Generating Abstract Comics

Chris Martens and Rogelio E. Cardona-Rivera

North Carolina State University
{crmarten,recardon}@ncsu.edu

Abstract. We investigate a new approach to comic generation that explores the process of generating the contents of a panel given the contents of all previous panels. Our approach is based on leading discourse theories for comics by McCloud (panel transitions) and Cohn (narrative grammar), unified by cognitive theories of inference in visual language. We apply these theories to comics whose panel paremeters are abstract geometric shapes and their positions, contributing a computational realization of McCloud's and Cohn's comics theories, as well as a modular algorithm that affords further experimentation and evaluation of visual discourse theories.

Keywords: intelligent narrative technologies, comics, narrative generation

1 Introduction

Interactive digital storytelling has traditionally been concerned with text-based modes of discourse. Despite the fact that the word *fiction* makes no commitments to medium (book, film, play, et cetera), the term *interactive fiction* is used nearly synonymously with text-based digital storytelling. Meanwhile, comics, a relatively unexplored domain of computational narratology [12], present a wide range of expressive opportunities for interactive storytelling not afforded by text. Interactive comics invite many of the same questions addressed by interactive textual narrative research: what modes are available for a machine to tell visual stories in collaboration with a human player or author? How can algorithms introduce novel, generative content into a partially-constructed comic? Which decisions will the human and algorithmic components have available to them?

This work begins to address these questions by exploring generation of *purely visual*, *abstract* comics. "Purely visual" means that panels do not use words and text to communicate narrative, as in Figure 1. "Abstract," as in abstract art, means that, aside from repetition of elements with a shape, color, and size, and spatial placement of said entities within panels, we do not intentionally ascribe literal meaning to the components of generated panels. Our purpose in choosing an abstract representation is partially due to ease of implementation, but partially also to separate issues of *structure* of comics (which can be perceived with abstract representations) from issues of contextual, reference-laden meaning that would be found especially in anthropomorphic figures. (XXX cite)

While most approaches to text generation in the tradition of interactive storytelling research has focused on a "pipeline model" [21], first generating a plot structure and then developing a discourse after the fact, this is not the only available nor most suitable approach for visual narrative. Leading theories on comics have studied their sequence structure directly. McCloud [14] proposes a taxonomy of *transitions* between a panel and the one following it, enumerating six different ways readers might make sense of the connection between two consecutive panels "across the gutter" (gap between panels). Cohn [1], on the other hand, eschews transitions as a basis for comic structure and asserts that grammar-like syntax trees govern the formation of legible comics. We present a small-scale computational system [15] that operationalizes a hybrid of these two approaches.

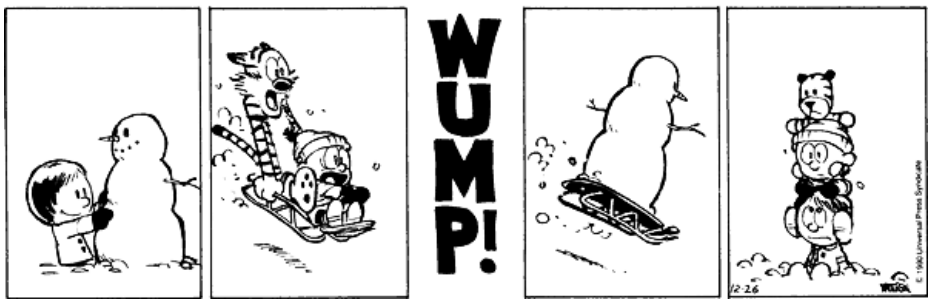

Fig. 1. This wordless Calvin and Hobbes comic strip (© Bill Watterson) exemplifies the kind of output we are targeting with our generator. The strip also illustrates how little plot structure informs this kind of short-form visual storytelling.

In the remainder of this paper, we discuss theoretical aspects of comics authoring, our computational implementation of a comic generation system, and our experience with refining our model with linguistic constraints. Our primary takeaway is that both *global and local* reasoning contribute important techniques to narrative generation: local reasoning is important for maintaining narrative coherence, and global reasoning is important for maintaining satisfying narrative structure. Both are thus important parts of creating comprehensible comics, and we present an outline of future work designed to explore the human interpretation of our generated artifacts.

2 Comics Theory

The basis of McCloud's theory about making sense of panel sequences across the gutters, later validated experimentally, is that readers of comics optimize their consumption of relevant information [20], and work to construct inferences [11] about story content in these liminal spaces of discourse. Inferences for story

content are constructed when they are needed for comprehension and enabled by what has been narrated thus far [17]. The dynamic between story authors and audiences parallels the dynamics of people engaged in cooperative conversation as outlined by the philosopher of language Grice [7]: the storyteller, as the active contributor to the ongoing communicative context, is expected to make her contributions to the discourse based on what is relevant to her narrative intent.

McCloud's introduced six *panel transition types* for comics [14], enumerate the different roles that the reader may infer from a well-written comic. These transition types are *moment-to-moment*, *action-to-action*, *subject-to-subject*, *aspect-to-aspect*, *scene-to-scene*, and non sequitur. While it is tempting to think we could simply operationalize these transitions in a generator, as Cohn [1] (Chapter 4) points out, so much of their meaning relies on contextual, real-world-situated understanding that it lends little help to computational authoring.

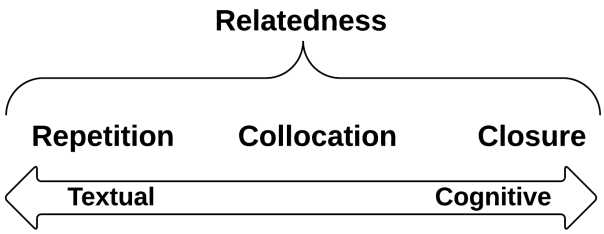

Fig. 2. The spectrum of *relatedness* as discussed by Saraceni [22]. Relatedness indicates how comic panels are connected or associated in the minds of readers, spanning from textual factors to cognitive factors. Along that spectrum, there are three distinguished categories of relatedness: *repetition*, *collocation*, and *closure*, which have demonstrably different effects on the construction of narrative mental models.

On the other hand, Saraceni [22] validated McCloud's hunch that readers create meaning from comics from perceived relationships between panels. Saraceni describes three notions of *relatedness* between comic elements, which are the building blocks from which readers may construct meaning inferences. Relatedness, a property of a comic that indicates how its panels are connected or associated, depends on a comic's *cohesion* – the lexico-grammatical features that tie panels together – and *coherence* – the audience's perception of how individual panels contribute to her mental model of the unfolding events. Relatedness emerges from a spectrum of *textual*[1] factors to *cognitive* factors, illustrated in Figure 2. Saraceni distinguishes three categories of relatedness. Closer to the textual end of the spectrum is the *repetition* of visual elements across panels. Beyond repetition is *collocation*, which refers to an audience's expectation that related visual elements

[1] *Textual* here does not mean the use of actual text, but rather is a shorthand for *surface code* [25].

will appear given the ones that have been perceived. Closer to the cognitive end of the spectrum is the *closure* over comic elements, which refers to the way our minds complete narrative material given to us. Closure is terminologically borrowed from the field of visual cognition, but is intended as the mental process of inference that occurs as part of an audience's *search for meaning* [5].

The comic in Figure 1 depends on these three aspects of relatedness: first, repetition of the sled, snowman, and other figures maintains cohesion across panels. Second, the humor of the sled carrying off the snowman depends on our (non-grammatical) domain knowledge that the snowman is not a living character in the same sense as the other figures. Finally, the comic depends on closure for the audience to "fill in the gaps" to infer what must have happened during the "WUMP!" panel: the sled maintained momentum to carry off the snowman, and the riders of the sled landed on top of the girl.

In our work we sought to develop a small scale computational model, and thus focused primarily on modeling discourse structure which lies on the textual side of the spectrum. However, our discourse model includes a minimal model of story, which is needed in order to account for some elements of the cognitive side of the spectrum: in particular, we assume chronological ordering between panels and track which visual elements have appeared previously in the panel sequence. We developed two compatible models of discourse structure: one based on McCloud's transition types and the other based on Cohn's [1] theory of visual language.

3 System Description

Our approach to generating visual narratives begins as a linear process that selects next comic panels based on the contents of previous panels, choosing randomly among indistinguishably-valid choices. The concepts we represent formally are *transitions*, *frames*, and *visual elements*, which we define below. There are two levels on which to make sense of these terms: the symbolic level, i.e. the intermediate, human-readable program data structures representing a comic, and the rendered level, designed to be consumed by human visual perception.

A **visual element (VE)** is a unique identifier from an infinite set, each of which is possible to map to a distinct visual representation. We do not explicitly tag visual elements with their roles in the narrative, such as characters, props, or scenery, making the symbolic representation agnostic to which of these narrative interpretations will apply. In the visual rendering, of course, our representation choices will influence readers' interpretation of VEs' narrative roles. A **frame** is a panel template; at the symbolic level, it includes an identifier or set of tags and a minimum number of required visual elements. The reason a frame specifies a minimum number of VEs is to allow for augmentation of the frame with pre-existing elements: for example, the *monologue* frame requires at least one visual element, indicating a single, central focal point, but other visual elements may be included as bystanding characters or scenery elements. At the rendering level, a frame includes instructions for where in the panel to place supplied VEs. A

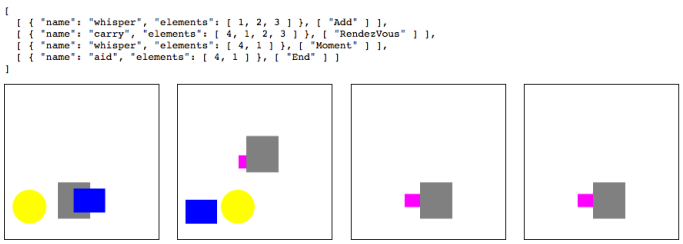

Fig. 3. Example of generator output. While the narrative here is ambiguous, we suggest the following readings: the repetition of the grey (largest) rectangle in every frame suggests it as a focal point, and the sudden appearance of the pink (smallest) rectangle suggests an interloper removing the grey rectangle from its initial context (established by the blue rectangle and yellow circle). Together with the names of the frames (reported in symbolic form above the comic), we can read the sequence as follows: the grey rectangle whispers to the blue rectangle, then is carried off by a pink rectangle, who whispers to the grey rectangle and then aids the grey rectangle.

panel is a frame instantiated by specific visual elements. Finally, a **transition** is a specification for how a panel should be formed as the next panel in a sequence. We took inspiration from McCloud transitions [14], developing a more syntactic notion defined purely in terms of frames and (abstract) visual elements, for which Saraceni's theory of relatedness [22] could be applied. For example, while McCloud could refer to an action-to-action transition as one where a character is depicted carrying out two distinct actions, we have no notion of *character* and *action* (these being semantic and contextual categories), so instead must refer to which visual elements appear, where they have appeared previously, and what their spatial relationships might be (potential frames). The rendering of a frame itself may position VEs in such a way that an audience would read certain actions or meaning into it; however, this kind of audience interpretation is not modeled to inform generation.

We introduce six formal transition types: *moment*, *add*, *subtract*, *meanwhile*, and *rendez-vous*, each of which specifies how a next panel should be constructed given the prior sequence. **Moment** transitions retain the same set of VEs as the previous panel, changing only the frame. **Add** transitions introduce a VE that didn't appear in the previous panel, but might have appeared earlier (or might be completely new). A new frame may be selected. **Subtract** transitions remove a VE from the previous panel and potentially choose a new frame. **Meanwhile** transitions select a new frame and show *only* VEs that did not appear in the previous panel, potentially generating new VEs. **Rendez-vous** transitions select a random subset of previously-appearing VEs and selects a new frame to accommodate them.

We implemented our generator in OCaml and additionally implemented a front-end, a web-based renderer (not linked here for anonymous review). The renderer assigns each frame type to a set of coordinates given by percentage of the vertical and horizontal panel size, and then renders panels by placing visual

elements at those coordinates. Visual elements are represented by randomly generated combinations of size, shape (circle or rectangle), and color. An example of the generator's output can be seen in Figure 3. The generator accepts as inputs length constraints (minimum and maximum) and a number of VEs to start with in the first panel. Its output is a sequence of panels (frame names and VE sets) together with a record of the transitions that connect them.

Generating random transition sequences may result in nonsensical output, such as ending a comic with a *meanwhile* frame in which completely new visual elements are introduced at the end of the comic, but not connected to previous elements; see Figure 4 for an example. To constrain output, we reached for work by Neil Cohn [3] and his colleagues on the linguistic structure of visual narratives. They claim that understandable comics follow a grammar that organizes its global structure. Instead of transition types, Cohn's grammar of comics consists of grammatical categories (analogous to nouns, verbs, and so on) indicating the role that each panel plays in the narrative. These categories are **establisher**, **initial**, **prolongation**, **peak**, and **release**, which allow the formation of standard narrative patterns including the Western dramatic arc of *initial – peak – release*. Formally, Cohn gives the following grammar as a general template for comic "sentences," or well-formed arcs:

$$\textit{(Establisher)} - \textit{(Initial (Prolongation))} - \textit{Peak} - \textit{(Release)}\ ^{2}$$

In our second iteration of the generator, we combine two approaches to discourse, using *global* Cohn grammars to guide the *local* selection of syntactically-defined transitions. In particular, we enumerate every possible category bigram in Cohn's grammar, such as *initial to prolongation*, *prolongation to peak*, and so on, and describe sets of transition types that could plausibly model the relationship. This mapping is given below:

[2] Symbols in parentheses are optional. In our expression of this grammar (and in several of Cohn's examples), we also assume that prolongations may occur arbitrarily many times in sequence.

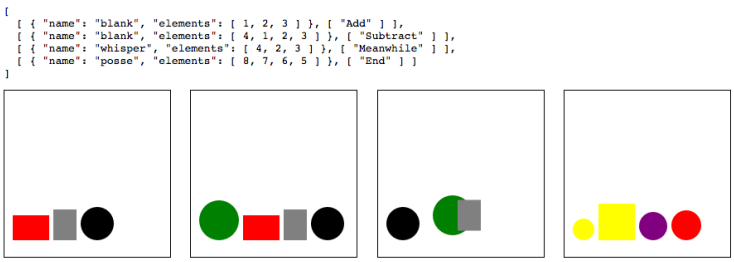

Fig. 4. Example of underconstrained output. The final panel does not maintain relatedness to the preceding sequence.

```
{
  "sequence": [ [ "Establisher" ], [ "Initial" ], [ "Peak" ], [ "Release" ] ],
  "comic": [
    [ { "name": "blank", "elements": [ 1, 2, 3, 4 ] }, [ "Moment" ] ],
    [ { "name": "whisper", "elements": [ 1, 2, 3, 4 ] }, [ "Meanwhile" ] ],
    [ { "name": "dialog", "elements": [ 6, 5 ] }, [ "RendezVous" ] ],
    [ { "name": "aid", "elements": [ 4, 3, 1, 2 ] }, [ "End" ] ]
  ]
}
```


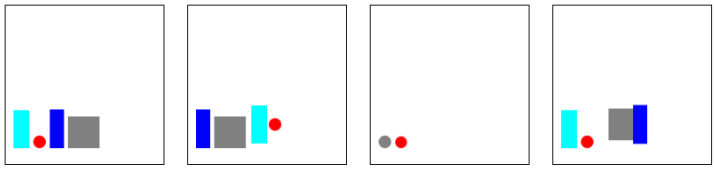

Fig. 5. Example of grammatically-constrained output. This example shows a common pattern in grammatically-constrained output, introducing a new visual element with a Meanwhile transition for the peak, then releasing with a Rendez-vous.

Establisher	Initial	{Moment, Subtract, Add, RendezVous}
Establisher	Prolongation	{Moment, Subtract, Add}
Establisher	Peak	{Add, Meanwhile}
Initial	Prolongation	{Moment, Subtract, Add}
Prolongation	Prolongation	{Moment, Subtract, Add}
Prolongation	Peak	{Subtract, Add, RendezVous}
Initial	Peak	{Subtract, Add, Meanwhile, RendezVous}
Peak	Release	{Subtract, Add, RendezVous}

With this mapping established[3], we randomly generate an instance of the arc grammar and populate it with an appropriate set of transitions, after which point we simply hook the transition sequence up to the same panel selector from before. An example of the constrained generator's output can be found in Figure 5.

4 Related Work

Montfort et al. [16] developed a blackboard architecture called Slant for textual story generation, which primarily specifies and refines plot structure. However, Slant includes a sub-component called Verso, which reasons over narrative discourse as a way to further constrain the narrative plot. Verso detects aspects of the verbs used during the generation of plot structure, and determines the

[3] This mapping is guided by our intuition rather than systematic symbolic reasoning. We aim to rule out obvious-seeming syntactic errors, e.g. a *meanwhile* transition at the end of an arc, but other constraints are not so easily expressed. For instance, perhaps a *prolongation* should be realized as a repetition of the previous transition, but this information is not available in the bigram model. In future work, we would like to refine the theoretical grounding of the relationship between transitions and grammatical categories.

in-progress story's match to a specific genre. Slant is thus not strictly a pipeline model architecture, but the constraints identified during discourse reasoning cannot themselves inform further discourse reasoning. In our approach, discourse reasoning constrains plot structure generation, and can potentially inform further narrative discourse generation.

Pérez y Pérez et al. [18] developed a visual illustrator to the MEXICA system [19], and verified the degree to which their 3-panel comic generator elicited in readers the same sense of story as a textual realization of the same MEXICA-generated plot. While this system still follows the pipeline model of narrative generation, we see their work as complementary: they developed an experiment methodology through which it is possible to empirically assess if their palette of designed visual elements denote story concepts as intended. Future work in discourse-driven comic generation will have to address this point going forward, and Pérez y Pérez et al. provide a step toward understanding the gap between story concepts and the pictorial symbols meant to encode them.

(XXX) Jhala and Young [10] - cinematic visual discourse

5 Future Work

Future directions for this project include expanding the set of visual elements beyond abstract, geometric shapes (one candidate being the modular XKCD sprite set (XXX ref)). We would also like to incorporate generation into an interactive framework, either for the purpose of interactive visual storytelling or mixed-initiative comics design.

We have several potential evaluation plans, each investigating distinct hypotheses about our approach. One candidate involves analyzing the style and variety of our comic generator's output; i.e. our system's *expressive range* [23]. For this, and as suggested by Smith and Whitehead, we would need to identify appropriate metrics for describing the generated output, which "should be based on global properties ... and ideally should be emergent qualities from the point of view of the generator." A textually-focused candidate metric is the number and type of transitions that are generated on average in a large sample of generated comics. A cognitively-focused candidate metric is the average number of unique readings that an audience comes up with for generated comics. Further, these metrics should be evaluated in the context of the discourse grammar's *cyclomatic complexity* [13], which in our case is low; such an analysis will yield insight into the representational power that the grammar has for generating narrative discourse, relative to the system's overall computational complexity. Another candidate evaluation involves analyzing the level of comprehension that our generated comics afford an audience. While there has been work in understanding how people read into narratives involving abstract shapes (e.g. the Heider-Simmel experiment [8]), this evaluation would be more concerned with whether the discourse categories (as discussed by Cohn) that guide the selection of transitions are recognizable by an audience during comprehension. Cohn [2] discusses a methodology through which panel discourse categories can be analytically identified; this analysis would

ask whether comic panel categories can be analytically identified by an audience when they are intentionally selected by our generative system.

6 Conclusion

In this work we have presented an approach to comic generation combining local, transition-driven decisions with global syntactic structure. We initially designed our system to tend only to textual factors in comic discourse: the repetition of comic actants across the narrative provides a minimal cohesive backbone on which to pin comic understanding. However, as discussed, this form of generation could generate non-sensical output (e.g. ending comics with a *meanwhile* discourse transition). We therefore appealed to more cognitively-oriented factors via the theory of visual grammar, which helped structure the output in a way that enables other senses of relatedness to contribute to the output's coherence. Thus, through our small-scale system, we have begun to explore the scale and limits of human story sense-making faculties, as well as how they come to bear on narrative generation systems: in our case, through both local and global procedures, which inform cohesion and coherence, respectively. Our algorithms and implementation offer a promising starting point for the computational investigation of discourse-driven narrative.

More broadly, our work highlights the importance of looking to human cognition as a point of departure for the design of narrative generators. Other scholars (e.g. Gervas [6] and Szilas [24]) have argued the same point; our system provides a computational system that demonstrates it. Concretely, the reason for this is that humans bring significant cognitive faculties to bear on the process of narrative comprehension [9]. An instance of this narrative intelligence is our unique ability to fill in the blanks in the liminal spaces of discourse, which (at least) relies on our focalized perspectives into the story world [4]. As our generated comics show, our narrative sense-making abilities allow us to intuit and impose narrative structure on the sequence of depicted images, due to how we fill in the blanks left unspecified in our comics. Therefore, this mental process has a significant role in our appreciation of the narrative artifact, and should have an equally significant role in the generation of it.

References

1. Cohn, N.: The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images. Bloomsbury, London, England, UK (2013)
2. Cohn, N.: Narrative conjunction's junction function: The interface of narrative grammar and semantics in sequential images. Journal of Pragmatics 88, 105–132 (2015)
3. Cohn, N. (ed.): The Visual Narrative Reader. Bloomsbury Publishing (2016)
4. Genette, G.: Narrative Discourse: An Essay in Method. Cornell University Press (1983)
5. Gerrig, R.J., Bernardo, A.B.I.: Readers as problem-solvers in the experience of suspense. Poetics 22(6), 459–472 (1994)

6. Gervás, P.: Computational Approaches to Storytelling and Creativity. AI Magazine 30(3), 49–62 (2009)
7. Grice, H.P.: Logic and conversation. In: Cole, P., Morgan, J.L. (eds.) Syntax and semantics 3: speech arts. Elsevier (1975)
8. Heider, F., Simmel, M.: An experimental study of apparent behavior. The American Journal of Psychology 57(2), 243–259 (1944)
9. Herman, D.: Storytelling and the Sciences of Mind. MIT Press (2013)
10. Jhala, A., Young, R.M.: Cinematic visual discourse: Representation, generation, and evaluation. IEEE Transactions on Computational Intelligence and AI in Games 2(2), 69–81 (2010)
11. Magliano, J.P., Kopp, K., Higgs, K., Rapp, D.N.: Filling in the gaps: Memory implications for inferring missing content in graphic narratives. Discourse Processes (2016)
12. Mani, I.: Computational modeling of narrative. Synthesis Lectures on Human Language Technologies 5(3), 1–142 (2012)
13. McCabe, T.J.: A complexity measure. IEEE Transactions on Software Engineering 2(4), 308–320 (1976)
14. McCloud, S.: Understanding Comics: The Invisible Art. Harper Collins, New York, NY, USA (1993)
15. Montfort, N., Fedorova, N.: Small-Scale Systems and Computational Creativity. In: Proceedings of the 3rd International Conference on Computational Creativity. pp. 82–86 (2012)
16. Montfort, N., Pérez, R., Harrell, D.F., Campana, A.: Slant: A blackboard system to generate plot, figuration, and narrative discourse aspects of stories. In: Proceedings of the 4th International Conference on Computational Creativity. pp. 168–175 (2013)
17. Myers, J.L., Shinjo, M., Duffy, S.A.: Degree of causal relatedness and memory. Journal of Memory and Language 26(4), 453–465 (1987)
18. Pérez y Pérez, R., Morales, N., Rodríguez, L.: Illustrating a computer generated narrative. In: Proceedings of the 3rd International Conference on Computational Creativity. pp. 103–110 (2012)
19. Pérez y Pérez, R., Sharples, M.: MEXICA: A computer model of a cognitive account of creative writing. Journal of Experimental and Theoretical Artificial Intelligence 13(2), 119–139 (2001)
20. Pirolli, P.: Information Foraging Theory: Adaptive Interaction with Information. Oxford University Press (2007)
21. Ronfard, R., Szilas, N.: Where story and media meet: computer generation of narrative discourse. In: Proceedings of the 5th Workshop on Computational Models of Narrative. pp. 164–176 (2014)
22. Saraceni, M.: Relatedness: Aspects of textual connectivity in comics. In: Cohn, N. (ed.) The Visual Narrative Reader, chap. 5, pp. 115–128. Bloomsbury (2016)
23. Smith, G., Whitehead, J.: Analyzing the Expressive Range of a Level Generator. In: Proceedings of the 2010 Workshop on Procedural Content Generation in Games at the 5th Interational Conference on the Foundations of Digital Games (2010)
24. Szilas, N.: Requirements for Computational Models of Interactive Narrative. AAAI Fall Symposium on Computational Models of Narrative (2010)
25. Zwaan, R.A., Radvansky, G.A.: Situation models in language comprehension and memory. Psychological Bulletin 123(2), 162–85 (Mar 1998)