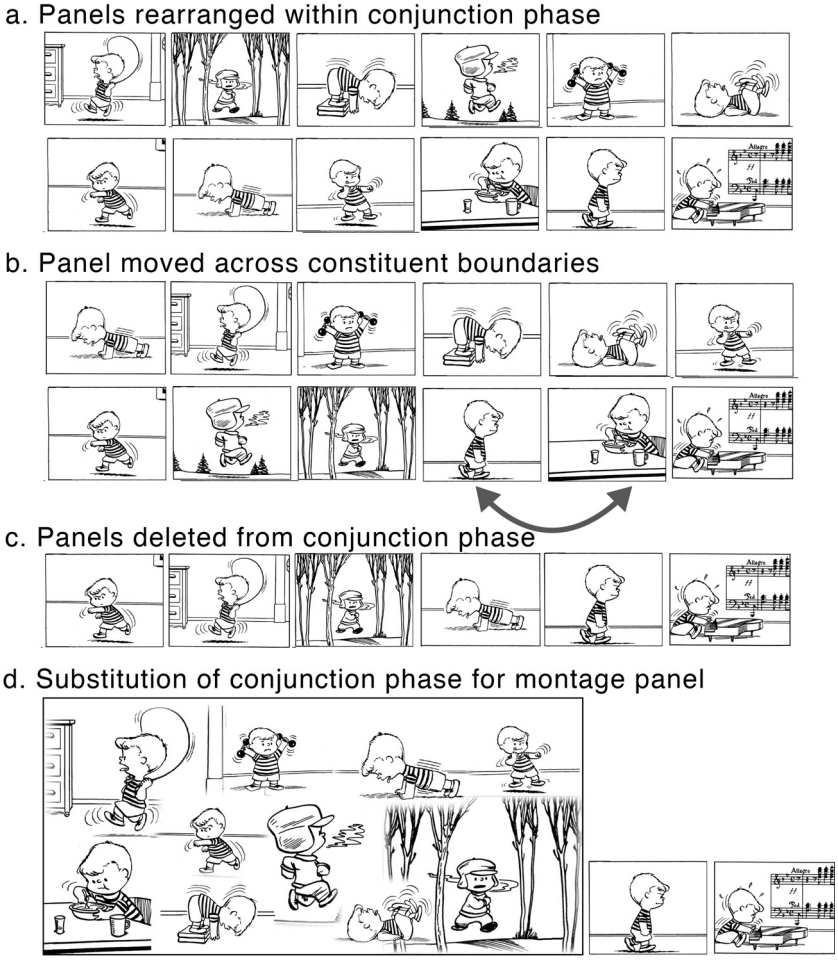

Fig. 14. Diagnostic tests applied to the S-Conjunction sequence in Fig. 13.

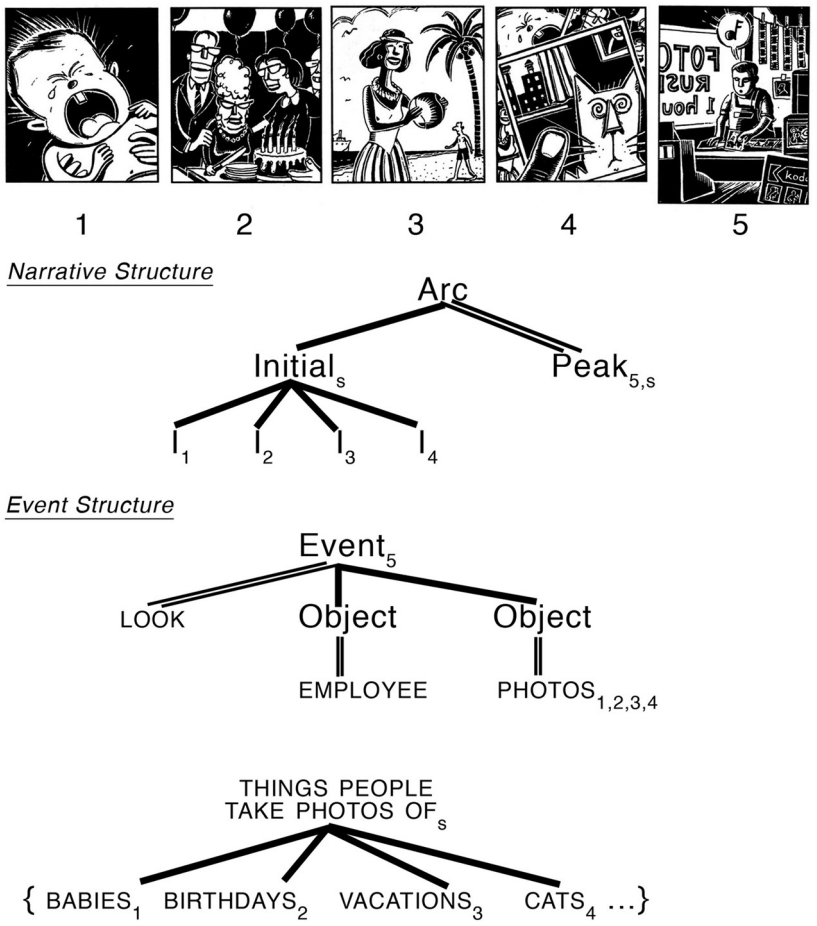

Fig. 15. A visual sequence using S-Conjunction depicting a list bound by the superordinate category ''things people take photos of'' (Kuper, 1996).

In Fig. 13, the panels all retain a broader semantic field, though S-Conjunction can also use a different relationship between categorical parts and wholes. Consider Fig. 15. The first four panels of this sequence appear to have no common theme, and are only united by the final panel, which reveals them as photos being looked at by a person in a photography store. These fronted panels could go in any order, and contain no overt markers regarding any sort of narrative roles. Indeed, only the final panel has positional relevance, and semantically it serves as a visual title for the ''graphic list'' provided in the prior panels: ''Things people take photos of.'' Thus, this panel reveals the superordinate category (Saraceni, 2003), comprised by the top-level Initial phase.

Without the final panel to serve as the top-down ''list title,'' these images would seem fairly disconnected and incongruous (Saraceni, 2000, 2001), and indeed locally the panels maintain only ''non-sequitur'' linear relations (McCloud, 1993). This is different than the training montage in Fig. 13, where the semantically associated bottom-up content within the morphology of each image suggests a superordinate category (i.e., exercise). In Fig. 15, the semantic network emerges once the ''title'' panel is revealed, and thus the component parts inherit the category of the ''title.'' Without this panel, the sequence would remain disjointed, similar to psycholinguistic experimental results where the absence of an overtly stated topic yields an incongruous discourse (Bransford and Johnson, 1972; Dooling and Lachman, 1971; St. George et al., 1994). In this case, the ''title'' is the final panel, a Peak, which creates a structural effect of ''reanalyzing'' the preceding sequence in light of the revealing categorical insight. However, this ''title'' panel could play a different role at the beginning to ground the sequence in its semantic field first, followed by those detailed listed items (but it could not go in the middle). Such a sequence would convey the same *semantic* information, but would do so with a different narrative *structure* for how it reveals that information to a reader.

It is worth mentioning another structural analogy to this type of construction in language (beyond actual lists). Noun-noun compounds sometimes feature a holistic meaning, though their constituent parts have little relation to each other.

For example, *Lou Gehrig's Disease* uses a connection of *Lou Gehrig* to *Disease*, but in the absence of knowledge of baseball history (Gehrig had ALS, the disease that would later take his name), this connection makes no sense (Jackendoff, 2009, 2010). Nevertheless, the compound as a whole has a coherent meaning. In an analogous way, the panels of the Initial ''list'' phase inherit their narrative roles from their phase (itself identified top-down from its relation to a Peak), relying wholly on membership to an *ad hoc* superordinate conceptual structure for their role in the narrative.

3.6. Multiple Conjunctions

We have now posited four schematized NS-CS interfaces involving conjunction. Because the conjunction schema in narrative stayed the same, these types differed depending on their interfaces to semantic information. Thus, mappings to other semantic information may also be possible beyond those discussed here. Nevertheless, these examples all used only a single conjunction constituent. However, the basic schema for conjunction is recursive, and should thereby allow additional embedding. This structure is useful when, for example, various NS-CS interfaces appear in the same surface sequence of panels. Because this model uses a parallel architecture, we could need only a single flat phase structure with several types of NS-CS interfaces. There are several reasons why this might not be theoretically preferable though, as can best be seen in examples.

Consider Fig. 16a, where a man and boy walk up to a family, who then bows to them. In this sequence, each group constitutes its own bound entity (e.g., Jackendoff, 2010; Talmy, 2000). The first two panels show only glimpses of the man/boy (feet, busts), using N-Conjunction to construct their whole conceptualization. Panels 3 and 4 show the family, first their whole bodies, then zoomed in on their faces (a Refiner). These pairs of panels then unite with E-Conjunction, where the man/boy and the family are recognized as belonging to the same location. These four panels are all Initials that set up the Peak, where the family gratefully bows to the man and boy for saving them in an earlier scene.

Diagnostic tests can show that this sequence embeds one structure inside another. First, panels 1 and 2 can be substituted for a single Mono (Fig. 16b), confirming this as N-Conjunction, while retaining the E-Conjunction with the subsequent panels. Second, we can test that E-Conjunction spans panels 1 through 4 by subsuming them into a single Macro that depicts all the characters (Fig. 16c). A deletion test reinforces this; by omitting panels 2 and 4 (Fig. 16d), this constituent retains an E-Conjunction with no additional embedding (N-Conjunction) or modification (Refiner). Finally, Fig. 16e shows that each subordinate constituent can be moved while retaining the same overall structural and semantic relationships (movement test). Altogether, these diagnostics suggest that this sequence embeds one type of conjunction within another, rather than all panels belonging to a single ''flat structure'' constituent.

Consider also Fig. 17, which depicts a man fixing a tire (a process) until he gives up (its termination). The sequence alternates between the mechanic and a highly schematized illustration of his actions with the tire. Each pair of panels unites the man to the tire using E-Conjunction, since they are never shown together in a single panel. Meanwhile, each conjoined pair forms a phase progressing in the step-by-step actions united by A-Conjunction within a superordinate phase. Thus, the broader constituent uses A-Conjunction to combine phases that use E-Conjunction.

In order to maintain an isomorphism between narrative structure and meaning, certain constraints operate on how conjunction phases embed within each other. Because whole entities belong within a broader scene, the construction of entities using N-Conjunction should prototypically be subordinate to E-Conjunction. Similarly, because actions are situated within an environment, then E-Conjunction should be subordinate to A-Conjunction. The overall preferences for *semantic* embedding descends as (Cohn, 2003, 2010): *Different times > Same time and different space/character (spatial environment) > Same time and space/character.* The prototypical embedding for types of conjunction is therefore isomorphic to this semantic constraint: *A-Conjunction > E-Conjunction > N-Conjunction*. Essentially, because entities belong within environments, and environments progress throughout actions, the interfaces of these meanings to conjunction follows this same embedding in the narrative grammar.

Studies from various fields suggest similar hierarchic relations between time/events, spatial locations, and characters/objects. Narratives on the whole typically progress from broader scenes to characters and their events (Cohn and Paczynski, 2013; Mandler and Johnson, 1977; Primus, 1993), i.e., scenes and characters are introduced first, then move through the larger structures of different event states (Mandler and Johnson, 1977). This organization is consistent with notions that discourse understanding involves a spatio-temporal situation model, whereby spatial regions containing focal entities move through changing temporal states (Zwaan, 2004). The segmentation of film episodes supports this, as participants are more sensitive to changes in action/time than those of spatial locations (Magliano et al., 2001; Magliano and Zacks, 2011), suggesting that action/time changes operate as a coarser structure than spatial locations (Magliano and Zacks, 2011). From a different field, research on scene perception of single images has suggested that observers recognize the gist of a scene's global-level location prior to the local-level objects and characters, which precede activation of event knowledge (e.g., Oliva, 2005; Oliva and Torralba, 2006).

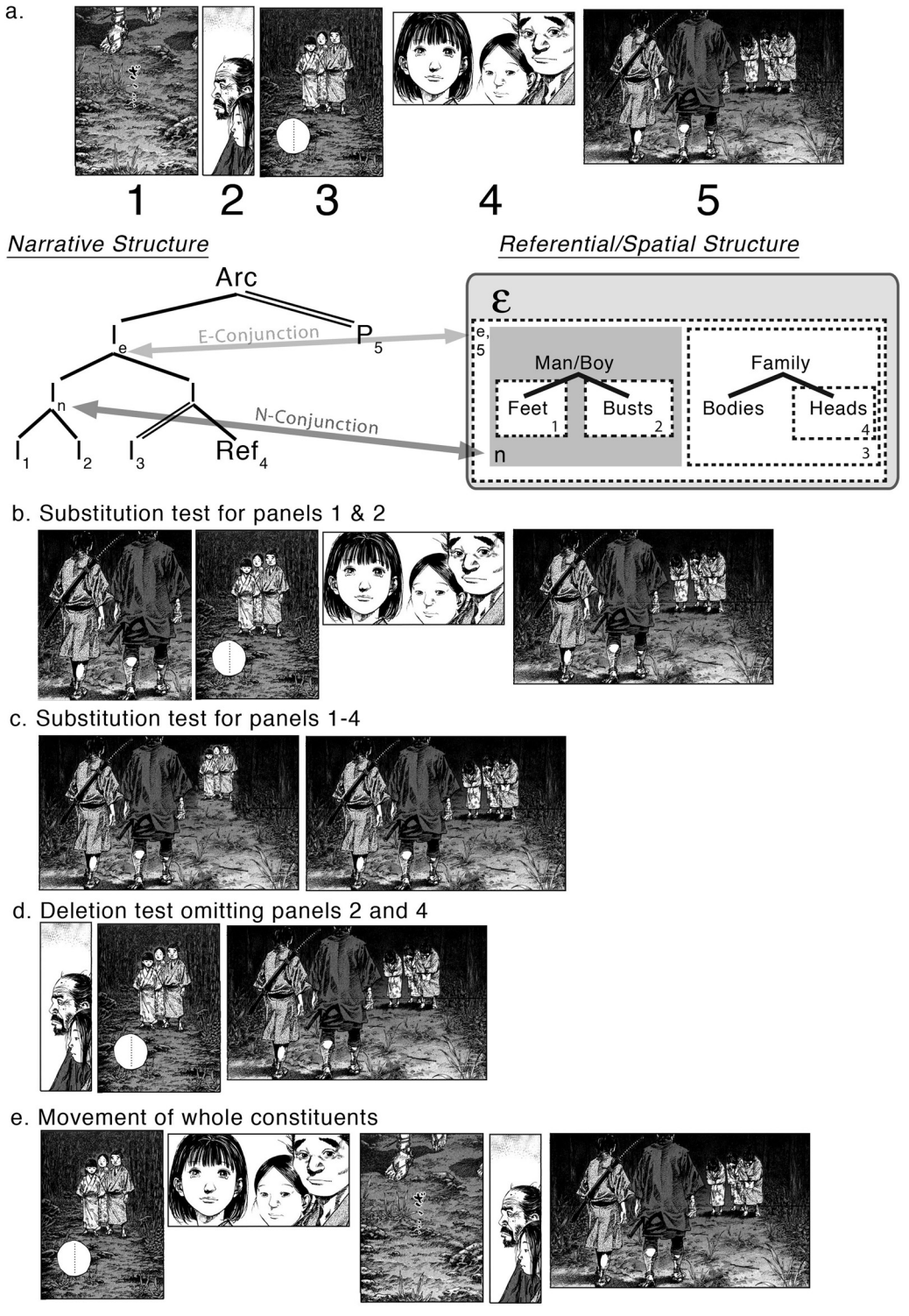

Fig. 16. (a) Sequence using both N-Conjunction and E-Conjunction (Takehiko, 2002) along with diagnostic tests examining its structure (b–e).

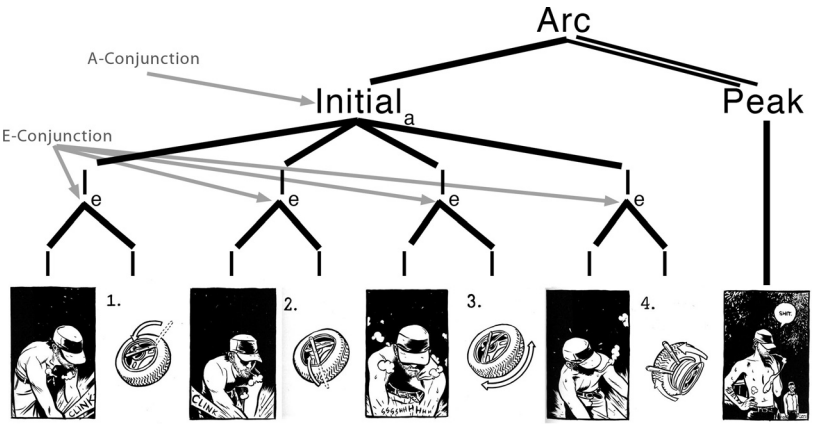

Fig. 17. Sequence using both E-Conjunction and A-Conjunction (Johnson, 2005:58).

These examples of multiple conjunctions serve several purposes. First, they show that, like other narrative constituents, conjunction is recursive. Second, they show that different correspondences to meaning can occur at different levels of conjunction. This recursive and multifaceted nature to conjunction enables the narrative grammar to produce even more interesting and complex sequences than the canonical narrative arc and conjunction alone. Third, these examples suggest constraints to the NS-CS relationships in embedded conjunctions.

4. Conjunction and attention structure

We can now consider how these connections between narrative and semantics link to the basic representations of information in a scene. As mentioned, panels can frame information in varying ways: Macros depict multiple entities, Monos depict individual entities, and Micros depict less than a single entity. Previously, these attentional categories were described as reflecting the interface between graphic structure and conceptual structure. Yet, different types of NS-CS interfaces using conjunction construct ''virtual'' versions of these attentional categories.

For instance, E-Conjunction often depicts the component parts of a scene (often Monos) without depicting the full scene with multiple interacting entities (a Macro). Rather, the notion of a full scene is constructed in spatial structure alone—a virtual Macro. Similarly, N-Conjunction uses various panels with less than a single entity (usually Micros), and constructs the notion of that entity—a virtual Mono. A-Conjunction unites several iterations or repetitions of an event or action—just as all that information can be conveyed in a single panel using polymorphic morphology (as in Fig. 12a). Finally, S-Conjunction depicts disparate information bound through only a common semantic network or superordinate category, which could be conveyed in a single montage panel that blends or layers these elements together in a collage.

Thus, as in Fig. 18, all of these interfaces provide options for framing information—either through individual panels or across sequences of panels: there are numerous ways to show the same semantic information. For example, if the creator of a visual narrative wanted to convey a whole scene, they have a choice: ''*Do I want to show the scene as a whole in a Macro? Do I want to highlight portions of a scene, and leave my reader to infer the scene as a 'virtual Macro'?*'' Both options convey similar conceptual information, though they achieve it by highlighting (or muting) aspects of that meaning through the framing in or across panels.

The distinctions laid out in Fig. 18 also provide a rubric for the expectations of substitution tests. Each attentional category should substitute for semantically corresponding conjunctions (i.e., a Macro for conjoined Monos, a Mono for conjoined Micros). However, panels of different levels should not apply across types of conjunctions: A Macro should not be able to substitute for N-Conjunction, and a Mono should not be able to substitute for E-Conjunction.

Note that these distinctions rely on clarity of the ''morphological'' information in panels' contents, and ambiguity could potentially lead to different construals. For example, imagine the boxing example with Initials showing a close up of a hand and another of eyes. If these body parts could be discriminated as different entities (hand belongs to one entity, eyes to another) it would use E-Conjunction. Such close ups may be harder to substitute with a Macro depending on their spatial relations in a scene, but they would still imply entities within an environment (E-Conjunction), not parts of entities as a single character (N-Conjunction). However, if these body parts could not be discriminated as different characters (hand of ambiguous referential entity, eyes of ambiguous referential entity), they may involve N-Conjunction to construct a single entity. Similar inference was reported by Kuleshov's (1974) ''experiments'' with film editing where he showed shots of different women's body parts, and people interpreted them as belonging to the same woman (whether these sequences used N-Conjunction or just general part-constructing-whole inference is unknown). If the sequence later disambiguates

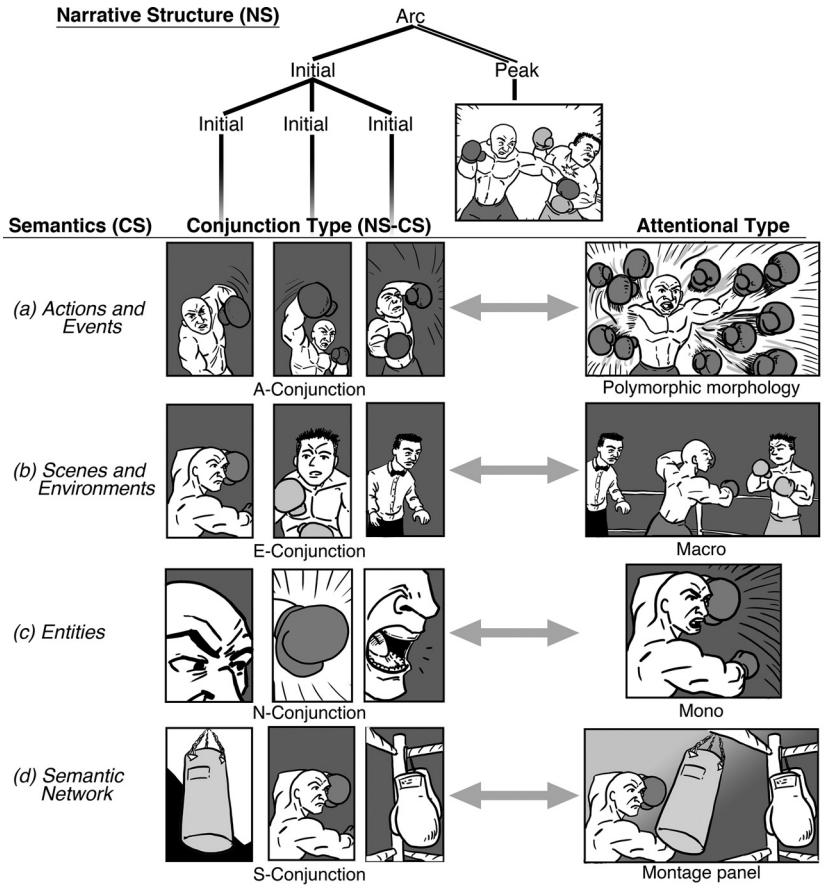

Fig. 18. Various types of meaning expressed either through an NS-CS interface (such as in Conjunction) or through a single panel using a particular attentional category.

that these body parts actually come from different entities, we would expect some form of reanalysis in processing the altered mental model. Thus, because of this parallel architecture, panels' content may allow for ambiguities in semantics without altering the narrative structure (i.e., in both cases they act as conjoined Initials). This would therefore be the inverse of Fig. 8 where ambiguity results in different narrative structures—here there is one narrative structure with different semantic interpretations.

4.1. Costs and benefits

As it is framed above, using conjunction or a single image provides different options for a creator of visual narratives to convey information. Why might one strategy be used over another then? What are the functional benefits or costs for using conjunction for both creators and comprehenders?

First, conjunction allows for panels to focus on individual characters, rather than providing characters embedded within a larger visual scene. By highlighting the component parts of a scene, panels function as an ''attention unit'' that put a ''window'' onto relevant information (Cohn, 2007; Cohn et al., 2012b). This allows an author to focus comprehenders directly to relevant information, rather than rely on comprehenders to reliably extract that information from a larger image. Consider Fig. 19a: all the characters appear in large Macros, which require a reader to discern what might be relevant within and across panels. By breaking up those scenes, Fig. 19b can directly depict the relevant information, thereby reducing the ''noise'' contributed by less important information. The effect of this increased focus may thereby modulate the strength at which those highlighted elements may be represented in the mental model of the scene—elements framed directly will be more salient than those falling outside this ''window of attention'' (Graesser, 1981; Langacker, 2001; Magliano et al., 2005; Zwaan, 2004).

A second function that arises from E-Conjunction is variation in the pacing of the narrative. Because the narrative grammar packages meaningful information in a coherent way, it modulates how this meaning is presented to a reader.

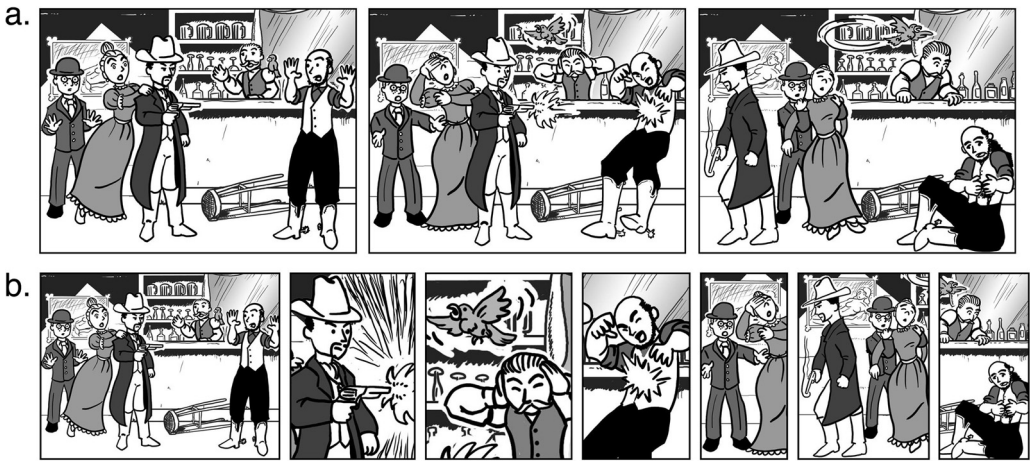

Fig. 19. Two narrative sequences with different attentional framing: (a) macro representation of a sequence and (b) sequence broken up into panels with less information.

Breaking apart panels results in an alternative pacing. For example, if Establishers and/or Initials use E-Conjunction, the additional units and bouncing between characters may add narrative tension before the Peak. This tension may be lost if simply depicting the scene in Macros. Some research has suggested that breaking up panels in this way may also vary the expectations about the subsequent sequence. Kaiser and Li (2013) found that the division of characters into individual panels using E-Conjunction elicits more predictions about subsequent events than single Macros. Thus, altering the narrative pacing may in turn affect the panel-by-panel comprehension of a sequence.

Finally, in line with these aforementioned functions, conjunction may also clarify sequences that would otherwise be more difficult to understand. If several Macro panels contain multiple entities, tracking all those characters across panels may be demanding. Fig. 19a does not use conjunction, and all six characters (including the bird) change across all three panels. Keeping track of all six entities across panels may be burdensome on attention and working memory. Compare this to Fig. 19b, where the individuation of entities should make it easier to track them across panels, highlighting the focal parts of the sequence, muting the others, and heightening narrative tension through increased narrative units. This difference may also be a tradeoff: As Macros are relied on less, more Monos are used. Some evidence suggests such a tradeoff has occurred over time in film narratives. The scale of framing of a film shot correlates with the number of characters per shot (Cutting, 2015), and increasingly smaller scales (and thus less characters) have been used over the past 70 years of film editing (Cutting, 2015), concurrent with a reduction in shot duration (Cutting et al., 2010).

Across all of these functional purposes, the narrative strategies in Fig. 19a and b should require cognitive resources to be allocated in different ways. An ''overload'' of characters in each frame and filtering out non-relevant ''noise'' likely makes demands on working memory and attention. Conjunction then allows for a reduction of such costs by providing comprehenders a ''manageable'' amount of characters to track across panels (while ''manageable'' may optimally be ''one'' per panel, this limit could be determined empirically). However, now, a comprehender must negotiate additional structure in the narrative grammar, and must inferentially construct a spatial structure that would otherwise be provided for them in full. Thus, conjunction also must bear the costs of building a narrative constituent and the inference of a broader scene. Both strategies are therefore hypothesized to incur costs and benefits to a comprehender, depending on how those resources might be allocated.

4.2. Cultural variation

The discussion above has framed the use or non-use of conjunction as a ''choice'' in authorship of visual narratives. However, because VNG argues that these elements are stored in memory as a schema, and not simply unfolding spontaneously in comprehension (Bateman and Wildfeuer, 2014a, 2014b; Magliano and Zacks, 2011), it opens the possibility that grammars of different visual languages use these schemas in varying ways. Thus, we might ask, how pervasive might conjunction be in visual narratives, and does it vary between different cultures' systems? Corpus analyses have shown that panels in Japanese manga use substantially higher proportions of Monos and Micros than those in American comics, which use equal if not greater numbers of Macros than Monos (Cohn, 2011; Cohn et al., 2012b). Because Monos only depict a portion of a scene, as opposed to the whole scene in Macros, the higher proportion of Monos in manga led to a preliminary interpretation that this narrative grammar used more E-Conjunction than the system used in American comics (Cohn, 2013a).

In addition, research on film editing may suggest a degree of learning for comprehending what is here called E-Conjunction. Individuals from a remote village in Turkey had difficulty inferring that film shots of individual characters were meant as simultaneously belonging to a common environment (i.e., shot/reverse shots), a deficit attributed to their inexperience with watching films (Ildirar and Schwan, 2015; Schwan and Ildirar, 2010). Beyond the general findings that basic aspects of comprehending and creating visual narratives are modulated by age and experience with comics (e.g., Cohn et al., 2012a; Nakazawa, 2005; Wilson and Wilson, 1987), these findings suggest that the comprehension of E-Conjunction—a specific part of narrative grammar—may require proficiency in the grammar of a visual language.

Altogether, the existing literature suggests that conjunction may involve a degree of fluency to understand, and that cultures' visual narratives may differ in its usage. Given this, we might hypothesize that individuals who read/view visual narratives that use more E-Conjunction (like Japanese manga) would differ in their comprehension of it versus those who read/view systems that use it less (like American comics), rather than it being an ''all or nothing'' feature of visual narrative fluency. Empirical experimentation on such a hypothesis would directly address the degree to which conjunction is stored in memory.

5. Conclusion

This paper has posited that ''conjunction'' occurs in visual narratives, whereby narrative constituents repeat several panels of the same category. This simple narrative structure allows for significant complexity via how these panels map to a conceptual structure, and the potential construction of meaning beyond the represented panels. Conjunction may map to a variety of meanings—part-whole inferences, repeated or iterated events, a broader semantic network, and potentially other meanings not explored herein. These semantic characteristics are determined by the content of the units used within a conjunction phase. This emphasis on the *interface* between structure and meaning helps explain how several images can play the same functional role in a sequence while conveying different semantic information. In Fig. 18, all of the conjunctions and attentional categories function as Initials, but the meaning changes based on the interface to semantics, and indeed further combination across these levels could yield additional complexity (such as combining Macros and Monos to create conjunction and Refiners within the same constituent). This interface between structures also allows for semantic ambiguity to arise from different semantic construal of the same narrative structure.

Finally, overall this model suggests that narrative structure uses three primary abstract patterns. First, a canonical narrative arc specifies the order of categories, based on their relative positions to a ''head'' Peak. Second, a Refiner schema specifies how categories can expand using spatial modifiers. Third, a conjunction schema allows narrative categories to repeat multiple times, mapping to various semantic meanings. These patterns, stored in long-term memory as ''constructions'' (Goldberg, 1995; Jackendoff, 2002), allow for significant complexity in the structure of sequential images. In addition, these constructions are consistent with the basic abstract schemas in the syntax of language—a head-modifier schema (x-bar) and a conjunction schema (Culicover and Jackendoff, 2005)—suggesting an underlying architecture of ''grammar'' that is similar across domains, yet differs in its domain-specific manifestations. Initial evidence for this domain-generality has been demonstrated by findings that similar neural responses appear to gross violations of ''grammar'' across the domains of language, music, and visual narrative (Cohn et al., 2014; Koelsch et al., 2005; Patel, 2003; Patel et al., 1998). However, explicit theorizing of the characteristics of such grammatical architecture, such as that provided here, can allow for more targeted experimentation on the specifics of these parallels between ''grammar'' across domains.

Acknowledgments

This research was supported by an NIH funded (T32) postdoctoral training grant for the Institute of Neural Computation at UC San Diego. Dark Horse Comics, Drawn and Quarterly, and Fantagraphics Books are thanked for their contributions to the Visual Language Research Library and the examples in this paper's analyses.

References

Asher, Nicholas, Lascarides, Alex, 2003. Logics of Conversation. Cambridge University Press, Cambridge.
Bateman, John A., 2007. Towards a grande paradigmatique of film: Christian Metz reloaded. Semiotica 2007, 13–64.
Bateman, John A., Schmidt, Karl-Heinrich, 2012. Multimodal Film Analysis: How Films Mean. Routledge, New York.
Bateman, John A., Wildfeuer, Janina, 2014a. Defining units of analysis for the systematic analysis of comics: a discourse-based approach. Stud. Comics 5, 373–403.
Bateman, John A., Wildfeuer, Janina, 2014b. A multimodal discourse theory of visual narrative. J. Pragmat. 74, 180–208.
Bordwell, David, Thompson, Kristin, 1997. Film Art: An Introduction, 5th ed. McGraw-Hill, New York.
Branigan, Edward, 1992. Narrative Comprehension and Film. Routledge, London, UK.

Bransford, John D., Johnson, Marcia K., 1972. Contextual prerequisites for understanding: some investigations of comprehension and recall. J. Verbal Learn. Verbal Behav. 11, 717–726.
Brown, G., Yule, G., 1983. Discourse Analysis. Cambridge University Press, Cambridge.
Buckland, Warren, 2000. The Cognitive Semiotics of Film. Cambridge University Press, Cambridge.
Carroll, John M., 1980. Toward a Structural Psychology of Cinema. Mouton, The Hague.
Cheng, Lisa Lai-Shen, Corver, Norbert, 2013. Diagnosing Syntax. Oxford University Press.
Chomsky, Noam, 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, MA.
Cohn, Neil, 2003. Early Writings on Visual Language. Emaki Productions, Carlsbad, CA.
Cohn, Neil, 2007. A visual lexicon. Public J. Semiot. 1, 53–84.
Cohn, Neil, 2010. The limits of time and transitions: challenges to theories of sequential image comprehension. Stud. Comics 1, 127–147.
Cohn, Neil, 2011. A different kind of cultural frame: an analysis of panels in American comics and Japanese manga. Image Narrat. 12, 120–134.
Cohn, Neil, 2013a. The visual language of comics: introduction to the structure and cognition of sequential images. Bloomsbury, London, UK.
Cohn, Neil, 2013b. Visual narrative structure. Cogn. Sci. 37, 413–452.
Cohn, Neil, 2014a. The architecture of visual narrative comprehension: the interaction of narrative structure and page layout in understanding comics. Front. Psychol. 5, 1–9.
Cohn, Neil, 2014b. You're a good structure, Charlie Brown: the distribution of narrative categories in comic strips. Cogn. Sci. 38, 1317–1359.
Cohn, Neil, Paczynski, Martin, 2013. Prediction, events, and the advantage of Agents: the processing of semantic roles in visual narrative. Cogn. Psychol. 67, 73–97.
Cohn, Neil, Paczynski, Martin, Jackendoff, Ray, Holcomb, Phillip J., Kuperberg, Gina R., 2012a. (Pea)nuts and bolts of visual narrative: structure and meaning in sequential image comprehension. Cogn. Psychol. 65, 1–38.
Cohn, Neil, Taylor-Weiner, Amaro, Grossman, Suzanne, 2012b. Framing attention in Japanese and American Comics: cross-cultural differences in attentional structure. Front. Psychol. Cult. Psychol. 3, 1–12.
Cohn, Neil, Jackendoff, Ray, Holcomb, Phillip J., Kuperberg, Gina R., 2014. The grammar of visual narrative: neural evidence for constituent structure in sequential image comprehension. Neuropsychologia 64, 63–70.
Colin, Michel, 1995a. Film semiology as a cognitive science. In: Buckland, Warren (Ed.), The Film Spectator: From Sign to Mind. Amsterdam University Press, Amsterdam, pp. 87–112.
Colin, Michel, 1995b. The grande syntagmatique revisited. In: Buckland, Warren (Ed.), The Film Spectator: From Sign to Mind. Amsterdam University Press, Amsterdam, pp. 45–86.
Culicover, Peter W., Jackendoff, Ray, 2005. Simpler Syntax. Oxford University Press, Oxford.
Cutting, James E., 2015. The framing of characters in popular movies. Art Percept. 3, 191–212.
Cutting, James E., DeLong, Jordan E., Nothelfer, Christine E., 2010. Attention and the evolution of Hollywood film. Psychol. Sci. 21, 432–439.
Dooling, D. James, Lachman, Roy, 1971. Effects of comprehension on retention of prose. J. Exp. Psychol. 88, 216–222.
Eisenstein, Sergei, 1942. Film Sense. Harcourt, Brace World, New York.
Friederici, Angela D., 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6, 78–84.
Goldberg, Adele, 1995. Constructions: A Construction Grammar Approach to Argument Structure. University of Chicago Press, Chicago, IL.
Graesser, Arthur C., 1981. Prose Comprehension Beyond the Word. Springer-Verlag, New York.
Graesser, Arthur C., Singer, Murray, Trabasso, Tom, 1994. Constructing inferences during narrative text comprehension. Psychol. Rev. 101, 371–395.
Hagmann, Carl Erick, Cohn, Neil (under review). The pieces fit: constituent structure and global coherence of visual narrative in RSVP.
Hagoort, Peter, 2003. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. Neuroimage 20, S18–S29.
Halliday, M.A.K., Hasan, Ruqaiya, 1976. Cohesion in English. Longman, London.
Halliday, M.A.K., Hasan, Ruqaiya, 1985. Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective. Deakin University Press, Victoria.
Haviland, Susan E., Clark, Herbert H., 1974. What's new? Acquiring new information as a process in comprehension. J. Verbal Learn. Verbal Behav. 13, 512–521.
Huff, Markus, Schwan, Stephan, 2012. Do not cross the line: heuristic spatial updating in dynamic scenes. Psychon. Bull. Rev. 19, 1065–1072.
Ildirar, Sermin, Schwan, Stephan, 2015. First-time viewers' comprehension of films: bridging shot transitions. Br. J. Psychol. 106, 133–151.
Jackendoff, Ray, 1987. Consciousness and the Computational Mind. MIT Press, Cambridge, MA.
Jackendoff, Ray, 1990. Semantic Structures. MIT Press, Cambridge, MA.
Jackendoff, Ray, 1991. Parts and boundaries. Cognition 41, 9–45.
Jackendoff, Ray, 1996. Semantics and cognition. In: Lappin, Shalom (Ed.), The Handbook of Contemporary Semantic Theory. Blackwell, Oxford, pp. 539–559.
Jackendoff, Ray, 2002. Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press, Oxford.
Jackendoff, Ray, 2007. Language, Consciousness, Culture: Essays on Mental Structure (Jean Nicod Lectures). MIT Press, Cambridge, MA.
Jackendoff, Ray, 2009. Compounding in the parallel architecture and conceptual semantics. In: Lieber, R., Stekauer, P. (Eds.), Oxford Handbook of Compounding. Oxford University, Oxford, pp. 105–128.
Jackendoff, Ray, 2010. Meaning and the Lexicon: The Parallel Architecture 1975–2010. Oxford University Press, Oxford.
Jahn, Manfred, 1997. Frames, preferences, and the reading of third-person narratives: towards a cognitive narratology. Poetics Today 18, 441–468.
Johnson-Laird, P.N., 1983. Mental Models. Harvard University Press, Cambridge, MA.
Kaiser, Elsi, Li, David Cheng-Huan, 2013. Visuospatial grouping influences expectations about upcoming discourse. In: 26th Annual CUNY Conference on Human Sentence Processing, Columbia, SC.
Kennedy, John M., 1982. Metaphor in pictures. Perception 11, 589–605.
Kintsch, Walter, 1998. Comprehension: A Paradigm for Cognition. Cambridge University Press.

Koelsch, Stefan, Gunter, Thomas C., Wittfoth, Matthias, Sammler, Daniel, 2005. Interaction between syntax processing in language and in music: an ERP study. J. Cogn. Neurosci. 17, 1565–1577.
Kuleshov, Lev, 1974. Kuleshov on Film: Writings of Lev Kuleshov. University of California Press, Berkeley.
Langacker, Ronald W., 2001. Discourse in cognitive grammar. Cogn. Linguist. 12, 143–188.
Magliano, Joseph P., Graesser, Arthur C., 1991. A three-pronged method for studying inference generation in literary text. Poetics 20, 193–232.
Magliano, Joseph P., Zacks, Jeffrey M., 2011. The impact of continuity editing in narrative film on event segmentation. Cogn. Sci. 35, 1489–1517.
Magliano, Joseph P., Miller, Jason, Zwaan, Rolf A., 2001. Indexing space and time in film understanding. Appl. Cogn. Psychol. 15, 533–545.
Magliano, Joseph P., Taylor, Holly, Kim, Hyun-Jeong Joyce, 2005. When goals collide: monitoring the goals of multiple characters. Mem. Cognit. 33, 1357–1367.
Mandler, Jean M., Johnson, Nancy S., 1977. Remembrance of things parsed: story structure and recall. Cogn. Psychol. 9, 111–151.
Marr, David, 1982. Vision. Freeman, San Francisco, CA.
Marslen-Wilson, William D., Tyler, Lorraine Komisarjevsky, 1980. The temporal structure of spoken language understanding. Cognition 8, 1–71.
Martin, James R., 1983. Conjunction: the logic of English text. In: Petöfi, J.S., Sözer, E. (Eds.), Micro and Macro Connexity of Discourse (= Papers in Textlinguistics 45). Helmut Buske Verlag, Hamburg, pp. 1–72.
McCloud, Scott, 1993. Understanding Comics: The Invisible Art. Harper Collins, New York.
McKoon, Gail, Ratcliff, Roger, 1992. Inference during reading. Psychol. Rev. 99, 440–466.
McNamara, Danielle S., Magliano, Joe, 2009. Toward a comprehensive model of comprehension. Psychol. Learn. Motiv. 51, 297–384.
Metz, Christian, 1974. Film Language: A Semiotics of the Cinema. Oxford University Press, New York.
Nakazawa, Jun, 2005. Development of manga (comic book) literacy in children. In: Shwalb, David W., Nakazawa, Jun, Shwalb, Barbara J. (Eds.), Applied Developmental Psychology: Theory, Practice, and Research from Japan. Information Age Publishing, Greenwich, CT, pp. 23–42.
Nakazawa, Jun, 2015. Manga literacy and manga comprehension in Japanese children. In: Cohn, Neil (Ed.), The Visual Narrative Reader. Bloomsbury, London.
Oliva, Aude, 2005. Gist of the scene. Neurobiol. Atten. 696, 251–258.
Oliva, Aude, Torralba, Antonio, 2006. Building the gist of a scene: the role of global image features in recognition. In: Martinez-Conde, S., Macknik, S.L., Martinez, L.M., Alonso, J.M., Tse, P.U. (Eds.), Progress in Brain Research. Elsevier, pp.23–36.
Osterhout, Lee, Nicol, Janet L., 1999. On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. Lang. Cogn. Process. 14, 283–317.
Patel, Aniruddh D., 2003. Language, music, syntax and the brain. Nat. Neurosci. 6, 674–681.
Patel, Aniruddh D., Gibson, Edward, Ratner, Jennifer, Besson, Mireille, Holcomb, Phillip J., 1998. Processing syntactic relations in language and music: an event-related potential study. J. Cogn. Neurosci. 10, 717–733.
Primus, Beatrice, 1993. Word order and information structure: a performance based account of topic positions and focus positions. In: Jacobs, Joachim (Ed.), Handbuch Syntax. de Gruyter, Berlin/New York, pp. 880–895.
Pustejovsky, James, 1991. The syntax of event structure. Cognition 41, 47–81.
Rinck, Mike, 2005. Spatial situation models. In: Shah, Priti, Miyake, Akira (Eds.), The Cambridge Handbook of Visuospatial Thinking. Cambridge University Press, Cambridge, pp. 335–382.
Rosch, Eleanor, Mervis, Carolyn B., Gray, Wayne D., Johnson, David M., Boyes-Braem, Penny, 1976. Basic objects in natural categories. Cognit. Psychol. 8, 382–439.
Rumelhart, David E., 1975. Notes on a schema for stories. In: Bobrow, Daniel, Collins, Allan (Eds.), Representation and Understanding. Academic Press, New York, pp. 211–236.
Sacerdoti, Earl D., 1977. A Structure for Plans and Behavior. American Elsevier, New York.
Saraceni, Mario, 2000. Language Beyond Language: Comics as Verbo-Visual Texts, Applied Linguistics. University of Nottingham, Nottingham.
Saraceni, Mario, 2001. Relatedness: aspects of textual connectivity in comics. In: Baetens, Jan (Ed.), The Graphic Novel. Leuven University Press, Leuven, pp. 167–179.
Saraceni, Mario, 2003. The Language of Comics. Routeledge, New York.
Schank, R.C., Abelson, R., 1977. Scripts, Plans, Goals and Understanding. Lawrence Earlbaum Associates, Hillsdale, NJ.
Schwan, Stephan, Ildirar, Sermin, 2010. Watching film for the first time: how adult viewers interpret perceptual discontinuities in film. Psychol. Sci. 21, 970–976.
Sitnikova, Tatiana, Holcomb, Phillip J., Kuperberg, Gina R., 2008. Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. J. Cogn. Neurosci. 20, 1–21.
St. George, Marie, Mannes, Suzanne, Hoffinan, James E., 1994. Global semantic expectancy and language comprehension. J. Cogn. Neurosci. 6, 70–83.
Stein, Nancy L., Nezworski, Teresa, 1978. The effects of organization and instructional set on story memory. Discourse Process. 1, 177–193.
Talmy, Leonard, 2000. Toward a Cognitive Semantics, Vols 1 & 2. MIT Press, Cambridge, MA.
Thorndyke, Perry, 1977. Cognitive structures in comprehension and memory of narrative discourse. Cogn. Psychol. 9, 77–110.
van Dijk, Teun, 1977. Text and Context. Longman, London.
van Dijk, Teun, Kintsch, Walter, 1983. Strategies of Discourse Comprehension. Academic Press, New York.
van Leeuwen, Theo, 1991. Conjunctive structure in documentary film and television. Continuum 5, 76–114.
Van Petten, Cyma, Kutas, Marta, 1991. Influences of semantic and syntactic context on open- and closed-class words. Mem. Cogn. 19, 95–112.
West, W. Caroline, Holcomb, Phil, 2002. Event-related potentials during discourse-level semantic integration of complex pictures. Cogn. Brain Res. 13, 363–375.
Willats, John, 1997. Art and Representation: New Principles in the Analysis of Pictures. Princeton University Press, Princeton.
Wilson, Brent, Wilson, Marjorie, 1987. Pictorial composition and narrative structure: themes and creation of meaning in the drawings of Egyptian and Japanese Children. Vis. Arts Res. 13, 10–21.
Zacks, Jeffrey M., Tversky, Barbara, 2001. Event structure in perception and conception. Psychol. Bull. 127, 3–21.
Zacks, Jeffrey M., Tversky, Barbara, Iyer, Gowri, 2001. Perceiving, remembering, and communicating structure in events. J. Exp. Psychol. 130, 29–58.

Zwaan, Rolf A., 2004. The immersed experiencer: toward an embodied theory of language comprehension. In: Ross, B.H. (Ed.), The Psychology of Learning and Motivation. Academic Press, New York, pp. 35–62.
Zwaan, Rolf A., Radvansky, Gabriel A., 1998. Situation models in language comprehension and memory. Psychol. Bull. 123, 162–185.

Graphic References[5]

Johnson, Kikuo R., 2005. Night Fisher. Fantagraphics Books, Seattle, WA.
Kuper, Peter, 1996. Eye of the Beholder. NBM, New York.
Larsen, Erik., 2000. Savage Dragon, vol. 78. Image Comics, Orange, CA.
Mignola, Mike, Sook, Ryan, et al., 2003. Mike Mignola's B.P.R.D.: Hollow Earth & Other Stories. Dark Horse Comics, Milwaukie, OR.
Sakai, Stan, 2008. Usagi Yojimbo: Book 22: Tomoe's Story. Dark Horse Comics, Milwaukie, OR.
Schulz, Charles M., 2004. In: Groth, Gary (Ed.), The Complete Peanuts: 1953–1954. Fantagraphics Books, Seattle, WA.
Sturm, James, 2003. The Golem's Might Swing. Drawn and Quarterly, Montréal.
Takehiko, Inoue, 2002. Vagabond, vol. 15. Kodansha, Japan.
Takehiko, Inoue, 2003. Vagabond, vol. 17. Kodansha, Japan.

[5] All images are created and copyright© Neil Cohn, except those cited throughout the text. Cited images are copyright their respective owners (listed below) and used purely for analytical, critical and scholarly purposes. Most examples herein have altered the layouts to be in a linear sequence, while those marked ''abridged'' may have omitted panels from the sequence for simplicity.