COGNITIVE SCIENCE
A Multidisciplinary Journal

Visual Narrative Structure

Neil Cohn

Department of Psychology, Tufts University

Abstract

Narratives are an integral part of human expression. In the graphic form, they range from cave paintings to Egyptian hieroglyphics, from the Bayeux Tapestry to modern day comic books (Kunzle, 1973; McCloud, 1993). Yet not much research has addressed the structure and comprehension of narrative images, for example, how do people create meaning out of sequential images? This piece helps fill the gap by presenting a theory of Narrative Grammar. We describe the basic narrative categories and their relationship to a canonical narrative arc, followed by a discussion of complex structures that extend beyond the canonical schema. This demands that the canonical arc be reconsidered as a generative schema whereby any narrative category can be expanded into a node in a tree structure. Narrative ''pacing'' is interpreted as a reflection of various patterns of this embedding: conjunction, left-branching trees, center-embedded constituencies, and others. Following this, diagnostic methods are proposed for testing narrative categories and constituency. Finally, we outline the applicability of this theory beyond sequential images, such as to film and verbal discourse, and compare this theory with previous approaches to narrative and discourse.

Keywords: Narrative; Discourse; Visual language; Comics; Film

1. Introduction

Sequential images take many forms and are ubiquitous in society. Beyond the context of comics, the main subject of this study, we find sequential images in places as diverse as airplane safety manuals and the Stations of the Cross in churches (McCloud, 1993).The question to be addressed here is: What mental representations does a reader construct in the course of understanding visual narratives, and on the basis of what principles?

On the surface, images in sequence appear simple to understand: Images generally look like objects in the world, and actions in the world are understood perceptually; thus, the argument would go, understanding sequential images should be just like seeing events.

Correspondence should be sent to Neil Cohn, Psychology Department, Tufts University, 490 Boston Ave., Medford, MA 02155. E-mail: neilcohn@emaki.net

Although such an explanation appears intuitive, it ignores a great deal of potential complexity. Consider Fig. 1.

Fig. 1 might be interpreted as a man lying awake in bed, while a clock ticks away the passage of time, until he talks on the phone (either calling someone or being called). What factors involved in this sequence allow us to understand it this way?

First, a reader must be able to comprehend that the drawings mean something. How do we know that the lines and shapes depicted in panels create objects that have meaning? Physical light waves hit our retinas, and our brains decode them as meaningful, not just nonsense lines, curves, and shapes. We decode them in terms of what we will call a ***graphic structure*** of lines and shapes that underlies our recognition of drawn objects in perceptually salient ways. On a larger level, we also must be able to recognize visually this is not just one image, but a ***sequence*** of images, facilitated by the visual shapes of the panel borders. This already creates a problem: How do we know which direction the sequence progresses? Left-to-right? Right-to-left? Center outward? One aspect of graphic structure must be a ***navigational component*** that tells us where to start the sequence and how to progress through it.

Beyond the visual surface of physical lines and shapes of the panels and sequence, we must also recognize that the individual images mean something. How do we create meaning out of visual images? This must involve connecting graphic marks to ***conceptual structures*** that encode meaning in (working and long-term) memory (e.g., Jackendoff, 1983, 1990). For example, we understand that the first and last panels of this sequence depict a man (indeed, the *same* man) with a bed and a phone, the second and fourth panels depict a clock, and the third panel depicts a window with clouds and the sun. These elements compose the objects and places involved in the sequence's meaning.

Additionally, how do we know that these images are not simply flat drawings on a page? We know that these 2D representations reference 3D objects, and thereby can vary in perspective, such as between the aerial point of view in the first panel and the lateral angle in the final panel. We know that both depict the same person, despite being from different viewpoints. These are all aspects of a ***spatial structure***, which combines geometric information with our abstract knowledge of concepts. We know that the first panels depict a man and a clock because we know what men and clocks look like (iconic reference), and we retain what *this* man and *this* clock look like as we read. In fact, the same characters appear in different states of a continuous progression across panels and each image does not depict wholly new people in new scenes. Other visual narratives use more symbolic aspects of graphic ***morphology***. Things like stars above the head to indicate pain, hearts in the eyes to

Fig. 1. Visual narrative.

show lust, bubbles to show thoughts, and lines to depict motion are all conventional signs with little or no resemblance to their meaning.

Furthermore, how is it that, despite the fact that the first and second panels show only individual characters (man, clock), we recognize that they belong to a common overall environment? We construct this information in our minds through a higher level of spatial structure. This is the unseen spatial environment that we create mentally. Panels can thus be thought of as ''attention units'' that graphically window parts of a mental environment (Cohn, 2007). Within a frame, attention can be guided to the different parts of a depicted graphic space: the whole scene (man and clock), just individual characters (man or clock), or close-up representations of parts of an environment or an individual (man's hands or eyes).

Beyond just objects, how do we understand that these images also show objects engaged in *events* and *states*? For example, the first panel does not just depict a man; it depicts a man *lying awake in bed*. The final panel depicts that man *talking on the phone*. The clock *hangs on the wall* at a particular *state in time*. These concepts are aspects of the **event structure** for each panel, and an event might extend across several panels. For example, we may infer that the man lies in bed until the final panel. This event is not depicted, but we might infer this duration because we have no contrasting information until the new event in the final panel. We may also construct an additive meaning for the whole collection of events depicted: A man lies in bed, *while* the clock ticks *until* he gets up and talks on the phone. By the end of the sequence, we understand that all of these things have taken place and they are not just isolated glimpses of unconnected events.

Finally, how do we understand the pacing and presentation of events? Why does the sequence start with a state and end with an event? Could not it start with the phone call? Why bother showing the clocks and the window in the middle panels? What effect do these panels create? These questions relate to the sequence's **narrative structure**, which guides the presentation of events. We cannot understand this sequence by virtue of the individual events alone, because there are actually several possible ways to construe it (Cohn, 2003). Under one interpretation, each panel depicts its own independent time frame. Here, the first and last panels connect as one progression of events, while the succession of the clocks embeds within that of the man (as in Fig. 2A). A second interpretation might be that the juxtaposed pairs of panels of the man and the clock depict the same place *at the same time*. Here, these events occur simultaneously, despite being depicted linearly. Mentally, one must group panel 1 with 2, and panel 4 with 5, which are then connected in a singular shift in time (as in Fig. 2B).

What role does the third panel play here in the overall meaning of the sequence? Semantically, it tells the time of day, which is also made explicit from the time on the clocks. It can be understood as simultaneous with the state of either clock panel, or at its own separate time between them. However, it also provides an extra unit of pacing that prolongs the action of the man lying in bed before using the phone. This is not a prolongation of ''time'' within the event structure of the character's actions or the situation. Rather, this is a prolongation in the *narrative* pacing: It builds the narrative tension leading to the event in the final panel.

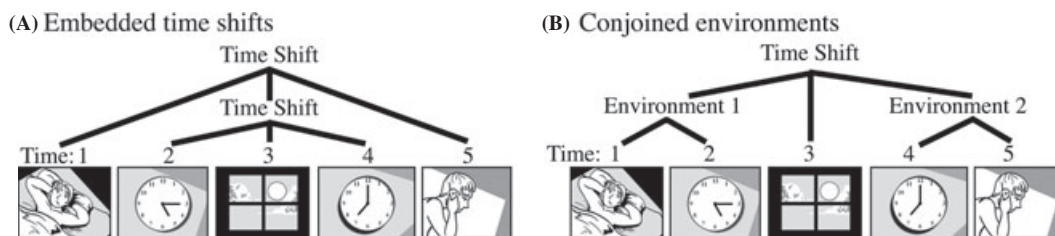**(A)** Embedded time shifts   **(B)** Conjoined environments

Fig. 2. Ambiguity in narrative structure.

So what overall issues are involved with understanding visual narratives? A graphic structure gives information about lines and shapes that are linked to meanings about objects and events at the level of the individual panel. The graphic structure also connects to a spatial structure that encodes the spatial components of these meanings, from which the reader constructs an environment in which they are situated. The narrative structure orders this information into a particular pacing, from which a reader can extract a sequence's meaning—both the objects that appear across panels and the events they engage in.

Importantly, this approach keeps narrative and event structures separate; while event structure is the knowledge of meaning, narrative structure organizes this meaning into expressible form. Previous approaches to narrative have varied in the extent to which narrative (presentation) and events (meaning) have been treated as the same or different. Some theories explicitly separate the underlying events from the narrative that orders them (e.g., Bordwell, 1985; Chatman, 1978; Genette, 1980; Tomashevsky, 1965), while this relationship is vaguer or not even addressed in other approaches (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977).

The present approach emphasizes the separation of narrative and meaning for sequential images and formalizes it in the theory of *Narrative Grammar* (specifically Visual Narrative Grammar for the visual-graphic modality). First, we establish the foundation for a theory of narrative structure by describing the basic narrative categories and their relationship to a canonical narrative arc. Following this, we explore complex structures in narrative that extend beyond the basic canonical schema. This demands that the canonical arc functions as a generative schema in which any element can be elaborated into a narrative arc of its own, recursively. Through this structure, we interpret narrative pacing as a reflection of various patterns of embedding: right- and left-branching trees, center-embedded constituents, and others. We propose a series of diagnostics for recognizing these categories and constituencies. Finally, we sketch how this Narrative Grammar can be applied to verbal discourse and film, and we conclude by discussing the connections of the proposed model with previous approaches to narrative.

2. Situating the approach

Prior to addressing the model itself, it is important to situate it in the context of previous approaches. ''Narrative'' as a whole involves many things, including the context and

circumstances surrounding a telling, the role of the author and∕or narrator and addressee, how a text constructs a world and immerses a reader into it, the emotive qualities that a text elicits, and the ordering of events into a coherent sequence which may include inferred events that have not been overtly specified (van den Broek, 1994; Herman, 2009a; Talmy, 1995, 2000a; Zwaan & Radvansky, 1998). Although all of these topics have important places in the study of (visual) narrative, this article focuses on the final facet of this overall picture: the structure of a narrative *sequence*.

The literature is consistent in finding that people prefer a particular type of sequencing in their narratives (see Table 1). However, the proposed models for the structure of sequential images vary greatly. The present approach argues that narrative categories organize sequential images into hierarchic constituents, analogous to the organization of grammatical categories in the syntax of sentences. This approach allows us to account for many important attributes that must be describable by a theory of visual narrative, all well highlighted by Fig. 1. These include the following:

1. Groupings of panels into constituents (e.g., panels 1 and 2 in Fig. 2B)
2. Interactions between the ''bottom-up'' content of panels and the ''top-down'' narrative schema
3. Description of narrative pacing through the structure of embedding (e.g., the effect of panel 3)
4. Ability to account for long distance dependencies between panels (e.g., the relation of panels 1 and 5)
5. Ability to account for structural ambiguities
6. Ability to account for how the structure of the representation facilitates inferences

Previous approaches to narrative structure have addressed some — but not all — of these traits. Notably, these traits are not unique to narrative; similar issues must be addressed in theories of syntax. The first five traits are concerned mostly with the structure of a narrative and will be the primary focus of this article, largely because of the paucity of research detailing them in visual narrative. On the other hand, the sixth trait is concerned with how structure interacts with meaning, and in fact many approaches to visual narrative have focused on the generation of inferences (e.g., Bordwell, 1985, 2007; Chatman, 1978; Eisenstein, 1942; McCloud, 1993; Saraceni, 2000). This article will incorporate some aspects of inferences; further details will appear in future work.

It is worth clarifying a few important, broad-scale differences between this approach and other models. First, the ''grammatical'' approach in Visual Narrative Grammar contrasts with approaches couched in terms of ''panel transitions'' (McCloud, 1993; Saraceni, 2000; Stainbrook, 2003), which focus on the semantic relations between adjacent images. The latter approach parallels theories of discourse that analyze the relations between pairs of sentences, either by categorizing the relationships of adjacent sentences to each other (Hobbs, 1985; Kehler, 2002; Mann & Thompson, 1987), by describing principles that combine their meanings (Halliday & Hasan, 1976, 1985; Trabasso & van den Broek, 1985), or by highlighting the types of semantic shifts that occur between them (Zwaan & Radvansky,

Table 1
Commonalities between various theories of narrative

Visual Narrative Structure	Arc, Phases	Orienter	Establisher	Initial	Prolongation	Peak	Release
Three-act play: Aristotle	Plot		Beginning (Protasis)		Middle (Epitasis)		End (Catastrophe)
Five-act play (Freytag, 1894)	Plot	Setup		Rising action		Climax	Falling action, Dénouement, Resolution
Japanese theatre (Noh, Kabuki, Bunraku) (Yamazaki, 1984)			Jo (introduction, slow beginning)		Ha (Change, speed up)		Kyu (Impact, rapid ending)
Theory of Japanese discourse (Hinds, 1976)	Discourse, paragraphs, segments	Transition	Set the stage	Evaluate		Peak	
Story grammars (Gee & Kegl, 1983; Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Nezworski, 1978; Thorndyke, 1977)	Story, episodes	Setting	Explanation of affairs; establishment of goal	Initiating event, internal response, Attempts at goal		Outcome	Reactions to outcome
Discourse theory (Clark, 1996)	Principles of embedding	Transitions: next, push, pop, return	Discourse topic, preface	Entry		Body	Exit
APA formatting			Introduction	Background, methods		Results, discussion	Conclusion
Film/comics (Arijon, 1976; McCloud, 2006)	Scene	Transitions: fade out, wipe, push, Iris, etc.	Establishing shot				

1998).The present approach is closer in spirit to that of Clark (1996) and Hinds (1976), who deal with global structure in extended conversation, as well as to syntactic theory, which is not just determined by word-to-word transitions: We examine the role that each unit plays relative to a global whole. Certain juxtapositions are indeed important for the detection of broader constituency (as will be discussed). However, the transitions between panels themselves are ultimately less important than the roles that panels play with regard to the whole narrative.

Narrative Grammar also differs from previous ''grammatical'' approaches to narrative in story grammars (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Nezworski, 1978; Thorndyke, 1977), where categories revolve around a protagonist striving for a goal. Not all narratives involve goal-direction — a story might be about an inanimate object, or it might climax in random events that interrupt a person's goals. We would still need to be able to describe such discourses in a theory of narrative structure. Goal-direction is a feature of characters *in* a story and thereby relates to its *meaning* (i.e., conceptual/event structures) rather than how the story is told (i.e., the narrative structure).The present approach is more ''form'' based: Narrative categories derive from the depiction of events in panels *and* from the contextual functions of their role in the narrative.

Finally, Narrative Grammar can be viewed as complementary to the agenda of ''cognitive narratology'' (for review, see Herman, 2003; Jahn, 2005), which invokes domain-general cognitive processes like frames, scripts, and schemas to describe narrative understanding (e.g., Bordwell, 1985, 2007; Herman, 2009a; Jahn, 1997). This domain-general approach has included comics in many of its narrative analyses (Herman, 2009a), while applications directly to comics have at least been sketched out (Bridgeman, 2004, 2005; Lefèvre, 2000). The *structural* model proposed here is not opposed to these *semantic* descriptions, and they could be integrated. However, the present work takes the view that the best way to capture the richness and complexity of narrative is through formalizing these schemas explicitly.

This article borrows methodology from linguistic analysis: Readers are asked to rely on their intuitions to assess the felicity of sequences referenced in the text.[1] This methodology has been used for decades in linguistic research, and it is acknowledged that such an approach has been criticized by some (e.g., Gibson & Fedorenko, 2010) while defended by others (e.g., Culicover & Jackendoff, 2010). Ultimately, this research program aims to extend beyond relying on intuitions. For example, experimental and corpus research is under way that seeks to validate and clarify the theory (e.g., Cohn, 2011; Cohn, 2012; Cohn, Taylor-Weiner, and Grossman, 2012; Cohn, unpublished data; Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012).

3. Narrative units

Panels make up the basic unit of visual narrative. In comics, they are discrete images ordered into a sequence with other images (Cohn, 2007; Duncan, 2000; Eisner, 1985; McCloud, 1993). Some research has described the ''grammar'' of individual images. For example, Kress and van Leeuwen (1996) propose a ''visual grammar'' for individual images based on the force

dynamic relationships between depicted elements. Engelhardt (2002, 2007) addresses the combinatorial processes and semantic elements within images, particularly maps, street signs, and other graphic displays. These visual grammars describe how composition factors into the semantics of understanding *individual* images. These elements no doubt play a role on their comprehension in a narrative. However, we will be concerned here with how this content relates to a broader *sequence*, not to the construction of this content on its own.

Panels serve two major narrative functions. First, panels act as ''attention units'' to window conceptual information in an individual image (Cohn, 2007). This is similar to how syntax windows conceptual structure in sentences (Talmy, 2000b), or how particular clauses window concepts in discourse (Hoey, 1991; Langacker, 2001; Zwaan, 2004). All of these cases use form to highlight or omit certain information. Panels do this by depicting varying amounts of information, ranging from scenes to individual characters and objects, to details of characters and objects (Cohn, 2007). The degree of content highlighted by panels can have inferential consequences for graphic sequences. For example, if two panels show different characters at the same state in time (e.g., the man and the clock in Fig. 2B), an inferential process must situate these elements into a constructed common environment (Cohn, 2003, 2010b). Corpus analysis has shown that panels between American and Japanese comics vary in the amount of information they depict (Cohn 2011; Cohn, Taylor-Weiner, and Grossman, 2012), thereby implying that these cultures' comics make different inferential demands on their readers (Cohn, 2010a). Thus, the framing of a scene can have ramifications on comprehension.

This piece will mainly focus on the second major role of panels as a *narrative* unit: What narrative roles can panels play relative to a sequence? And what features of an image's event structure cue this narrative role?

4. Narrative categories

As discussed at the outset, narrative structure orders the meaningful elements from conceptual structure. Narrative roles have been analyzed in plotlines and storytelling as far back as Aristotle's structure of theatre (Butcher, 1902). More recently, they have appeared in psycholinguistic theories of conversational discourse (Labov, 1972; Labov & Waletzky, 1967) and ''story grammars'' (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). Although ''storytelling'' is certainly a prototypical case, I wish to conceive of ''narrative structures'' simply as a method of *conveying concepts*, and, as such, they should be applicable beyond just ''stories'' (which may be an ''entertaining'' context of narrative broadly). In this context, ''stories'' are only a prototypical instance of narrative structure, and ''good stories'' are only a case of rhetorical skill.

4.1. Basic narrative categories

In Visual Narrative Grammar, there are five core categories, each of which will be discussed in detail. The core categories are as follows:[2]

Establisher (E) – sets up an interaction without acting upon it

Initial (I) – initiates the tension of the narrative arc

Prolongation (L) – marks a medial state of extension, often the trajectory of a path

Peak (P) – marks the height of narrative tension and point of maximal event structure

Release (R) – releases the tension of the interaction

Together, these categories form ***phases*** of constituency, which are coherent pieces of a structure, as in syntax. Just as phrases belong to a sentence in syntax, phases belong to a ''Arc'' in narrative. The canonical constituency structure and linear order for categories within a phase is:

Phase → *(Establisher) – (Initial (Prolongation)) – Peak – (Release)*

This rule states that a phase contains this ordering of narrative categories. The parentheses indicate optional categories; except for Peaks, they each can be left out of a sequence with no significant structural consequences. Peaks, and to a lesser degree Initials, are the most important components to the structure of narrative. In turn, each category can also serve as a phase. We will address this capacity for expansion in the next section. First, we must describe the properties of these categories in more detail, focusing on the semantic cues that motivate their categorization. For the sake of simplicity, most examples will come from basic short comic strips; but as we will see, these structures can be elaborated on to develop far more complex examples, as in sequences from long-form comic books, Japanese manga, or even other instances of sequential images, such as instruction manuals. As such, this approach extends beyond short comic strips or isolated visual sequences, though they make for easier examples.

4.1.1. Peaks

Although narratives have many parts, one panel often motivates the meaning of the sequence. Consider Fig. 3A, which shows a woman smacking a man in the head. The penultimate panel constitutes the ***Peak*** of the sequence: the location of the primary events of the sequence or phase.

The Peak is where the most important things in a sequence happen, and it motivates the context for the rest of the sequence. Prototypically, Peaks correspond with the culmination of an event, or the confluence of numerous events. They are the realization of Todorov's (1968) narrative disruption of equilibrium. Because of this, Peaks may show the interruption of events, which create alterations to the expectations of an event structure. This occurs in Fig. 3B: The soccer players interrupt the dog's chase. In this regard, Peaks best capture the crucial aspect of surprise in many narratives (Brewer & Lichtenstein, 1981; Sternberg,
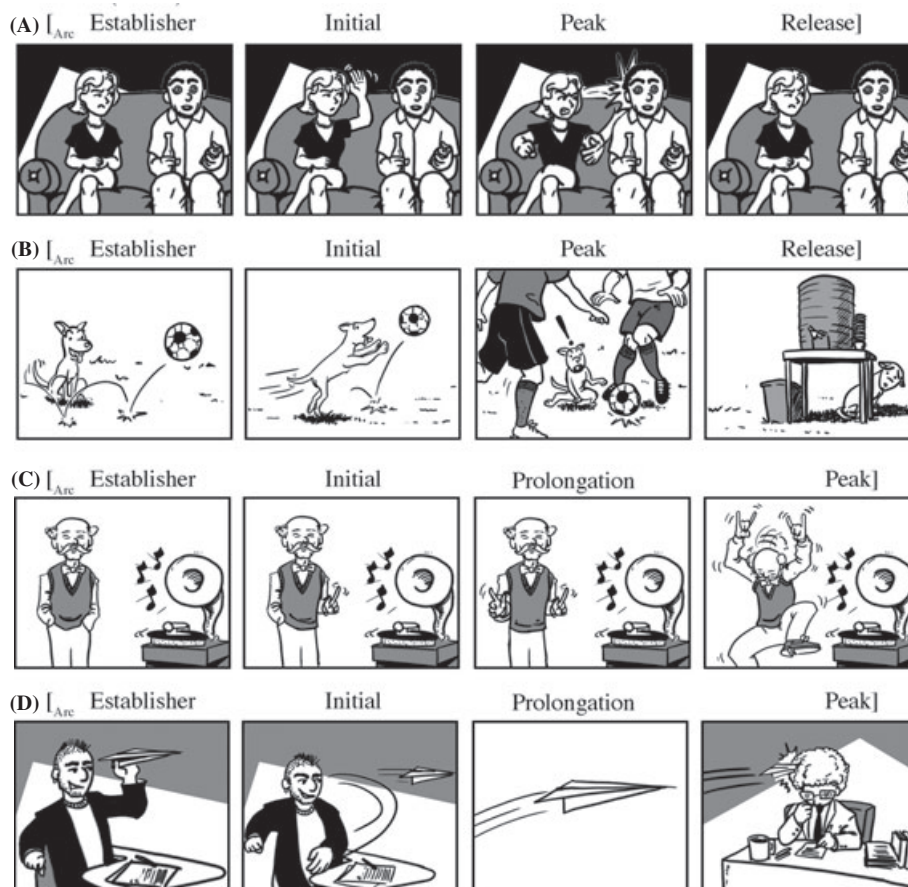
Fig. 3. Comic strips with narrative structures glossed.

2001). Indeed, a surprise would be difficult to reveal in a place other than the culmination of a narrative.

When an action involves a trajectory, Peaks prototypically map to the Goal of the Path. For example, when throwing a punch or cutting with a sword, the event is fulfilled at the endpoint of the object's path. The Peak in Fig. 3D shows this: The paper airplane's endpoint is in the teacher's hair. Such motion also aligns with the endpoint in a transference of energy, in the sense of Talmy's (1988, 2000b) force dynamics.

Similarly, Peaks might contain a change of one state to another, especially in an interruption or termination of a process (as in Fig. 3B). Also, a growing event may culminate in a Peak. Consider Fig. 3C, which shows an older man dancing passively until he breaks loose and starts headbanging. The final panel is the Peak, as it shows the events reach their apex (him fully rocking out). Also, Notice that panels 2 and 3 show virtually the same event, but extending it throughout two panels builds the narrative tension toward the culmination in the Peak. As will be discussed, this is a function of narrative, not just events.

4.1.2. Initials

Following Peaks, Initials are the second most important part of a phase, because they set the action or event in motion. They create the disequilibrium in Todorov's (1968) sense. Consider the panel just prior to the Peak in Fig. 3A. Here, the woman reaches back her arm in preparation to smack the man. This panel starts her action, but it does not climax until the next panel. This preparatory event maps to an ***Initial*** in narrative structure: It *initiates* the primary event of the sequence. Similarly, in Fig. 3C, the Initial shows the man start to groove to the music without yet fully rocking out.

Initials can be related to Peaks in several different ways. The prototypical Initial shows an inception or preparatory action that culminates in the Peak. For example, the woman's reaching action in Fig. 3A is a preparation to smack the man. As they contain the start of an action, Initials often mark the Source of a path, as in any event that involves a trajectory (as in the Initials in Fig. 3A and 3D). These properties of Initials derive bottom-up from the panel's content.

A second type of Initial relies more on the panel's context in the narrative than the depicted event structure. Consider the Initial in Fig. 3B. It shows the dog chasing a soccer ball prior to being interrupted by the soccer players. This Initial does not show a preparatory action; it shows the dog already chasing the ball. However, this process is interrupted in the Peak. Only after the Peak has been reached can the previous panel be recognized as an Initial. Thus, this type of Initial is defined more by its contextual relationship to the Peak than by its internal content.

4.1.3. Releases

In Fig. 3A, the final panel depicts the woman looking angrily at the man — the aftermath of the Peak's action. This panel is a ***Release*** for the narrative tension of the Peak and gives a ''wrap up'' for those events, often as an outcome or resolution. Prototypically, this aftermath involves the coda of an action, such as the retraction after throwing a punch or swinging a sword. Alternatively, it may show a passive state after an event, such as a person's return to standing *en garde* with his or her hands or a sword. In the case of Fig. 3A, the passive state is that the smack in the Peak had no effect on the man.

Releases also may involve a reaction to the events in the Peak. For example, the final panel in Fig. 3B shows the dog hiding behind a water cooler after being assaulted by soccer players. This panel does not relate directly to the actions of the soccer players: A Release of this nature might show the dog flattened on the grass. Rather, this Release provides the dog's reaction of running and hiding.

Finally, many strips are funny *because* of the Release. In Figs. 3A and 3B, the culmination occurs in the Peak, but the actual punchline is delivered in the Release. Thus, Releases provide an important panel for humor, perhaps because they convey an aftermath, response, or a (relatively) passive follow-up to the climax of the sequence.

4.1.4. Establishers

The first panel in Fig. 3A depicts the woman sitting next to the man, not doing anything in particular except looking at him. They are simply at a state of ''being'' prior to the actual

actions of the sequence. This ***Establisher*** provides referential information (characters, objects) without engaging them in the actions or events of a narrative. This most often involves a constant state or process that is changed by the events of the narrative.

Consider the first panel of Fig. 3B. This Establisher shows the dog watching a soccer ball bounce in front of him. The ball is in a process of bouncing, while the dog is surprised/curious. Despite the high degree of ''action'' in the panel, it functions to set up the relationship between the dog and the ball. This panel does not just *establish* the relationship narratively to the reader but also semantically in the strip: The dog here discovers the ball, as opposed to already engaging with it. Bordwell (2007) has noticed that film scenes often open with characters entering into the shot or moving toward the viewer (as in this panel in 3b). This allows the character to enter the scene concurrently with the viewer's entrance into the narrative — an act of Establishment. (The finale of a film often reverses this, with characters leaving a shot or moving away from the viewer.)

Establishers give the first glimpse of a scene and thereby set up the characters. This process can facilitate Gernsbacher's (1990) notion of ''laying the foundation'' for the building blocks of a discourse. Just as the first sentence of a discourse provides new information (Haviland & Clark, 1974), the Establisher lays the foundation of new information for a sequence. Establishers can also lay the groundwork for what Herman (2009a) describes as ''story building'' — the construction of a fictive environment in which a reader can be immersed.

Establisher and Release panels are often similar (as are Prolongations, discussed next). Some narratives even make this overt, as in Fig. 3A, where the first and last panels are identical. ''Returning to the start'' is a common narrative theme. This is what makes Freytag's (1894) model of plotlines ''triangular'' in shape: The ending connects back to the beginning. It also appears in Todorov's (1968) notion that narratives return to equilibrium after disruption.

Nevertheless, though the first and last panels of Fig. 3A are identical, they play different functional roles as Establisher and Release. Thus, here again top-down contextual information from the sequence interacts with a panel's intrinsic content to determine its narrative role. However, even when a sequence begins and ends with the same panel, the Release appears more important to the narrative than the Establisher. Deleting the Establisher in Fig. 3A would make little impact on the sequence compared with deleting the Release.

Establishers are parallel to several notions in discourse and narrative, listed fully in Table 1. Clearly, they relate to ''establishing shots'' in film (Arijon, 1976; Bordwell & Thompson, 1997; Carroll, 1980) and in comics (Madden & Abel, 2008; McCloud, 2006). However, they do not necessarily require an expansive long-shot viewpoint highlighting the broader environment, as film and previous comics work suggests. Notions similar to Establishers even appear in contexts without concepts, such as in the setting of a mood or rhythmic texture at the outset of musical pieces, including ''vamps'' in pop songs or the opening ''alap'' or ''alapana'' that establishes the ''raga'' of Indian music (Jackendoff & Lerdahl, 2006). More commonly, Establishers function similarly to discourse Topics, frame setters, or storytelling Prefaces (Clark, 1996; Jacobs, 2001; Krifka, 2007). They conform to the

Table 2
Primary correspondences between narrative categories and conceptual structures,
in order of importance to a narrative Arc

Narrative Category	Conceptual Structure
Establishers	Introduction of referential relationship
	Passive state of being
Initials	Preparatory action
	Process
	Departing a Source of a path
Prolongations	Position on trajectory of a path
	Sustainment of a process
	Passive state (delaying)
Peaks	Culmination of event
	Termination of a process
	Interruption of event or process
	Reaching a Goal of a path
Releases	Wrap up of narrative sequence
	Outcome of an event
	Reaction to an event
	Passive state of being

general observation that discourse prefers to describe who is doing an action before describing the action itself (Primus, 1993; Sasse, 1987).

4.1.5. Prolongations

Beyond the core narrative roles, a modifying category can be used to hold off the realization of a Peak. A ***Prolongation*** marks a medial state in the course of an action or event. Prolongations often depict the trajectory between a Source and Goal, sometimes clarifying the manner of the path. For instance, the third panel of Fig. 3D shows a medial state in the trajectory of the paper airplane from the student (Source/Initial) to the teacher's hair (Goal/Peak) and could easily be omitted with no semantic consequences for the sequence. However, narratively, it holds off the Peak for another panel. To this purpose, Prolongations can function as a narrative ''pause'' or ''beat'' for delaying the Peak, adding a sense of atmosphere, and/or building tension before the Peak (as in the third panel of Fig. 3C, or the central panels in Fig. 1). This allows an author to draw out a scene, or perhaps to end a page (or daily episode) with a Prolongation to leave readers in suspense until its resolution.

4.2. Summary

This section has established the basic categories of Visual Narrative Grammar. Peaks, Initials, and Releases appear to be core categories, while Establishers and Prolongations are more expendable. These categories fall into a canonical pattern within ''phases'':
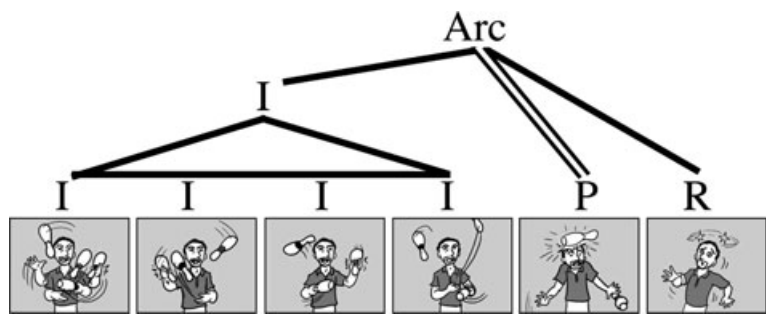
Fig. 4. Complicated narrative structure.

Phases → (Establisher) – (Initial (Prolongation)) – Peak – (Release)

The categories/functional roles are summarized in Table 2.

5. Combinatorial structure in narrative

We have now established several categories that serve as the ''parts of speech'' or grammatical functions in Visual Narrative Grammar. However, so far, these pieces use only a variant of the canonical narrative phase. The implication would be that sequences must use this pattern and could not be more than five panels long. However, consider Fig. 4.

In this example, a man is juggling and then gets hit in the head with his juggling pin. The Peak of this sequence is the penultimate panel, where he gets hit in the head, followed by a
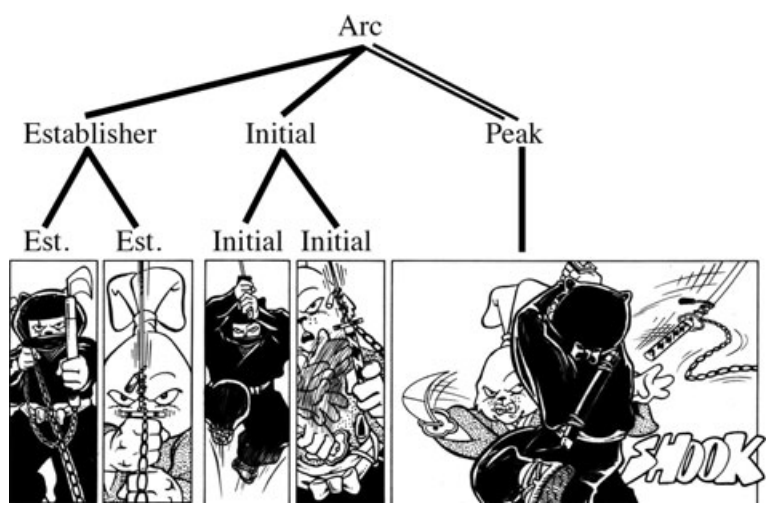
Fig. 5. Alternation between characters to build tension. UsagiYojimbo art © 1987 Stan Sakai.

Release of him stumbling dizzily. However, what exactly is happening in the first four panels? These panels all show roughly the same information: the man juggling. This repetition can be captured with what will be called ***Conjunction***: All four of these panels are co-Initials of an ''Initial Phase'' that sets up the Peak. We can formalize Conjunction as:

A phase uses Conjunction if…

…that phase consists of multiple panels in the same narrative role. Semantically, this often corresponds with…

1. various facets of an iterative process,
2. various viewpoints of a broader environment or individual, or
3. various images tied through a broader semantic field.

In the case of Fig. 4, the Conjunction shows an iterative process. Semantically, all these panels contain the same information, which could be achieved with fewer units (indeed, the first panel shows the full event). However, by extending this action across several panels, *narrative* tension and pacing builds until a culmination in the Peak. This example again highlights how narrative structures differ from events: The event structures constitute the meaning (juggling), while the choice to extend it across four panels involves how the meaning is conveyed, the narrative structure.

Conjunction can correspond to semantic content beyond showing iterative events. For example, in Fig. 5, the opening four panels alternate between the characters involved in the interaction, using pairs of Conjunctive phases. Flipping back and forth between characters builds the tension of the sequence, which then culminates in the final Peak panel, ***converging*** on both characters. This sense of narrative pacing would be lost if the Establisher and Initial each used only one panel with both characters together, as all panels would contain the same amount of information. Dividing the scene into parts allows the narrative rhythm to build until a culmination in the Peak.

There is a second way to build larger narrative structures. Consider Fig. 6. In this sequence, a man looks at a juggling pin and then throws it away. On its own, this sequence is a fairly banal narrative with an Initial and Peak. However, if it is added to the end of the sequence in Fig. 4, additional structure appears, as shown in Fig. 7.

By combining these strips, we create a larger narrative structure: The primary events in the first six panels set up the aftermath in the final two panels. Each grouping forms its own
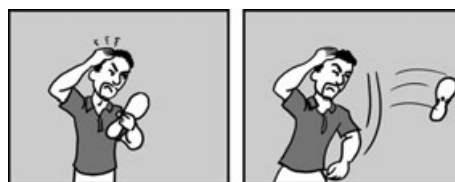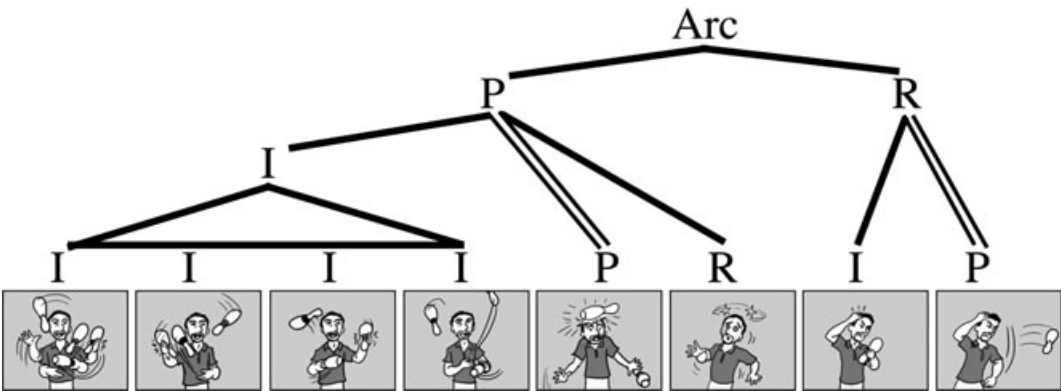

Fig. 6. Short narrative sequence.

Fig. 7. Embedding of phases.

phase. The first six panels together constitute a Peak — the main component of the overall narrative — while the final two panels combine as a Release. Each of these phases, which can stand alone, plays a larger narrative role relative to the other when combined. In this way, narrative categories apply to sequences of panels as well as individual panels.

The panel that motivates the meaning of a phase can be thought of as the ''head'' of that node. Normally, Peaks are the default heads of phases. As notated by double bar lines in Fig. 7, the Peak of the man being hit heads the first Peak phase, while the final Peak panel of him throwing the pin heads the Release phase. Thus, a Peak provides the key component in its local phase. The other categories simply support, lead up to, or elaborate upon a phase's Peak. In other words, Peaks drive the narrative sequence. Because of this importance, deletion of all non-heads from a sequence should adequately paraphrase the sequence's meaning, as in Fig. 8.

We see then that sequential images can be elaborated structurally in two primary ways. First, constituents can repeat the same category in a Conjunction that shares their narrative category at the phasal level. Second, a whole phase can take a narrative role in a broader structure, meaning that phases can also be embedded inside each other. These two strategies create numerous possibilities for structure.

Consider Fig. 9, which has a sequence of Peaks after the first panel. Instead of belonging to a singular Conjunction, they create a *left-branching structure* where each serves as an Initial for the next. It begins with ''Tarzan'' preparing to jump, which as an Initial sets up

Fig. 8. Paraphrase of Fig. 7 with only grammatical heads.

Fig. 9. Left-branching visual narrative.
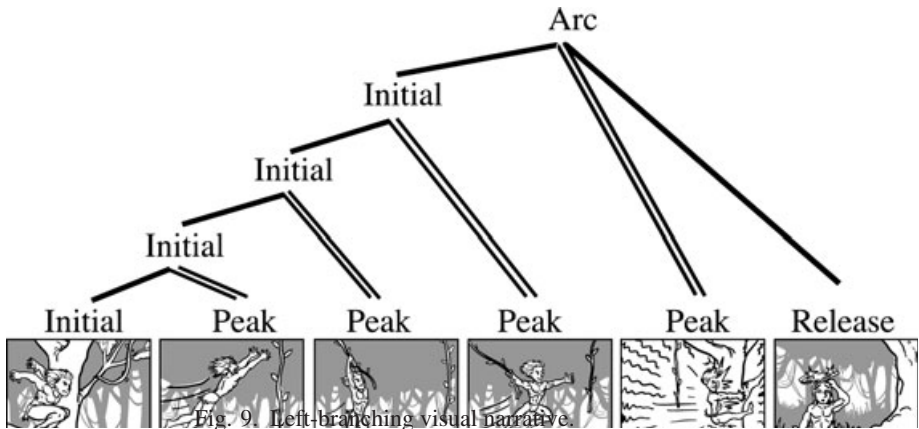
the Peak of his leaping for a vine. Together, these panels act as another Initial for a Peak of him swinging on that vine. These panels then set up a Peak of him reaching for a new vine, which culminates in the main Peak of the strip, where he slams into a tree. The resulting structure has recursive embedding, with the left-branching structure here creating the feeling of progressive building actions and/or increasing narrative tension.

In contrast to the left-branching structure, consider Fig. 10. Here, the first four panels alternate between sunny and rainy weather over the man in a sweater. Each pair acts as an Initial phase, and together they form a Conjunction that acts as an Initial for the final Peak, where the sweater shrinks. We can confirm this Conjunction because either of these Initial phases could be deleted to little overall effect. In contrast, deleting a whole phase in the left-branching structures in Fig. 9 would drastically alter the reading. The embedding of phases also reflects the narrative pattern: The ''on-off'' pattern of panels (sun-rain-sun-rain-sun...) builds until the final panel, where the pattern is broken (...sun-shrink).
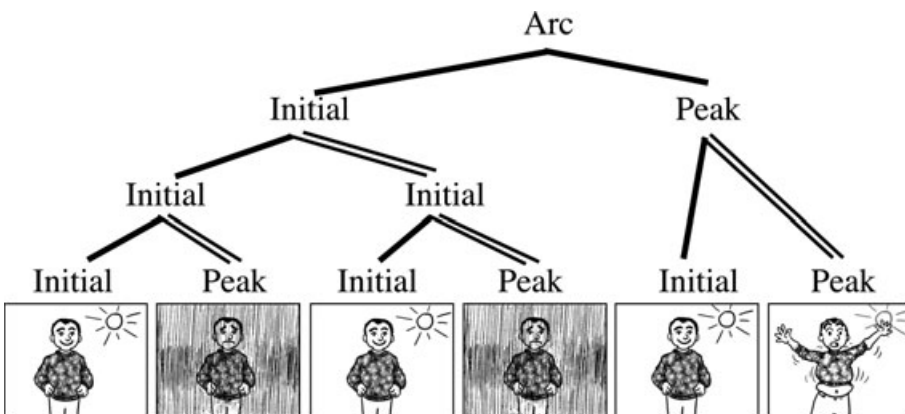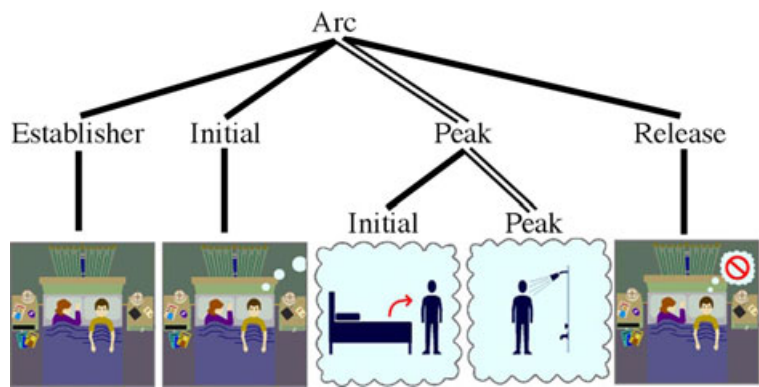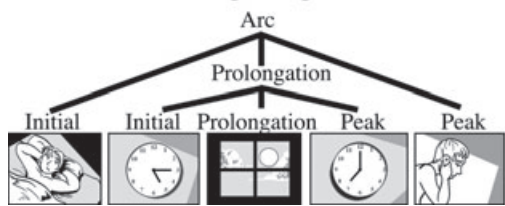

Fig. 10. Alternating initials.

Fig. 11. Embedded phase. One Night art © 2006 Tym Godek.

Narrative structure can also use ***center-embedded phases***, as in Fig. 11. This strip shows a man lying in bed. He thinks about getting up and going to the bathroom but decides not to get up. His thoughts stand alone as an embedded phase (here a Peak phase) that could be separated from the rest of the strip. In fact, both the embedded phase and the surrounding ''matrix'' phase could constitute their own felicitous sequences.

We now have enough machinery to adequately analyze Fig. 1, the ambiguous sequence of the man and the clock. The sequence opens with two Initial panels, which both show states that undergo a change in the final two Peak panels. The central panel of the windows is a Prolongation because it delays the Peak. As was mentioned before, the ambiguity of this sequence rests on whether the panels of the man and the clock belong to the same time frame. Under one interpretation (12a), the first and last panels form their own phase. The clocks then form another subordinate phase, which embeds as a Prolongation to delay the final Peak. This structure implies that all of the panels depict different times. A second interpretation (12b) conjoins the adjacent Initial and Peak panels. Here, the first two panels belong to a single Initial phase, while the last two panels belong to a single Peak phase. Each constituent depicts facets of a broader spatial environment at a single moment in time.[3]Thus, the two rules proposed by this model, for phases and conjunctions, allow us to parse a single ambiguous sequence into multiple interpretations.
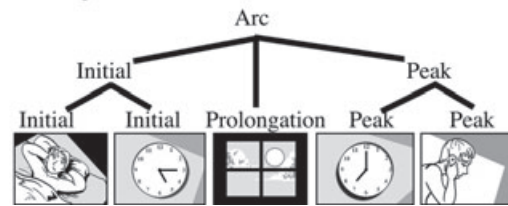

Fig. 12. Re-analysis of Fig. 1.

5.1. Summary

This section has explored how narrative categories extend beyond the panel level. We have described two types of relationships between phases and categories, which vary based on which category assigns a role to the phase (if no assignment, it becomes an Arc). A phase is an elaboration of its Peak taking the form of a subordinate narrative Arc, and it plays a narrative role in the larger structure in which it is embedded. A phase may contain multiple daughters of the same type, which function as co-heads. This results in only requiring two rules to organize numerous narrative categories. The system can generate an infinite array of larger patterns: left-branching trees, center-embedded phases, alternation, etc.

6. Diagnostics for the structure of narrative

Given a novel sequence, how can we test a panel's category, or where the boundaries of a constituent lie? In linguistics, various diagnostics have been developed for recognizing the category of a structure (be it phonemes, words, or phrases), as well as the constituents of that structure. We will now apply these methodologies to the structures involved with narrative. As before, the focus here will rely on judgments based on intuitions. However, these judgments can be tested experimentally as well (Cohn, unpublished data).

6.1. Narrative categories

An important issue is determining a panel's narrative category. One method is to rely on the semantic cues of its depiction. For example, we might expect any panel showing a preparatory action to automatically be an Initial. However, this might not always be the case, as the same unit can play multiple roles depending on its context (Jahn, 1997; Sternberg, 1982). For example, all the Initials in Fig. 10 are passive states. Thus, beyond semantic cues, how can we determine the category of a panel in a given strip?

Linguistics uses various diagnostics to test the syntactic category of a word in a sentence. These include substitution, alteration, deletion, or reordering of a word or phrase. By analogy, we can use similar diagnostics to identify narrative categories.

6.1.1. Substitution

Just as phrases can be replaced by a pronoun in syntax, some narrative categories can be replaced with another panel. Consider Fig. 13. Its fifth panel is an ''action star'' that plays the role of a Peak (here, where the security guard gets hit by the tossed backpack).

An action star panel indicates the culmination of an event (characteristic of a Peak), without revealing what it actually is. Action stars thus allow narrative structure to be retained without being specific about event structures. They thereby force an inference of a ''hidden'' event. This makes an action star almost like a ''pro-Peak,'' comparable to a pronoun or other pro-forms that have a grammatical category but minimal semantics. Thus, just as a

Fig. 13. Action star in a visual sequence. Grrl Scouts art © 2002 Jim Mahfood.

pro-form can replace its corresponding grammatical categories, substituting an action star for a panel serves as a diagnostic test for Peaks.

An action star can replace nearly any Peak panel, especially one that features some sort of impact, but switching it with any other category is infelicitous. Consider the sequences in Fig. 14, which insert action stars into Fig. 3B.

When the action star replaces the Peak (14a), the sequence reads acceptably. However, when moved to the Initial (14b), the felicity of the sequence worsens. Nevertheless, in 14a, we no longer see the soccer players; we only know that some event frightened the dog (perhaps the ball exploded?). In other words, an action star demands that the reader infer the missing event.

6.1.2. Alteration

Some panels can be altered in ways that reveal their narrative categories. For example, humorous sequences often have punchlines in the Release as a response to the Peak. This quality can be playfully highlighted by the fact that a word balloon saying ''Jeez, what a jerk!'' can be added to the final panel of nearly any comic strip without it losing its felicity (Sinclair, 2011).[4] More specifically, such a balloon can be attached to any Release, not to any final panel. It works because this phrase is pragmatically a response to an action, and thus only makes sense in panels that share this context.
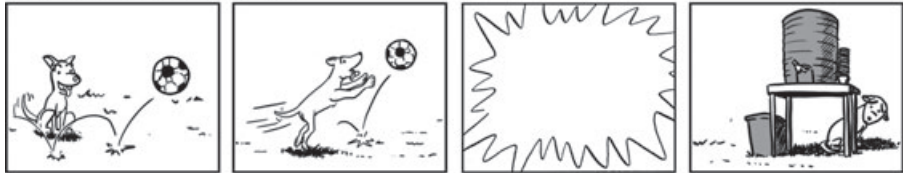
Fig. 14. Action stars substituting for narrative roles.

6.1.3. Deletion

Deletion can also be used as a diagnostic. For example, as discussed, deletion of non-Peak panels from a phase results in a narrative Arc with about the same sense. As a result, a simple diagnostic for Peaks is that they are the only panels in a phase capable of paraphrasing the meaning of the entire phase, as in the paraphrase of Fig. 7 with Fig. 8.

Deletion of individual panels offers insight both into the characteristics of those panels and into the inference created by their omission. First, because Peaks are so important for the sequence, their deletion creates large inferential demands. Take for example Fig. 15.

The first panel shows a man skating backward past two spectators, which is an Establisher setting up the scenario. The skater then opens his eyes and notices something in the second panel, an Initial. The final panel then shows his legs sticking out of a broken window of an antique store, while a sexy woman looks on. This panel is a Release, showing a prototypical aftermath of an event. However, the primary action (the Peak) of the sequence is missing: We never see him crash into the window. We only infer this event by seeing its result in the Release. Furthermore, depicting the woman in the Release reveals that she distracted the skater in the Initial. Thus, the final panel demands that the *given* graphic sequence be reanalyzed and the unseen information inferred.

In other cases, it is impossible to infer a deleted Peak. Fig. 16A shows a sequence where the Peak has been deleted from Fig. 3B. It no longer makes sense without the Peak. Why is the dog suddenly scared? Inference alone cannot fill in this information, as the Peak was an unexpected interruption. By and large, deletion of Peaks damages a sequence's felicity.

16b deletes the Initial and information is noticeably missing; it jumps from an established relationship to a culminating event. This is particularly pronounced where the Peak is an interruption. Studies of verbal narrative echo this feeling: participants report that the deletion of initiating actions in a discourse creates a more ''surprising'' narrative (Brewer & Lichtenstein, 1981). Like Peaks, deletion of Initials often strains the comprehension of sequences.

In contrast, the deleted Release in 16c renders the events of the sequence fairly complete, with less indication that something is missing. Nevertheless, the sequence ends abruptly and leaves the reader expecting that something should come next. Unlike Initials, the aftermath in a Release may not be inferable from other panels' contents.
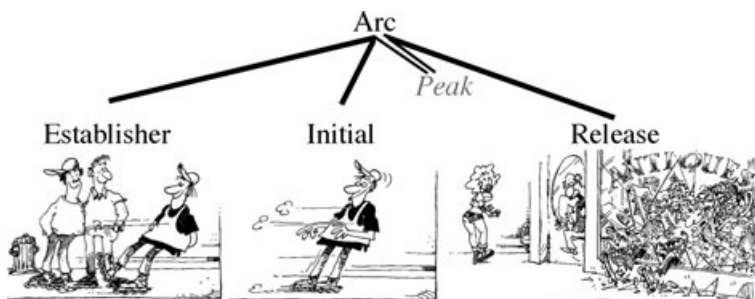

Fig. 15. Sequence requiring a Peak to be inferred. Actions Speak art © 2002 Sergio Aragonés.
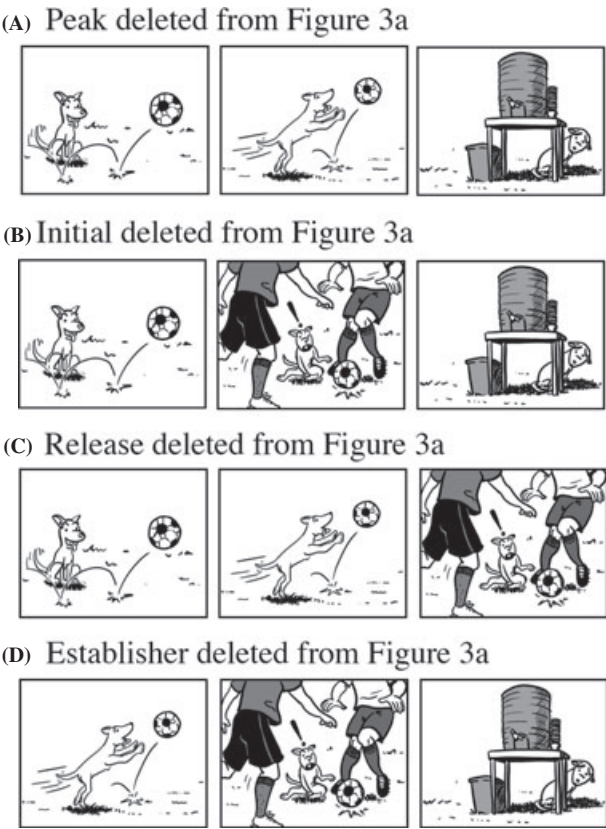
Fig. 16. Deleted narrative categories.

Finally, 16d shows a sequence without its Establisher. This alteration has almost no impact on the sequences — you can hardly tell that the panel is missing! Studies of film have shown similar results when establishing shots are deleted; they have almost no effect on the overall comprehension of the film (Kraft, Cantor, & Gottdiener, 1991). Because Establishers set up characters and interactions, this information is often redundant with subsequent panels, where those characters engage in the actions of the sequence. So deleting Establishers should make little difference to the sequence's meaning. However, it can impact the narrative *pacing*. By leaving out an Establisher, the actions immediately appear at the first panel. This leaves no ''lead in'' time for the reader to be acclimated to the elements involved prior to their events.

In sum, deletion can reveal the characteristics of categories and the inferences that omissions create. The generation of inference is the same across all panels in a sequence and greatly depends on the content of what is deleted as well as the surrounding context. That is, the understanding of inferential processes benefits from detailing the formal properties of a narrative.

6.1.4. Reordering

Categories can also be distinguished through reordering. As mentioned earlier, a panel that acts as a Release can often also start a sequence as an Establisher, as in Fig. 17A. Here, the dog starts off scared, only then becomes curious about the ball and chases it. The reverse should be just as good: An Establisher can be reordered to the end of a sequence, to where it acts as a Release. In Fig. 17B, a panel that once introduced the dog and the ball as an Establisher now shows the dog's frightened response to an object that earlier led to danger.

In contrast, other reorderings work less well. Bringing a Peak to the front worsens the sequence, as the culmination then precedes its lead up. Fig. 17C should feel weird because of this, as the Peak makes no sense out of context. A sequence-final Initial also feels unusual. In Fig. 17D, the dog's enjoyment in chasing the ball seems odd given the prior context.
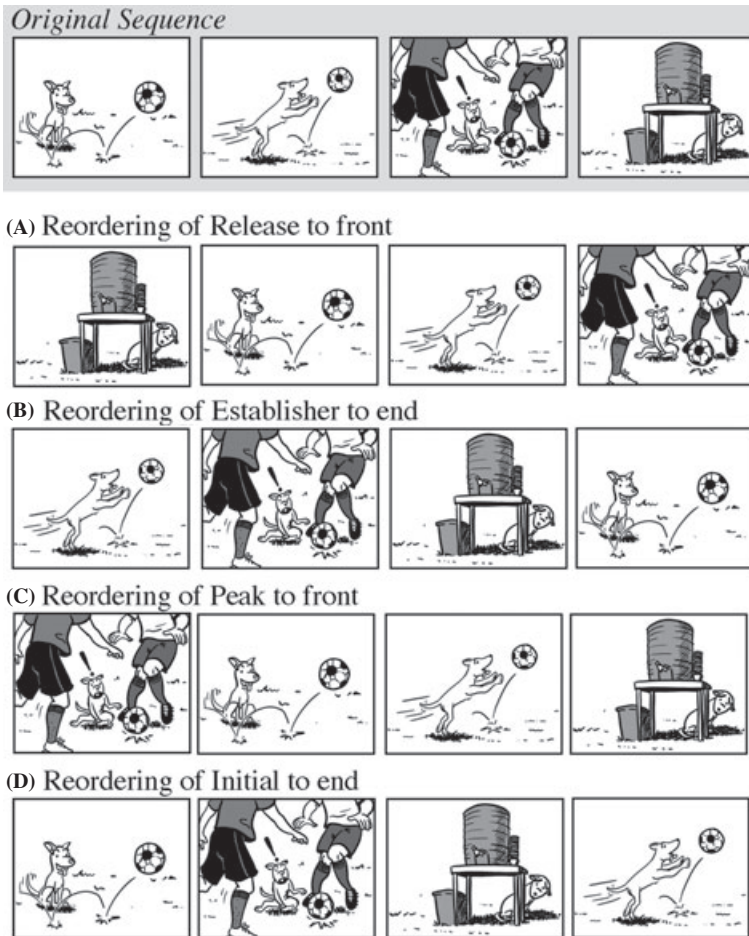

Fig. 17. Reordering of categories to other places in a sequence.

6.1.5. Summary

In summary, various techniques can be used to identify the category of a panel beyond its intrinsic semantics. Substitution, alteration, deletion, and reordering all provide ways to assess the attributes of different panels. Altogether, several diagnostic questions can be asked of each narrative category:

Establishers – Can it be deleted with little effect? Can it possibly be reordered to the end of a sequence and function as a Release?

Initials – Does its deletion affect felicity (though its absence may be inferred by the content of a Peak)? Is it impossible to reorder it in the sequence without disrupting felicity?

Prolongations – Can it be deleted with little effect?

Peaks – Can this panel paraphrase a sequence? Can an action star replace it? Does its deletion make large inferential demands and adversely affect felicity?

Releases – When a sequence's Release is deleted, can it still show a coherent event, yet be an infelicitous narrative? Does the panel carry a punchline? Can you insert the phrase ''Jeez, what a jerk'' as a speech balloon?

6.2. Testing for constituency

Beyond recognizing individual categories, we also need a way to detect the boundaries of constituents. Like categories, one method may be to rely on semantics. Research on event structure has long noted that the boundaries of events align with changes in characters, locations, or causation (Newtson & Engquist, 1976). These semantic cues may aid narrative structure as well. Take for instance Fig. 18. Here, the first phase is about a batter, while the second phase is about an interaction between a base runner, a catcher, and an umpire. The boundary between constituents features a change in characters.

Studies of film support that significant semantic shifts occur at the boundaries of events. Linear shifts in narrative along dimensions of goals, space, causes, and locations frequently correlate with the ending of one event and the beginning of the next (Zacks & Magliano, 2011; Zacks, Speer, & Reynolds, 2009; Zacks, Speer, Swallow, & Maley, 2010). Here, we find a connection between the hierarchic narrative grammar outlined here and approaches that focus on linear semantic relationships between images (e.g., Magliano, Miller, & Zwaan, 2001; McCloud, 1993): Significant semantic changes between juxtaposed images signal the boundaries between hierarchic constituents.
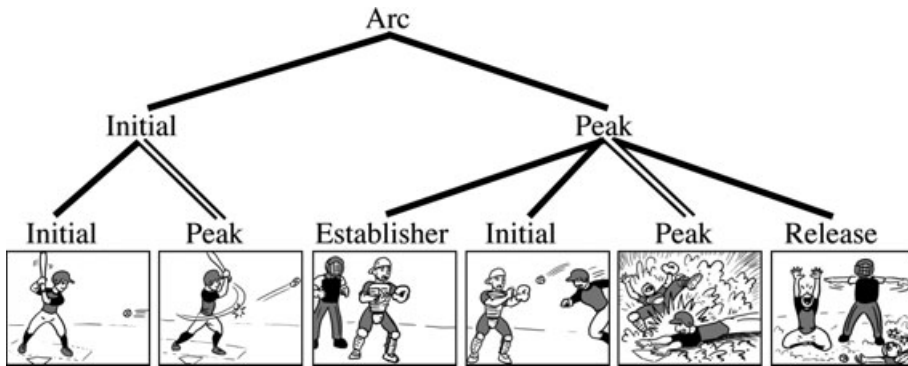
Fig. 18. Narrative sequence with two constituents.

Nevertheless, semantic shifts might not always indicate the edge of a boundary. For example, all four opening panels in Fig. 5 alternate between two characters, yet they do not *all* mark constituent boundaries because they use Conjunction. Thus, while they correlate strongly, transitions between panels do not map one-to-one with phase boundaries and cannot be relied on as the sole indicator of constituents.

As in the previous section, several additional diagnostics can be used to test for the boundaries of constituents. Like the investigation of phrases in sentences, these techniques include windowing, deletion, reordering, and alteration.

6.2.1. Windowing

The boundaries of a constituent can be made salient by extracting a subsequence and seeing if it makes sense. This technique is used in syntactic analysis of phrases. For example, selecting only three words at a time from the sentence ''My lazy roommate watched the television'' can show that ''My lazy roommate'' and ''watched the television'' are complete phrases, but ''lazy roommate watched'' and ''roommate watched the'' cannot grammatically stand alone.We can use a similar technique with visual narratives.

Consider Fig. 19, where the sequence from Fig. 18 is windowed into three-panel segments. Fig. 18 involves a two-panel phase followed by a four-panel phase. We can confirm these constituents as segments crossing the phase boundary (as in 19a and 19b) cannot stand alone as sequences. In contrast, the segments in 19c and 19d can stand alone, as they feature a complete phase.

6.2.2. Reordering

Conjunctions can be tested by reordering panels. Conjunction joins together panels that are either: (a) at the same temporal state, (b) iterative parts of a larger event, or (c) fragments of a broader semantic field. Rearrangement of panels should not dramatically impact these phases, because no temporality would be violated. Thus, reordering adjacent panels can test for Conjunction. If the reordering has little impact on the meaning of the sequence, it is likely a conjunctive phase. If reordering does change the meaning or temporality of the sequence, then it is likely not a Conjunction.
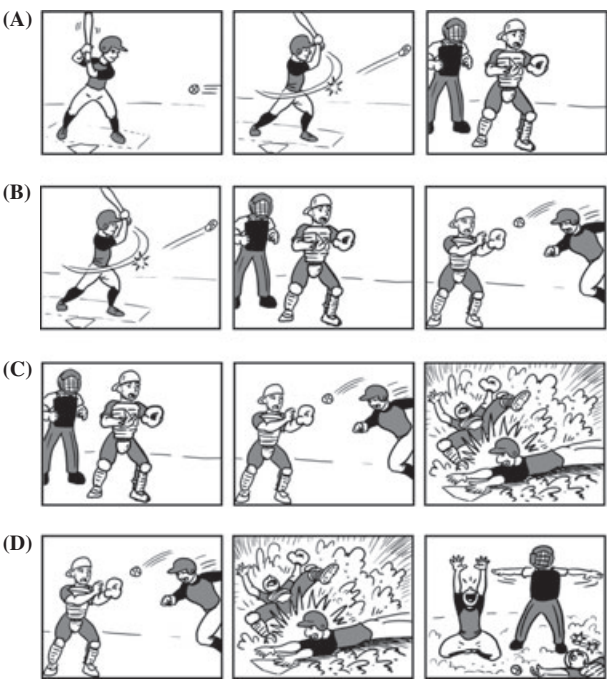
Fig. 19. Windowing of three-panel segments of Fig. 18.

For example, Fig. 5 features two successive phases that use Conjunction. Fig. 20A reverses the order of panels in the first phase. This changes the alternation of characters in narrative pacing, but it makes no impact on the sequence's meaning. In contrast, Fig. 20B exchanges panels from across the constituency boundaries. This change makes the shift between characters appear strange; panels 1 and 3 in this sequence clearly belong at the same time, yet they are separated by another action.

6.2.3. Summary

Like categories, constituents can be identified through semantic criteria (transitions between panels signaling phase boundaries) as well as through diagnostic tests. Regular phases can be tested using windowing and deletion, while Conjunction can be tested using reordering. These tools allow for sequences to be assessed within this theoretical model.

7. Beyond visual narrative

Thus far the discussion of narrative has centered on the visual-graphic modality. However, if narrative structure transcends a single modality, we would expect the same structures to be used in the comprehension of verbal[5] and filmic narratives, an assumption held across theoretical models (Bordwell & Thompson, 1997; Branigan, 1992; Carroll, 1980)
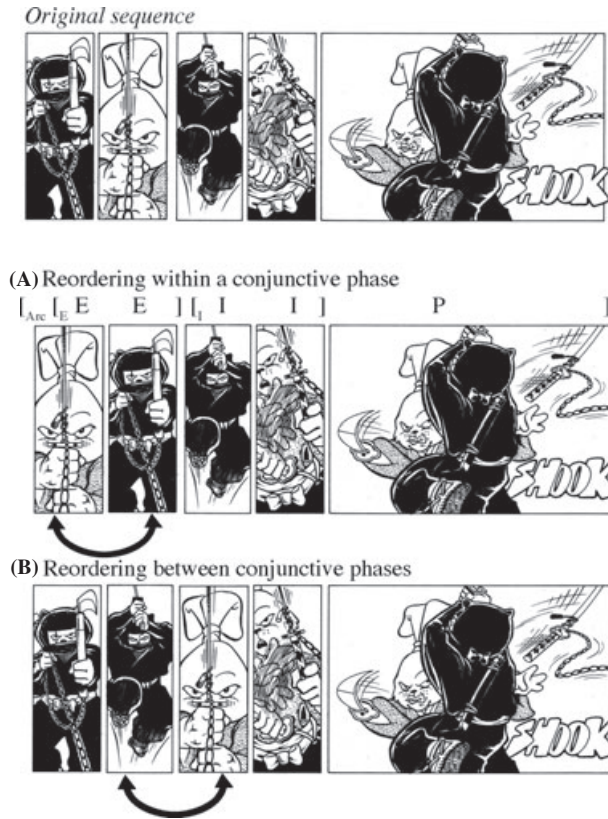
Fig. 20. Reordering within and between conjunctive phases.

and empirical experimentation (Gernsbacher, 1985; Gernsbacher, Varner, & Faust, 1990; Magliano, Dijkstra, & Zwaan, 1996; Magliano et al., 2001; Zacks & Magliano, 2011). Here, I outline how Narrative Grammar, designed for sequential images, can be adapted to other domains.

7.1. Discourse and film

Just as panels in the graphic form window a scene, sentences allow discourse a way to package semantic information (Hoey, 1991; Langacker, 2001; Zwaan, 2004). Panels and sentences have been compared in various approaches applying discourse theory to graphic narrative (Saraceni, 2000, 2001; Stainbrook, 2003). Several sentences and phrases exemplify prototypical narrative categories: ''There once was an X…'' is a prototypical Establisher, providing a frame for referential entities to be established. ''And they all lived happily ever after'' is a prototypical Release, a generic aftermath appropriate for all happy endings. These phrases could be added to the beginning or end of nearly any narrative and retain their felicity, because of their status as narrative units.

Strings of sentences are used to combine multiple events in the same way that strings of panels are used in the graphic form. A discourse structure organizes the meanings in individual sentences with respect to each other. We can apply the present model of narrative grammar to discourse by verbally translating a comic, as in Fig. 21, which is a verbal version of Fig. 7.

This narrative conveys the same meaning as Fig. 7, in the same narrative structure: A series of Initials culminate with the Peak of the juggler being hit in the head. This phase serves as a Peak, which then resolves in the Release of him looking at the pin and throwing it away. It is important to note that sometimes a whole sentence can act as a category (as in the Release panel), but often a subclause alone can serve as a narrative category.

Narratives also appear in the visual modality through film. In film, cameras capture events as they unfold in time—just as in perception. This ongoing temporal progression records a single unbounded stream of events from one camera's viewpoint. Filmmakers then break up this recording into ''shots'' in the editing process, which combine with other shots to create a novel sequence in which a new temporality emerges, dictated by the shots themselves (Block, 2008; Bordwell & Thompson, 1997; Brown, 2002). Narrative roles are assigned during the process of recombining shots into a novel sequence. In fact, the filming and editing process often begins with ''storyboarding,'' where shots are drawn out in a form similar to the visual language used in comics. In other words, film uses the same narrative grammar, except the units are not static panels, but rather moving segments of film. The result is a hybrid: The narrative grammar organizes captured perceptual events in shots. Because film uses motion, this temporality can ''gloss over'' what in the static form would be individuated narrative units. A single shot may include both a preparatory action and primary event, thereby combining what statically would be discrete Initials and Peaks. In fact, because a camera can just be left recording, an entire action or even a full scene could be captured in one continuous shot, thereby concatenating what would be a whole Arc or more in drawn form. Complicating matters further, not only do the elements within a film shot move (i.e., characters and objects move around), but the camera itself can move. Panning and zooming create alterations to a graphic scene that is continuous rather than discrete.

These differences create an area of debate: Can non-discrete shots constitute ''narrative categories'' or not? Most definitely, a continuous single shot would show an *event* structure.
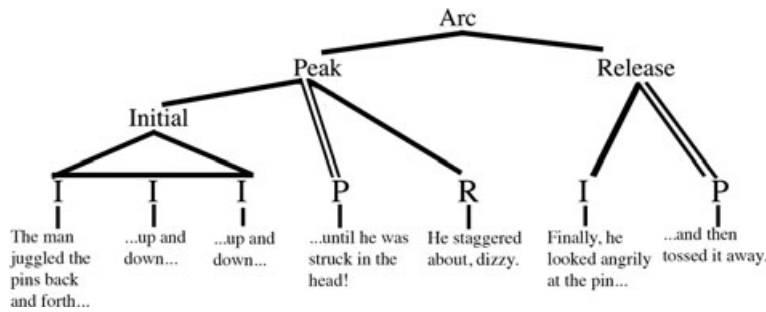

Fig. 21. Verbal discourse structure.

However, is narrative dependent upon discrete units that organize those potentially continuous events, or is a continuous representation merely a variation in ''performance'' but not ''competence''? Exploring such questions is important for cross-modality understandings of narrative.

7.2. Previous approaches

Although relatively little work has focused on narrative in sequential images, ample research has examined narratives in the verbal and filmic domain. Thus, it is worth comparing the theory of Narrative Grammar to previous research.

As mentioned earlier, previous approaches to narrative have differed in their treatment of the relationship of narrative (presentation) and events (meaning). In early research, the Russian Formalists quite explicitly separated the underlying events (''fabula'') from their presentation in narrative (''syuzhet'') (Tomashevsky, 1965). This tradition was maintained by Structuralist approaches in France (Genette, 1980) and America (Chatman, 1978), and has continued in theories of cognitive narratology (e.g., Bordwell, 1985; Herman, 2009a). On the other hand, theories of story grammar (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977) blurred the distinction between narrative and events or neglected it altogether (with some exceptions, e.g., Brewer & Lichtenstein, 1981). By and large, story grammar categories described the conveyance of events.

Approaches to narrative also differ in another respect. One line of research has focused on the semantic relationships between individual discourse units. A second line of research has used formal models of global schemas to describe narrative sequences. A third line has eschewed formal models, choosing instead to focus on general cognitive principles guided by inference and schemas. We take them up in turn.

7.2.1. Local theories

Several theories of narrative have focused on the pairwise semantic relationships between discourse units, adjacent or non-adjacent. In the verbal domain, the issue is how a sequence of sentences establishes meaningful continuity across a discourse. These theories either describe the technique used to create connections, such as using an anaphor to refer to something in a previous sentence (Halliday & Hasan, 1976), or characterize entire sentences' roles relative to each other, such as one sentence being an ''elaboration'' of another (Asher & Lascarides, 2003; Hobbs, 1985; Kehler, 2002; Mann & Thompson, 1987). Other approaches concentrate on the creation of causal inference throughout a narrative (Black & Bower, 1979; Trabasso, Secco, & van den Broek, 1984; Trabasso & van den Broek, 1985). These ''causal networks'' extend beyond just adjacent sentences to characterize relationships between pairs of sentences throughout a discourse. However, all of these approaches describe only the meaningful connections of individual sentences and do not establish global structure or constituency.

Similarly, McCloud's (1993) popular theory of ''panel transitions'' characterizes the semantic characteristics between two adjacent panels in terms of temporal change, shifts within and between characters and scenes, and completely non-sequitur relations. These

transitions operate through an inferential process that fills in the ''gap'' between images. Similar approaches have had a long-standing tradition in film theory. For example, Eisenstein's (1942) theory of ''montage'' argued that two film shots can unite to create a third inferred meaning, while Metz's (1974) ''grande syntagmatique'' outlined a taxonomy of relationships between film shots. Like McCloud's approach, these theories have largely characterized the semantic relationships between adjacent visual units.

A more recent theory focuses on the impact of local relationships on the actual processing of verbal and visual discourse. The *event-indexing model* (Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998) identifies five domains that readers actively monitor when reading a text: space, time, entities, motivation (i.e., characters' intentionality), and causation. If a text features a change in one of these domains, a ''processing shift'' marks the demand in comprehension that readers face as they integrate this new information into their working memory (Zwaan & Radvansky, 1998). Research has shown that such processing shifts can be detected in verbal discourse and films. Across several studies, research by Zacks and colleagues has shown that viewers can consciously identify the changes in characters, spatial location, and time between individual film shots (Magliano & Zacks, 2011; Magliano et al., 2001; Zacks et al., 2009). They appear to be most sensitive to changes between film shots depicting actions at one location and shots showing actions at another location (Magliano & Zacks, 2011). These findings echo McCloud's panel transitions.

As discussed previously, additional studies of filmed events have found that these types of shifts often co-occur with the boundaries of actual events. Zacks and colleagues (Zacks et al., 2009, 2010) have shown that changes in temporal, spatial, and causal coherence align with the end of one event and the start of another. They hypothesize that these changes are effective because viewers make predictions about what might occur next. These shifts confound those expectations by making a change in some semantic domain, thereby signaling the start of a new structure (Zacks & Magliano, 2011; Zacks et al., 2009, 2010).

How might these theories describe the structures in actual sequences of images? Return to the example in Fig. 1 of the man lying in bed with a clock on the wall. First, how would a linear approach, such as McCloud's panel transitions, handle such a sequence? If given only the local relationships between panels, the transitions would all be non-sequiturs, or at best random transitions through parts of a scene. What relationship does a man lying in bed have to a clock? The clocks to a window? A later time on a clock to a man making a phone call? Without a global view of the sequence, there would be no narrative here at all. These panels give no intrinsic cues about their narrative roles: Understanding comes entirely top-down from the global sequence.

This example illustrates the broader problems with theories based on local relationships. By only describing pairwise relationships between units (whether adjacent or as a network throughout a discourse), such approaches cannot account for groupings of units into constituents. However, it has long been shown that people intuitively agree upon where visual and verbal narratives divide into constituent segments (e.g., Gee & Grosjean, 1984; Gernsbacher, 1985; Mandler, 1987). Without this notion of structure, locally constrained approaches are unable to describe the type of embedding that this narrative grammar handles easily. For example, center-embedded phases require one sequence to stop for an aside and then

continue later on, all while maintaining a relationship to the broader sequence. Any approach that looks only at pairs of units has no way to express this relationship.

Nevertheless, while these locally constrained approaches cannot adequately describe the structure of sequential images on their own, they may provide valuable insights for particular aspects of narrative structure. As mentioned, the boundaries of narrative constituents often coincide with semantic changes between characters or locations, or the end of one event and the beginning of the next. These cues align well with McCloud's (1993) panel transitions and the event-indexing model's processing shifts (Zwaan & Radvansky, 1998; Zwaan et al., 1995). In this light, certain panel transitions may cue the crossing of narrative boundaries (Speer & Zacks, 2005). This would be analogous to the way that phrasal boundaries in verbal sentences serve as indicators of constituent structure (Fodor & Bever, 1965). Thus, continued study of these local relationships can complement — and be studied alongside — the theory of Narrative Grammar presented here.

7.2.2. *Global theories of narrative*

Other approaches have focused on how units play roles related to the broader structure of a narrative sequence. Observations about a global narrative schema have a long standing tradition in analyses of plotlines and storytelling, particularly for the theatre. Over 2,000 years ago, Aristotle described the *Beginning-Middle-End* schema for plays (Butcher, 1902), while in the 13th century, Zeami described similar structure for Japanese Noh drama (Yamazaki, 1984), and in the 19th century, Freytag (1894) outlined the contemporary notion of a narrative arc for five-act plays. More recently, a canonical schema emerged as central to theories of story grammars (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). These theories use narrative categories based around the achievement of a protagonist's goals, organized through specific phrase structures. For example, in Mandler and Johnson's (1977) well-known model, a canonical story structure involves the following rewrite rules (among others):

STORY $\rightarrow$ SETTING AND EVENT STRUCTURES

SETTING $\rightarrow$ {STATE* (AND EVENT*), EVENT*}

EVENT STRUCTURE $\rightarrow$ EPISODE ((THEN EPISODE)^n)

EPISODE $\rightarrow$ BEGINNING – CAUSE – DEVELOPMENT – CAUSE – ENDING

The first three rules state that a Story consists of a Setting — characters and environment — and Event structures. These Events consist of Episodes that use a canonical story structure organized around the achievement of a goal. These rules provide story grammars with a hierarchic generative grammar with numerous levels of comprehension.

Several experiments have supported story grammar's top-down approach to narrative, particularly with memory paradigms asking participants to recall written stories. Stories following the canonical story grammar episode structure were remembered with better accuracy than those with changes in temporal order (Mandler & Johnson, 1977), inversion of sentence order (Mandler, 1978, 1984; Mandler & DeForest, 1979), or fully scrambled sentences (Mandler, 1984). Recall also worsened in correlation with the degree to which a story rearranged the order of events (Stein & Nezworski, 1978): The further a structure departed from the canonical order, the harder it was to comprehend. However, children relied on this canonical structure more than adults. Adults recalled the surface structure of altered stories more accurately than children, who were more likely to reconstruct altered stories back into their canonical patterns (Mandler, 1978; Mandler & DeForest, 1979).

Outside of the theory outlined here, few global theories of static sequential images have been proposed. However, scholars have described sequences of film shots using grammatical models. For example, Carroll (1980) and Colin (1995) propose phrase structures to organize basic constituents of film shots, and Carroll (1980) and Buckland (2000) appeal to transformational rules to handle more complicated aspects of film sequencing. Although there is some diversity between these models, they differ from story grammars in that their rules do not detail a canonical narrative arc. Rather, they focus on how films order units of actions directly. However, in this way they are also like story grammars, as they conflate aspects of meaning (events) with those of structure, a point to which I return below.

How might a story grammar approach describe Fig. 1? The first panel might be considered some sort of Beginning or even an Initiating Event, while the final panel could possibly be considered an Outcome. However, such categorization is difficult given that story grammar categories are motivated by goal-directed events, and the goals in this sequence are vague. Also, how would the three central panels be described? Clocks and a window do not factor into the attempted achievement of goals. For a story grammar, these panels would be irrelevant or difficult to categorize. Thus, despite story grammars being solely based on top-down schemas, they cannot describe what is happening in Fig. 1, which is an example that *requires* top-down information to be understood as a *narrative*.

The issues faced by story grammars with Fig. 1 highlight the deeper problems with the theory. Because story grammars do not distinguish event structure and narrative structure, goal-driven behavior and meaningful aspects like ''setting'' and ''characters'' are folded into the ''narrative'' structure. This conflation has led to critiques that these models described semantic rather than truly structural relationships (de Beaugrande, 1982; Black & Wilensky, 1979). Subsequently, much psychological research on discourse comprehension has shifted toward studying the semantic aspects of discourse alone (e.g., Zwaan & Radvansky, 1998).[6]

However, it is important to mention that one approach concurrent with the story grammar tradition explicitly divides narrative and event structures. Brewer and colleagues (Brewer, 1985; Brewer & Lichtenstein, 1981, 1982; Ohtsuka & Brewer, 1992) have emphasized, as does the narrative grammar discussed here, that the separation of structures allows for different mappings between narrative and events. This research effectively characterizes how narratives create affective states based on how much information about event structure is

withheld or provided to a reader. Surprise narratives withhold critical information that is only reinterpreted, when the information is finally revealed. Suspenseful narratives provide an initiating event that causes a reader to be concerned about the outcome. Finally, narratives that provoke curiosity provide only enough information to let the reader know that something is missing. Brewer and colleagues do not formalize the mappings between narrative and events, and their approach does not include a way to discuss hierarchic structures, but the overall insights of their approach could be well integrated with the theory of Narrative Grammar.

Another limitation of the story grammar approach is the reliance on numerous levels of unique phrase structures instead of a generalizable recursive schema. Critiques of story grammars argued that these phrase structure rules do not adequately provide constraints to yield proper sequences (de Beaugrande, 1982), which led to widespread abandonment of hierarchic approaches altogether (Black & Bower, 1979; Trabasso et al., 1984; Zwaan & Radvansky, 1998). Furthermore, story grammars also do not allow for modifiers of base categories, as accomplished here by Prolongations or Conjunction, thereby leaving no room for elaborations on structure. By comparison, Narrative Grammar is flexible and internally recursive, requiring only a singular fundamental phase structure along with a generalized rule for repeating categories (Conjunction). Variations in the phase structure simply reflect the relationship between head and phase, not entirely new phase structures. This abstractness brings the combinatorial properties of narrative structure close to the ''X-bar'' schema underlying contemporary understandings of syntax (Culicover & Jackendoff, 2005; Jackendoff, 1977): The idea that a phrase is a generalized schema that is ''headed'' by one of its constituents.

These different views on narrative structure reflect, in part, the differences between linguistic models of syntax. Story grammars like Mandler and Johnson's (1977), and film grammars like Carroll's (1980), follow early approaches to transformational generative grammars (such as Chomsky, 1957, 1965) that used specific phrase structure rewrite rules to generate strings. Syntactic theory has changed dramatically since those times, particularly with the development of X-bar syntax (Jackendoff, 1977), which allowed for a multiplicity of phrase structures to be reduced to a single schema. Because story grammars quickly disappeared from the theoretical landscape due to criticism (de Beaugrande, 1982; Black & Wilensky, 1979; Garnham, 1983), they likely did not benefit from these historically concurrent advances in syntactic theory (not to mention subsequent innovations). Thus, in many ways Narrative Grammar picks up where story grammars left off — integrating more contemporary views of grammar (specifically Culicover & Jackendoff, 2005; Jackendoff, 2002) with those of narrative structure.

7.2.3. Cognitive narratology

Finally, the growing field of ''cognitive narratology'' in the humanities has analyzed narrative using general cognitive processes, such as frames, scripts, and schemas (for review, see Herman, 2003; Jahn, 2005). This broad approach takes story grammars as a precedent, but it does not propose a specific formulation for narrative sequencing. However, cognitive frames and scripts have been effectively applied to a wide range of narrative phenomena, particularly to generate inference. These analyses have included discussions of

comics directly (Bridgeman, 2004, 2005; Herman, 2009a,b, 2010; Lefèvre, 2000). This cognitive inferential approach for sequential images has owed a great deal to the film studies by Bordwell (1985, 2007), who has eschewed approaches that compare the structure of visual narrative to language (e.g., Metz, 1974). Instead, Bordwell focuses on how people draw upon scripts and schemas to create inferences while watching a movie. This involves a general conception of schemas described by story grammars, or possibly just the scripts of events (e.g., Minsky, 1975; Schank & Abelson, 1977). For example, people naturally must draw upon scripts about soccer to understand Fig. 3B, or about baseball to understand Fig. 18. Such a view is not incompatible with the approach here.

Nevertheless, such general processes alone are not enough. For example, theories of scripts generally do not factor in the interruptions of events (as in the Peaks of the majority of strips shown here), though these are important aspects of understanding both events and narratives. Furthermore, how would inferences alone describe the structural ambiguities in Fig. 1? First, is there a script involved with lying awake in bed before talking on the phone? Second, as was described, the first and last pairs of panels can be inferentially grouped into common environments. However, this is only one option; each panel could also depict its own unique time frame. How can a model without an explicit notion of constituency describe inferences that require grouping panels (as with the environmental Conjunction)? Also, how can it differentiate between one interpretation that requires localized inference, and another that requires a nested temporality? These phenomena require an explicit model of constituency, not simply general semantic relationships.

In Narrative Grammar, these *semantic* principles no doubt motivate inferences and possibly other facets of narrative. Yet this knowledge alone is not enough to describe the *structure* of visual narratives. These are complementary approaches, which both require rigorous elaboration as formal systems to fully understand them, independently and in combination.

8. Conclusion

This discussion began with the question of ''how do people make meaning out of sequential images?'' To answer it, this piece has outlined a theory of Narrative Grammar. In contrast to most recent approaches to discourse and narrative in the cognitive sciences, this model has emphasized separate narrative structures (presentation) from semantic structures (meaning). This separation allows us to describe how the same meaning can be conveyed in different surface presentations, as well as the opposite: how a single surface presentation can convey multiple meanings (as in Fig. 1). Formally, Narrative Grammar uses several core narrative categories that map to prototypical features of events or play functional roles related to the sequence. These categories are organized through a basic canonical phase rule similar to traditional notions of narrative. This canonical rule allows each category to also expand into its own phase, making the structure recursive. Along with a Conjunction rule, this recursion allows for complex narrative structures and pacing. Although we have focused on the visual-graphic domain, these structures permeate across film and verbal discourse as well. Exploring this permeability can offer us a better understanding of the way that

narrative changes to adapt to the unique properties of each modality, as well as the general function and structure of narrative as a modality general cognitive process.

Throughout, a broad analogy has been made between structure at the narrative level and syntax at the sentence level. This analogy has allowed us to apply techniques from linguistic analysis directly to properties of narrative in visual sequences. However, this analogy also raises concurrent questions about the structure and processing of visual narratives. Can we find evidence of a separation between structure and meaning in visual narrative processing? Can we find evidence of constituent structure? Do similar behavioral and neurocognitive responses appear to manipulations of narrative structure in sequential images as to manipulations of syntax in sentences?

Just as this analogy allows us to ask such questions, it can also provide a method for answering them. Guided by this theory, experimentation on structure and processing of sequential images can emulate paradigms for studying sentences. For example, in a recent study (Cohn et al., 2012), we replicated the paradigms of two classic psycholinguistic studies of sentence processing (Marslen-Wilson & Tyler, 1980; Van Petten & Kutas, 1991) using analogously constructed visual sequences based on this theory. Future research can draw from the wealth of previous research on language processing, as well as use the diagnostics outlined here as experimental manipulations. Thus, not only does the comparison between sentences and visual narratives provide a guide for theoretical study (as in linguistics), it can also provide methods for studying processing (as in psycholinguistics).

9. Graphic references

All images are created and copyright © 2012 Neil Cohn, except those cited throughout the text. Cited images are copyright their respective owners and used purely for analytical, critical, and scholarly purposes.

Aragonés, S. (2002). *Actions speak*. Milwaukie: Dark Horse Comics.

Godek, T. (2006).*One night*. Originally posted on March 20, 2006. Available at: http://www.yellowlight.scratchspace.net/comics/onenight/onenight.html.

Mahfood, J. (2002). Grrl Scouts in ''Just another day''. In D. Schutz (Ed.), *Dark Horse maverick: Happy endings*. Milwaukie: Dark Horse Comics.

Sakai, S. (1987). *UsagiYojimbo: Book one*. Seattle, WA: Fantagraphics Books.

Notes

1. Importantly, intuitions of felicity should not be mistaken for artistry or aesthetics. We should expect narratives in artistic contexts occasionally to push the limits of felicity. Similarly, in poetry, language that might otherwise be rejected as infelicitous is judged as acceptable because of the context. The present approach seeks only to describe the nature of comprehension; it makes no express claims on what is, or should be, considered of artistic or aesthetic merit.

2. Many of these categories should be reminiscent of notions from other models of narrative or discourse. For example, the canonical *E-I-P-R* sequence resembles the classic story structure for five-act plays proposed by Freytag (1894): *Set up-Rising Action-Climax-Falling Action-Denouement*. It also echoes Todorov's (1968) notion that narratives move from a state of equilibrium to disruption and back to equilibrium. More specifically, individual categories share traits with other approaches. For example, Establishers share qualities with discourse topics in verbal stories (Clark, 1996) or ''Settings'' in story grammars (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Nezworski, 1978; Thorndyke, 1977). Rather than describe all of the similarities of this approach to others for each category, a full listing is provided in Table 1.

3. The middle panel of a window could potentially group in several different patterns. For simplicity, I leave it isolated.

4. The actual contents of the balloon he used were a bit more risqué!

5. Both written and spoken forms will be subsumed here as ''verbal.'' For completeness, sign languages in the visual-manual modality should be included here as well.

6. Story grammars' methodological use of memory paradigms may have contributed to this conflation. Structure (syntax/phonology) is lost after comprehension, while only semantic information is retained in memory (van Dijk & Kintsch, 1983). Thus, if story grammars were based on recall measures, researchers may have unwittingly framed their experiments to pick up on semantic cues, not structural ones. Notably, after this criticism, some experiments using parsing techniques (Mandler, 1987), self-paced reading (Mandler & Goodman, 1982), and produced narration (Gee & Grosjean, 1984) offered psychological evidence for the story grammar hierarchies. However, these measures did not test the basic categories that comprised this hierarchy.

Acknowledgments

References

Arijon, D. (1976). *Grammar of the film language*. London: Focal Press.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge, England: Cambridge University Press.

de Beaugrande, R. (1982). The story of grammars and the grammar of stories. *Journal of Pragmatics*, *6*, 383–422.

Black, J. B., & Bower, G. H. (1979). Episodes as chunks in narrative memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 187–198.

Black, J. B., & Wilensky, R. (1979). An evaluation of story grammars. *Cognitive Science*, *3*, 213–230.

Block, B. (2008). *The visual story*. Oxford, England: Focal Press.

Bordwell, D. (1985). *Narration in the fiction film*. Madison: University of Wisconsin Press.

Bordwell, D. (2007). *Poetics of cinema*. New York: Routledge.

Bordwell, D., & Thompson, K. (1997). *Film art: An introduction* (5th ed.). New York: McGraw-Hill.

Branigan, E. (1992). *Narrative comprehension and film*. London: Routledge.

Brewer, W. F. (1985). The story schema: Universal and culture-specific properties. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning* (pp. 167–194). Cambridge, England: Cambridge University Press.

Brewer, W. F., & Lichtenstein, E. H. (1981). Event schemas, story schemas, and story grammars. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 363–379). Hillsdale, NJ: Erlbaum.

Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics*, *6*(5–6), 473–486.

Bridgeman, T. (2004). Keeping an eye on things: attention, tracking, and coherence-building. *Belphégor*, *4*(1), 1–18.

Bridgeman, T. (2005). Figuration and configuration: mapping imaginary worlds in BD. In C. Forsdick, L. Grove, & L. McQuillan (Eds.), *The Francophone Bande Desinée* (pp. 115–136). Amsterdam: Rodopi.

van den Broek, P. (1994). *Comprehension and memory of narrative texts: Inferences and coherence*. New York: Academic Press.

Brown, B. (2002). *Cinematography theory and practice*. Oxford, England: Focal Press.

Buckland, W. (2000). *The cognitive semiotics of film*. Cambridge, England: Cambridge University Press.

Butcher, S. H. (1902). *The poetics of Aristotle* (3rd ed.). London: Macmillian and Co. Ltd.

Carroll, J. M. (1980). *Toward a structural psychology of cinema*. The Hague, The Netherlands: Mouton

Chatman, S. (1978). *Story and discourse*. Ithaca, NY: Cornell University Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague, The Netherlands: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Cohn, N. (2003). *Early writings on visual language*. Carlsbad, CA: Emaki Productions.

Cohn, N. (2007). A visual lexicon. *Public Journal of Semiotics*, *1*(1), 53–84.

Cohn, N. (2010a). Japanese Visual Language: the structure of manga. In T. Johnson-Woods (Ed.), *Manga: An anthology of global and cultural perspectives* (pp. 187–203). New York: Continuum Books.

Cohn, N. (2010b). The limits of time and transitions: challenges to theories of sequential image comprehension. *Studies in Comics*, *1*(1), 127–147.

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

Cohn, N. (2012). *Structure, meaning, and constituency in visual narrative comprehension*. Doctoral Dissertation, Medford, MA: Tufts University.

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

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: structure and meaning in sequential image comprehension. *Cognitive Psychology*, *65*(1), 1–38. doi: 10.1016/j.cogpsych.2012.01.003

Colin, M. (1995). The grande syntagmatique revisited. In W. Buckland (Ed.), *The film spectator: From sign to mind* (pp. 45–86). Amsterdam: Amsterdam University Press.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford, England: Oxford University Press.

Culicover, P. W., & Jackendoff, R. (2010). Quantitative methods alone are not enough: Response to Gibson and Fedorenko. *Trends in Cognitive Sciences*, *14*(6), 234–235.

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

Duncan, R. (2000). Toward a theory of comic book communication. In K. Fudge & M. R. Lloyd (Eds.), *Academic forum* (Vol. 17, pp. 71–88). Arkadelphia, AR: Henderson State University.

Eisenstein, S. (1942). *Film sense* (J. Leyda, Trans.). New York: Harcourt, Brace World.

Eisner, W. (1985). *Comics & sequential art*. Tamarac, FL: Poorhouse Press.

Engelhardt, Y. (2002). *The language of graphics*. Doctoral Dissertation, Amsterdam: University of Amsterdam.

Engelhardt, Y. (2007). Syntactic structures in graphics. *Computational Visualistics and Picture Morphology*, *5*, 23–35.

Fodor, J., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*(5), 414–420.

Freytag, G. (1894). *Technique of the drama*. Chicago: S.C. Griggs & Company.

Garnham, A. (1983). What's wrong with story grammars? *Cognition*, *15*, 145–154.

Gee, J. P., & Grosjean, F. (1984). Empirical evidence for narrative structure. *Cognitive Science*, *8*, 59–85.

Gee, J. P., & Kegl, J. A. (1983). Narrative/story structure, pausing, and American Sign Language. *Discourse Processes*, *6*, 243–258.

Genette, G. (1980). *Narrative discourse* (J. E. Lewin, Trans.). Ithaca: Cornell University Press.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, *17*, 324–363.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Earlbaum.

Gernsbacher, M. A., Varner, K. R., & Faust, M. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 430–445.

Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, *14*(6), 233–234.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Victoria: Deakin University Press.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*, 512–521.

Herman, D. (2003). *Narrative theory and the cognitive sciences*. Stanford, CA: CSLI.

Herman, D. (2009a). *Basic elements of narrative*. West Sussex, UK: Wiley-Blackwell.

Herman, D. (2009b). Cognitive approaches to narrative analysis. In G. Brône & J. Vandaele (Eds.), *Cognitive poetics: Goals, gains and gaps* (pp. 30–43). New York: Walter de Gruyter.

Herman, D. (2010). Multimodal storytelling and identity construction in graphic narrative. In D. Schiffrin, A. D. Fina, & A. Nylund (Eds.), *Telling stories: Language, narrative, and social life* (pp. 195–210). Washington, DC: Georgetown University Press.

Hinds, J. (1976). *Aspects of Japanese discourse*. Tokyo: Kaitakusha Co., Ltd.

Hobbs, J. R. (1985). *On the coherence and structure of discourse*. Stanford, CA: CSLI Technical Report 85–37.

Hoey, M. (1991). *Patterns of lexis in text*. Oxford, England: Oxford University Press.

Jackendoff, R. (1977). *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, England: Oxford University Press.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*, 33–72.

Jacobs, J. (2001). The dimensions of topiccomment. *Linguistics*, *39*(4), 641–681.

Jahn, M. (1997). Frames, preferences, and the reading of third-person narratives: Towards a cognitive narratology. *Poetics Today*, *18*(4), 441–468.

Jahn, M. (2005). Cognitive narratology. In D. Herman, M. Jahn, & M.-L. Ryan (Eds.), *Routledge encyclopedia of narrative theory* (pp. 67–71). London: Routledge.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kraft, R. N., Cantor, P., & Gottdiener, C. (1991). The coherence of visual narratives. *Communication Research*, *18*(5), 601–616.

Kress, G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London: Routledge.

Krifka, M. (2007). Basic notions of information structure. In C. Féry, G. Fanselow, & M. Krifka (Eds.), *The notions of information structure* (pp. 13–54). Potsdam, Berlin: SFB.

Kunzle, D. (1973). *The history of the comic strip* (Vol. 1). Berkeley: University of California Press.

Labov, W. (1972). The transformation of experience in narrative syntax. In W. Labov (Ed.), *Language in the inner city* (pp. 354–396). Philadelphia: University of Pennsylvania Press.

Labov, W., & Waletzky, J. (1967). Narrative analysis: oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle: University of Washington Press.

Langacker, R. W. (2001). Discourse in cognitive grammar. *Cognitive Linguistics*, *12*(2), 143–188.

Lefèvre, P. (2000). Narration in comics. *Image [&] Narrative*, *1*(1).

Madden, M., & Abel, J. (2008). *Drawing words and writing pictures*. New York: First Second Books.

Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes*, *22*, 199–224.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, *15*, 533–545.

Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, *35*(8), 1489–1517.

Mandler, J. M. (1978). A code in the node: the use of story schema in retrieval. *Discourse Processes*, *1*, 14–35.

Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Mandler, J. M. (1987). On the psychological reality of story structure. *Discourse Processes*, *10*, 1–29.

Mandler, J. M., & DeForest, M. (1979). Is there more than one way to recall a story? *Child Development*, *50*, 886–889.

Mandler, J. M., & Goodman, M. S. (1982). On the psychological validity of story structure. *Journal of Verbal Learning and Verbal Behavior*, *21*, 507–523.

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

Mann, W. C., & Thompson, S. A. (1987). *Rhetorical structure theory: A theory of text organization*. Marina del Rey, CA: Information Sciences Institute.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.

McCloud, S. (1993). *Understanding comics: The invisible art*. New York: Harper Collins.

McCloud, S. (2006). *Making comics*. New York: Harper-Collins.

Metz, C. (1974). *Film language: A semiotics of the cinema* (M. Taylor, Trans.). New York: Oxford University Press.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*, 436–450.

Ohtsuka, K., & Brewer, W. F. (1992). Discourse organization in the comprehension of temporal order in narrative texts. *Discourse Processes*, *15*, 317–336.

Primus, B. (1993). Word order and information structure: a performance based account of topic positions and focus positions. In J. Jacobs (Ed.), *Handbuch syntax* (Vol. 1, pp. 880–895). Berlin and New York: de Gruyter.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding* (Vol. 211–236). New York: Academic Press.

Saraceni, M. (2000). *Language beyond language: Comics as verbo-visual texts*. Doctoral Dissertation, Nottingham: University of Nottingham.

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

Sasse, H.-J. (1987). The thetic/categorical distinction revisited. *Linguistics*, *25*, 511–580.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Sinclair, R. (2011). Christ, it works for everything. Available at: http://www.robertsinclair.net/comic/asshole.html. Accessed March 12, 2011

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*, 125–140.

Stainbrook, E. J. (2003). *Reading comics: A theoretical analysis of textuality and discourse in the comics medium*. Doctoral Dissertation, Indiana, PA: Indiana University of Pennsylvania.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing* (pp. 53–119). Norwood, NJ: Ablex.

Stein, N. L., & Nezworski, T. (1978). The effects of organization and instructional set on story memory. *Discourse Processes*, *1*(2), 177–193.

Sternberg, M. (1982). Proteus in quotation-land: Mimesis and the forms of reported discourse. *Poetics Today*, *3*(2), 107–156.

Sternberg, M. (2001). How narrativity makes a difference. *Narrative*, *9*(2), 115–122.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*, 49–100.

Talmy, L. (1995). Narrative structure in a cognitive framework. In G. Bruder, J. Duchan, & L. Hewitt (Eds.), *Deixis in narrative: A cognitive science perspective* (pp. 421–460). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Talmy, L. (2000a). *Toward a cognitive semantics* (Vol. 2). Cambridge, MA: MIT Press.

Talmy, L. (2000b). *Toward a cognitive semantics* (Vol. 1). Cambridge, MA: MIT Press.

Thorndyke, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*, 77–110.

Todorov, T. (1968). La Grammaire du récit. *Langages*, *12*, 94–102.

Tomashevsky, B. (1965). Thematics. In L. T. Lemon & M. I. Reis (Eds.), *Russian formalist criticism: Four essays* (pp. 61–95). Lincoln: University of Nebraska Press.

Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Erlbaum.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory and Cognition*, *19*, 95–112.

Yamazaki, M. (1984). *On the art of the no drama: The major treatises of Zeami* (J. T. Rimer, Trans.). Princeton: Princeton University Press.

Zacks, J. M., & Magliano, J. P. (2011). Film, narrative, and cognitive neuroscience. In D. P. Melcher & F. Bacci (Eds.), *Art and the senses* (pp. 435–454). New York: Oxford University Press.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307–327.

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*, 1–15.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). New York: Academic Press.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*, 292–297.

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