Shall I Compare Thee to Another Story?—An Empirical Study of Analogy-Based Story Generation

Jichen Zhu and Santiago Ontañón

***Abstract*—Despite their use in traditional storytelling, analogy-based narrative devices have not been sufficiently explored in computational narrative. In this paper, we present our analogy-based story generation (ASG) approach in the Riu system, focusing on analogical retrieval and projection. We report on an empirical user evaluation about Riu's capability to retrieve and generate short noninteractive stories using the story analogies through mapping (SAM) algorithm. This work provides the foundation for exploration of ASG in more complex and interactive computational narrative works.**

***Index Terms*—Computational analogy, empirical evaluation, force dynamics (FD), story generation.**

I. Introduction

NARRATIVE, as it evolves with technological developments, constantly reinvents itself to better capture individuals' experience and new social orders. Over the past decades, an increasing number of computational narrative artifacts have been developed in various areas, such as entertainment (e.g., computer games), training and education (scenario-based training simulation), or artistic expression (electronic literature). Similar to the history of film and many other traditional media, unleashing the full potential of computational narrative requires a close collaboration between technical innovation and expressive exploration. Among others, developments in AI research provide new possibilities for enhancing storytelling with user interaction, personalization, procedurally generated content, etc. These elements help to reshape the boundary and the poetics of computational narrative, whether the resulting narrative closely remediates [1] traditional forms of stories [2]–[5] or evolves into something radically different [6], [7].

Story generation, one of the active research areas in computational narrative, has made considerable progress in recent decades, notably in planning-based approaches [8]–[10] and multiagent simulation-based ones [11]. New algorithmic improvements, often aided by narratology theories, have allowed computer systems to produce increasingly complex stories. Although there are several exceptions [12], [13], most computer-generated stories occupy a very similar space in the expressive spectrum defined by traditional narrative. Elsewhere, we have observed that different story generation techniques have specific built-in narrative affordances and constraints [14]. Stories generated using planning, for instance, often embody a strong action-based, goal-driven aesthetic. We believe that the long-term success of computational narrative lies in whether it can communicate both the breadth and depth of being-in-time of human experiences. Although planning-based story generation is an important direction, the further development of computational narrative as a mature form of cultural expression calls for broadening its expressive range by exploring alternative technical approaches.

In this paper, we present our work toward this goal in a relatively under-explored area—analogy-based story generation (ASG). In traditional forms of narrative, metaphor, simile, free association, parallel narrative, and other similarity-based narrative devices are commonly used. For instance, metaphors (e.g., "Juliet is the sun") and similes (e.g., "My love is like a red, red rose," or "Shall I compare thee to a summer's day?") frequently appear in literature, especially in poetry. At a coarser level of granularity, free association has been used, notably by stream of consciousness writers such as Joyce and Woolf, as a means of depicting characters' trains of thought. In these works, a character's thoughts do not always follow the logical cause-and-effect order; rather, they shift fluidly, from one topic to another, with similar elements/traits as the bridge. At the plot level, parallel narratives, sometimes called tandem narratives [15] or double-scope stories [16], can be used to connect seemingly separate stories through a common theme. For example, in Guillermo del Toro's film *Pan's Labyrinth* (2006), the narrative shifts between the reality and the protagonist's imaginary fantasy world. The two highly contrasting worlds are intertwined through similar events such as a near escape, although these events were carried out by different characters in different settings. In prose fiction, Haruki Murakami masterfully used a similar technique in his novel *Hard-Boiled Wonderland and the End of the World* [17].

In the above examples, similarity-based narrative devices offer an alternative to the cause-and-effect-based narrative world of actions and well-defined character goals. What is foregrounded here are the characters' rich inner worlds as well as the authors' subjectivity. In order to explore these narrative possibilities, we need a set of computational tools capable of establishing appropriate associations between different narrative elements. In this paper, we explore and evaluate the use of computational analogy, particularly with different types of domain knowledge, for noninteractive short story snippets.

Manuscript received October 06, 2012; revised March 31, 2013; accepted June 10, 2013. Date of publication July 30, 2013; date of current version June 12, 2014.

J. Zhu is with the Digital Media Program, Antoinette Westphal College of Media Arts & Design, Drexel University, Philadelphia, PA 19104 USA (e-mail: jichen.zhu@drexel.edu).

S. Ontañón is with the Computer Science Department, Drexel University, Philadelphia, PA 19104 USA (e-mail: santi@cs.drexel.edu).

Digital Object Identifier 10.1109/TCIAIG.2013.2275165

The long-term goal of our research is to broaden the expressive range of computational narrative by exploring different story generation techniques. In this paper, we present our approach for analogy-based story generation through our interactive narrative system Riu, and particularly its ASG components: memory retrieval and the story analogies through mapping (SAM) [18] analogical projection algorithm, which completes partially specified stories by analogy. As ASG is a relatively new direction for computational narrative, the focus of this paper is on developing and evaluating the technical foundations necessary for our long-term goal. Specifically, we focus on the story representation formalism as well as how Riu utilizes analogy in its analogical retrieval and projection processes. This paper extends our prior work by empirically evaluating the major hypotheses of our system design and the effectiveness of our system. Our user study confirms that: 1) the assessment of story similarity and analogical mappings used in our system aligns with human readers' perceptions of them; 2) the force-dynamics-based story representation contributes significantly to the performance of our system; and 3) the quality of the stories generated by our system is high. We believe that the results presented in this paper provide us with a solid foundation for further investigating narrative aesthetics and user interactivity in ASG.

This paper is organized as follows. First, we provide a theoretical framework on computational analogy and the cognitive semantic theory of force dynamics (FD), the basis of our story representation. Then, we introduce Riu, focusing on the SAM algorithm, with examples of generated stories. Next, we present our user study and analyze the results. Finally, we compare Riu with other related ASG systems.

II. Theoretical Framework

This section presents the theoretical framework foundational to our approach. Related ASG systems are discussed in relation to our Riu system in Section V.

A. Computational Analogy and Structure Mapping

Drawn upon the human cognitive process of analogy making, computational analogy operates by identifying similarities and transferring knowledge between a source domain S and a target domain T. The intuitive assumption behind analogy is that if two domains are similar in certain key aspects, they are likely to be similar in other aspects. Given a target domain T, this process is composed of four stages [19]:

1) recognition of a candidate analogous source S;
2) elaboration of an analogical mapping and inferences between source domain S and target domain T;
3) evaluation of the mapping and inferences;
4) consolidation of the outcome for other contexts.

Existing systems that implement some or all of these stages can be classified into three classes based on their architecture [20]. Symbolic models (e.g., ANALOGY [21] and the structure-mapping engine (SME) [22]) heavily rely on concepts from the "symbolic AI paradigm": symbols, logics, planning, search, means-ends analysis, etc. Connectionist models (e.g., ACME [23], LISA [24], and CAB [25]), on the other hand, adopt the connectionist framework of nodes, weights, spreading activations, etc. Finally, the hybrid models (e.g., COPYCAT [26], TABLETOP [27], and LETTER–SPIRIT [28]) blend elements from the previous two classes.

Of particular relevance here is the SME algorithm [22]. Its cognitive foundation is Gentner's structure-mapping theory on the implicit biases and constraints by which humans interpret analogy and similarity [29]. Built upon psychological evidences, Gentner's central idea is that human analogical reasoning favors the relations between entities, rather than their surface features. SME implements this view of analogical reasoning as a structure-preserving process. Focusing exclusively on the elaboration stage of analogy, SME receives two domains as input, each represented as a series of entities and relations, and outputs an analogical mapping of the entities and relations between the domains. In Riu, we use SME as the analogy-mapping component [18], however, other algorithms could also be used.

Despite the psychological plausibility of structure-mapping theory, critics often point out that SME is very sensitive to the representation formalism being used [30]. For this reason, we based the story representation in our system on an established cognitive semantic model, namely, FD.

B. Force Dynamics

FD is a semantic category defined by cognitive linguist Leonard Talmy [31]. It is based on the observation that a wide range of human linguistic and cognitive concepts are understood by considering them as if they were physical forces. When representing the semantics of a given sentence or situation, FD captures fundamental structures such as "the exertion of force, resistance to such a force, the overcoming of such a resistance, blockage of the expression of force, removal of such blockage, and the like" [31, p. 409]. Some of these constructs are key to narratives, and are hard to represent using the traditional notions of causality.

A basic FD pattern contains two entities: an agonist (the focal entity) and an antagonist, exerting force on each other. An agonist has a tendency toward either motion/action or rest/inaction. It can only manifest its tendency if it is stronger than the opposing antagonist. For example, to represent "The ball kept rolling because of the wind blowing on it," the agonist's (ball) intrinsic tendency toward rest is overcome by the antagonist's (wind) greater force, and hence the result is the motion of the agonist. In other FD structures, the antagonist can function as a facilitator and help the agonist. At the temporal level, Talmy uses the concept of phase to describe the interaction between agonist and antagonist at a particular point in time. A story, therefore, can be represented as a sequence of phases.

Important to narratives, FD describes not only physical forces, but also psychological and social interactions. Conceiving such interactions as psychological "pressure," FD patterns can manifest themselves in various semantic configurations, such as the "divided self" (e.g., "He held himself from responding") and complex social interactions (e.g., "She gets to go to the park"). Additionally, certain linguistic structures are force-dynamically neutral (e.g., "He did not respond").

To illustrate how we use FD to represent stories, consider the following example: "Ales always wanted to be a painter, despite his long working hours. But his job got more demanding, and he eventually gave up his practice." Here, the agonist is "Ales" and the antagonist is his "job." The story may be divided into two phases. First, the agonist has the tendency to move, and he is stronger than the antagonist. In the second phase, their relative force strength shifts—the antagonist strengthens and sets the agonist at rest. Details of how FD is integrated into Riu's story representation formalism are discussed in Section III-B.

As we argued elsewhere [32], FD can enhance existing story representations in two main ways. First, it can express complex relations such as "hindering," "helping," and "leaving alone," some of which are hard to represent by planing-based representations. Second, FD's level of abstraction helps SME find better analogies (as discussed in our study in Section IV).

III. SAM AND THE RIU SYSTEM

This section presents Riu, focusing on its story representation and on its ASG components: memory retrieval and SAM.

A. Riu

Riu is a text-based interactive narrative system designed to explore the connection between an external story world and the character's inner world of memories and imagination. So far, computational analogy-based techniques have been mostly used to exploit associations between individual self-contained stories, as in MINSTREL [33] and the Story Translator [34]. Within each story, however, analogical connections play a limited role, if any. Part of Riu's goal is to explore analogy both as the generative technique and as the storytelling device in one story.

Motivated by the aforementioned literary examples, Riu focuses on the analogical connection between characters' external and inner worlds. We explore how to procedurally allow the two worlds to intersect and influence one another. Our most recent interactive story world, *Evening Tide*, is about the last diving expedition of marine biologist Julian Champagne. In the full story, the player character's actions can trigger related memories, analogous to the current state, and cause the generation of narratives about his inner thoughts (using SAM). These thoughts can, in turn, affect the actions available in the external world. In other words, ASG is used to create an intertwined parallel narrative structure, connected via analogy. A sample interaction with Riu can be found in [18].

The high-level architecture for Riu is diagrammed in Fig. 1. The story engine component interprets user input, and coordinates the other components in the system. The memory retrieval component identifies memories similar to a given scene, typically the current state of the story. The imaginative projection component uses the SAM algorithm to generate stories by analogically transferring knowledge from source to target domain. Finally, the character model uses the previous two modules to determine the main character's behavior, given the user's input. Riu has two types of preauthored content: memories and main story graph, which define the character's inner world and the external story world, respectively.

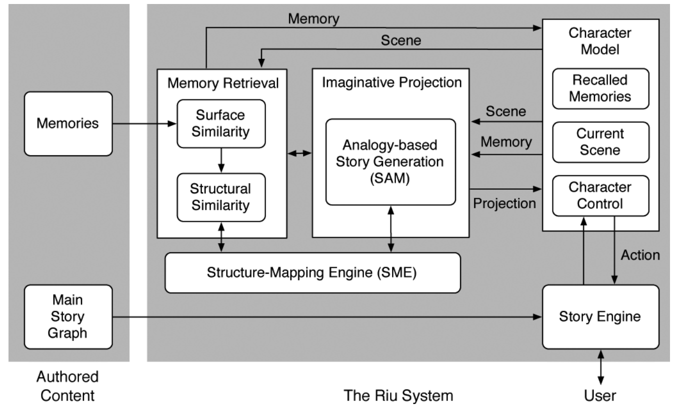

Fig. 1. High-level architecture of the Riu system.

Computational analogy is used in two major aspects of Riu: memory retrieval and imaginative projection. Memory retrieval is based on the recognition stage of computational analogy (Section II-A). When Riu needs a memory related to the current situation, it searches a repository of preauthored memories and retrieves the most relevant one. Once retrieved, the memory becomes part of the character model and influences her disposition toward the world. The imaginative projection process is based on the elaboration stage of analogy. In this process, SAM uses analogical projection to infer the consequences of a particular user-selected action in the story world by transferring knowledge from one of the recalled memories. In other words, the system generates a story of the possible consequences of the action, which might influence the behavior of the main character.

In order to evaluate the effectiveness of using ASG to make analogical connections and generate stories, the focus of this paper, we divided the full *Evening Tide* story into smaller, noninteractive snippets. In our user evaluation, we further disassociate them by changing the protagonist to a different one in each story. By doing so, we intend to test the intrinsic analogical connections between these stories. It is important to keep in mind that, in Riu, these story snippets are integrated into a single interactive piece, highlighting the analogy-based interconnection between the main story world and the player character's inner world. Further evaluation of the entire story with user interaction will be conducted in our future work.

B. Story Representation

As observed by others [30] and confirmed in our previous work [35], the choice of representation formalism has a substantial impact on computational analogy. In Riu, the basic story representation element is a scene. It is a small encapsulated piece of story, typically involving one main character in a single location. In Riu, each memory is a scene and the main story is represented as a graph where each node corresponds to a scene and each directional link represents a user action that triggers the next scene from the current one. A scene is composed of a series of phases. As in FD, a phase represents the state of the scene in a particular point in time. The sequence of phases represents the temporal development. Each scene is represented in two parallel parts: a computer-understandable description and a human-understandable description.

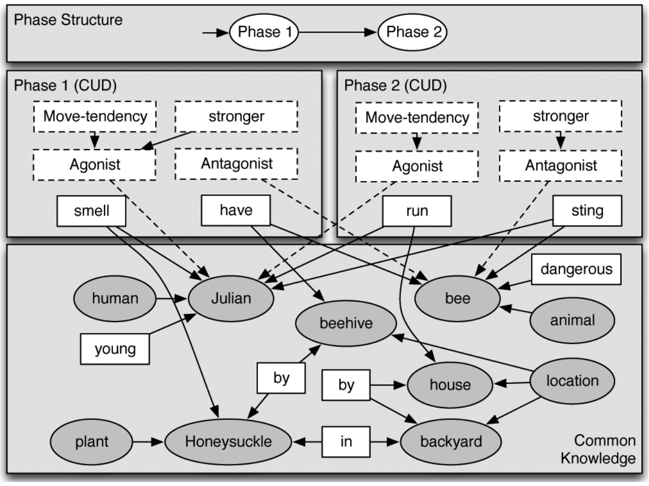

Fig. 2. CUD of a scene in Riu.

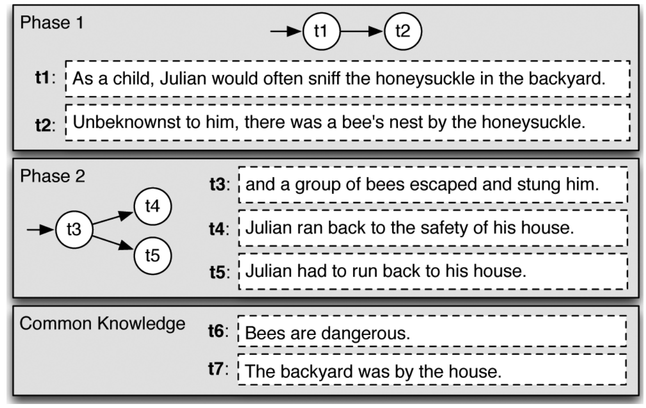

Fig. 3. HUD of a scene in Riu.

1) Computer-Understandable Description (CUD): The CUD of a given scene is composed of three main parts. First, a phase structure is used to specify the temporal relation between the phases. Second, each phase is depicted by a frame-based representation consisting of entities (e.g., a character or a prop) and relations (e.g., actions, properties, or relations between entities). Finally, elements shared between phases are placed in a common knowledge container, which also contains any additional domain knowledge we want to include. Notice that the separation between the common knowledge and individual phases is only relevant for the human author creating the stories. At runtime, the common knowledge is incorporated into each phase.

Fig. 2 shows the CUD representation of a scene from the *Evening Tide* story, used for the empirical evaluation presented below. Here, gray ovals represent entities, and white rectangles represent relations. In the story, Julian tried to sniff the honeysuckle plant near a beehive; in the second phase, the bees stung him, and he had to run back to the house. Here, "Julian" is represented as an entity of type "human" (entity types form a hierarchy, which we exclude from the figure for clarity). The action of "smelling" in the first phase is represented as a relation between "Julian" and "honeysuckle."

At the core of the CUD representation is an ontology, that is, a set of concepts relevant to the story world. In *Evening Tide*, we created an ontology containing 114 concepts (e.g., "human," "have," "run," "by").[1] We avoided synonymous concepts as much as possible in order to reduce the difficulty for SME to find analogies. For example, in Fig. 2, we used the property "young" to represent that Julian was a child.

In addition to cross-phase entities and relations, the common knowledge container includes domain knowledge and common sense knowledge implicit in the story. For instance, we use it to specify that the backyard is by the house, and that bees are dangerous. In order to be consistent when adding implicit common sense knowledge, we only inserted three types of knowledge in *Evening Tide*: spatial relations (e.g., the backyard is by the house), which entities are edible, and which ones are dangerous (e.g., bees are dangerous).

[1]The ontology used for our study can be found along the full source code of Riu and SAM at: https://sites.google.com/site/santiagoontanonvillar/software.

Each phase is annotated with FD elements (Section II-B) to depict plot/structural-level information such as agonist and antagonist, which one is stronger, the agonist's move or rest tendency, and the antagonist's helping or hindering role. As shown in our experimental evaluation in Section IV-B, FD increases the quality of the analogies found and the quality of the generated stories under certain circumstances.

2) Human-Understandable Description (HUD): The HUD representation is inspired by that of the GRIOT system [12]. In our system, it consists of a collection of preauthored natural language sentences called templates. They capture the same information present in the CUD, but in a format that is closer to natural language. Inheriting the phase structure in the corresponding CUD, the HUD for each phase consists of a set of templates $\{t_1, \ldots, t_n\}$, and a directed acyclic graph (DAG) that specifies the order in which they can be sequenced to generate the final output. Each node in the DAG represents a template, and edges specify their sequential order. The DAG allows us to specify several alternative templates for similar thematic content.

Fig. 3 shows the HUD of the scene represented in Fig. 2. Notice it contains two alternative templates to express similar content. For example, t_4 and t_5 are alternatives to be selected by Riu randomly at runtime. When equipped with more content, this DAG structure allows us to increase the variability and replayability of the story. Additionally, the common knowledge container can also have templates. This can be exploited by SAM if the text output of the generated story contains information from the common knowledge container.

The CUD and HUD representations are connected through links between corresponding elements. This process is currently done manually. Fig. 4 illustrates how a template in Fig. 3 is connected to the nodes from the graph in Fig. 2. Thanks to these links, SAM's manipulations of the CUD can be translated to the HUD and eventually to natural language output. This template-based text generation for narrative representation may be less flexible than other generative natural language-processing-based methods. However, our CUD–HUD configuration is interesting in that it allows an algorithm designed to work primarily with the CUD to also automatically generate text. In our case, this configuration allows us to focus mainly on exploring the narrative range of computational narrative at the plot

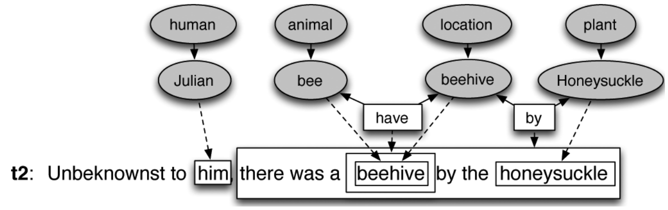

Fig. 4. Links between the CUD and the HUD.

level, while still having output that the user can interface with directly.

C. Memory Retrieval

The memory retrieval component of Riu finds the source stories most similar to a given target scene. Unless otherwise specified, source stories in Riu are memories, and the target story is the current scene from the external story world (simply story world from now on). Memory retrieval is used both for retrieving memories analogous to a given scene in the story world and for providing a source memory for the imaginative projection process. It consists of two steps.

1) Surface similarity: Riu first extracts a series of keywords from the target scene and all candidate memories (as potential source scenes), and selects the k memories which share the largest number of overlapping keywords with the current scene (in *Evening Tide*, $k = 3$). Keywords are extracted by taking all the concepts from CUD except for the FD-based ones. For example, the keywords in Fig. 2 are: "human," "Julian," "young," "beehive," "bee," "animal," "by," "house," "location," "plant," "honeysuckle," and "backyard."
2) Structural similarity: Then, SME is triggered to compute analogical mappings between each of k selected memories and the target scene as well as a numerical score representing the strength of the mappings. This SME score indicates how well a given mapping aligns with the structure-mapping principles, such as systematicity [22]. It is computed by a set of rules, each of which provides a belief value between -1 and 1, which are aggregated using Dempster–Shafer's rule [36]. Since SME favors deeper, structural similarity over the surface one (i.e., isolated nodes), the memory that shares the largest structures with the target scene will have the highest mapping strength, and thus will be retrieved.

Other similarity measures than the SME score can also be used. Elsewhere we have experimented with a measure that exploited domain knowledge from WordNet [35]. The rationale behind this two-step process is to minimize the use of the computationally expensive structural similarity by filtering the candidates first via surface similarity. This is a well-established procedure used in other analogy-based memory retrieval computational models such as many are called/few are chosen (MAC/FAC) [37].

D. Analogy-Based Story Generation: SAM

SAM takes two input parameters T and S (the target and source scenes, respectively), and outputs a new scene R, as the completion of T by analogy with S. We say that R is an analogical projection of S over T. For the rest of the paper, we will use the following terminology: an analogical connection is an individual one-to-one correspondence between a single entity or relation in the source domain and another one in the target domain. By contrast, a mapping is the complete set of connections found between the two domains.

The execution of SAM consists of four main steps:

1) Generate all possible phase mappings: Let P_T and P_S be the sets of phases of the two input scenes. We say that an injective mapping m from P_T to P_S (a mapping in which each element in P_T is mapped to one in P_S, and not two elements of P_T are mapped to the same element in P_S) is consistent when there are no inconsistencies in the ordering of each pair of phases $p_1, p_2 \in P_T$ with the corresponding pair of phases in P_S [$m(p_1)$ and $m(p_2)$]. This means if p_1 happens before p_2 in P_T, $m(p_1)$ must also happen before $m(p_2)$ in P_S.
   SAM computes M as the set of all the possible consistent injective mappings from P_T to P_S. An example of a consistent injective mapping between two scenes is shown in Fig. 5. Note that there might be a very large number of these mappings if the number of phases is large. With the typical number of phases used in *Evening Tide*, this number is manageable. Optionally, SAM allows the user to specify the desired phase mapping as an optional third input parameter m_i.
2) Find the analogical mappings: For each phase mapping $m \in M$, SAM does the following.
   - Let $P_S^m = \{p \in P_S | \exists p' \in P_T : m(p') = p\}$, i.e., all the phases from S in the mapping m.
   - e_S^m is constructed as all the entities in the CUDs of the phases in P_S^m and in the common knowledge of S. e_T is all the entities in the CUDs of T.
   - r_S^m is constructed as all the relations in the CUDs of the phases in P_S^m and in the common knowledge of S. r_T is all the relations in the CUDs of T.
   - SME is called using $e_T \cup r_T$ as the target domain and $e_S^m \cup r_S^m$ as the source domain. SME returns two things: an analogical mapping g_m from the target to the source domain, and a numerical score s_m.
   - $m^* \in M$ is selected as the phase mapping in M that maximizes s_m. If M is empty, m^* is not defined, and SAM returns an error token.
3) Construct a resulting scene R: A new scene R is constructed in the following way.
   - The phase structure DAG of R is copied from S.
   - The set of phases in R is $P_R = P_T \cup (P_S \setminus P_S^{m^*})$. The CUD of the common knowledge of T is added to all the phases in P_R that came from P_T, and the CUD of the common knowledge of S is added to all the phases in P_R that came from P_S.
   - The common knowledge in R is empty.
4) Transform R using the analogical mapping: The reverse of the analogical mapping g_{m^*} is applied to all the phases in P_R. For each phase $p \in P_R$, the following two steps are executed.

- For each entity or relation $e \in p$ such that $\exists e' \in S_T$: $g_{m^*}(e') = e$, we substitute e by e' in p.
- Every time an element $e \in p$ is substituted by another element e', we substitute the corresponding sentence fragment of e in the HUD of p by the corresponding sentence fragment of e' using the links between the CUD and HUD.

In all the phases in R that are from P_S, if there are any entities that do not appear in the mapping g_{m^*}, they are removed. Sentences and relations that refer to those removed entities are also removed. For example, irrelevant characters in the source, and sentences that refer to those, are removed. This last step prevents transferring irrelevant information to the generated scene.

An illustration of this process is shown in Fig. 6. We can see parts of the CUD and the HUD of a phase, a mapping g generated by SME, and the result of transforming the phase using such mapping.

The exploration of different mappings between phases in S and T gives SAM significant flexibility. For instance, if the last phases of S are mapped to the first phases of T, SAM can project backwards in time by generating past events that lead to the current situation in T. Another example is filling in the "temporal holes" between phases in T.

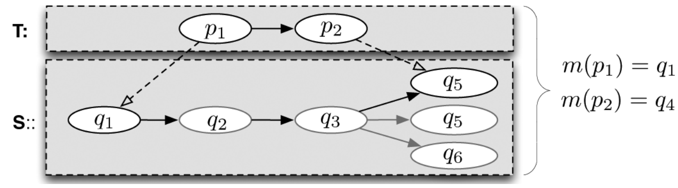

Fig. 5. Phase mapping between target T and source S.

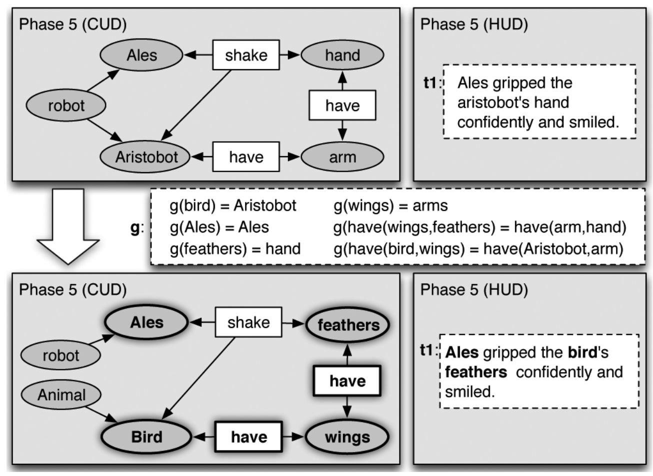

Fig. 6. Transformation of a phase using the analogical mapping g.

E. A Sample Output

In this section, we will show one of the stories generated by SAM, used in our user study discussed below. A larger sample story can be found in [38]. Using the following source story:

> "Julian hadn't eaten all day and was counting on the crab trap to provide him with a feast of hearty shellfish. When he pulled the trap off the water it was filled with the largest crabs he had ever seen. So large in fact, that the weight of them caused the rope to snap just before Julian could pull the trap onto the deck."

and the following target story:

> "Zack is on deck, ready to face a storm. There's a flash of lightning in the distance. Suddenly, there's a bump on the side of the boat. Zack looks over. It is a gigantic cod! He's never seen one this large and close to the surface before. The storm is closing in. He races to get some fishing gear and try to catch it."

SAM completes the target story by generating the following extra phase at the end:

> "When Zack pulled the fishing gear off the water it was filled with the largest cod Zack had ever seen. So large in fact that the weight of cod caused the rope to snap just before Zack could pull the fishing gear onto the deck."

To generate this story, step 2 of SAM maps elements from the source to the target using SME. For instance, "fishing gear" is mapped to "crab trap," as we can see in the generated story. In step 4, SAM takes the second phase of the source story, and replaces the appearances of "crab trap" by "fishing gear." In this case, SAM completed the story by adding one additional phase at the end.

IV. Evaluation: A User Study

In order to evaluate the ASG components in Riu, including memory retrieval and analogical projection, we conducted a user study to answer the three research questions below.

1) How effective is the ASG component in identifying analogical connections in ways similar to readers in our user group? We compare our system's performance with the readers' in several aspects of analogy identification.
2) Is our choice of FD suitable in the context of computational narrative? Here we compare the performance of our system with FD and with other domain knowledge.
3) What is the quality of the stories generated by our system from the perspective of the readers? We directly ask the readers to rate these stories.

As mentioned earlier, our study does not cover the interactive aspects of Riu. The stories we used are noninteractive snippets from the larger *Evening Tide* interactive story.

A. Study Design

Our user study is targeted at the general public between the ages of 18 and 65. Measures are taken to minimize our influence on the readers' interpretation. The participants are only informed of the broad topic of the study—computational narrative; no information about the system or whether/which stories were generated by the computer are revealed. We also avoid any unnecessary technical jargon in the phrasing of the survey. For instance, "source" and "target stories" are simply referred to as "story A" and "story B." There is no mention of anything related to FD. (For consistency, we will continue using the technical terminology here.)

The narrative texts in the study are small excerpts from the *Evening Tide* story, authored by a recent graduate from the Screen Writing program at Drexel University (Philadelphia,

TABLE I
PROPERTIES OF THE S/T PAIRS USED IN OUR STUDY

	FD similarity	Surface similarity
S/T 1	low	low
S/T 2	low	high
S/T 3	high	low
S/T 4	high	high

PA, USA). In the overall design of the stories, we intentionally kept the tone of the writing close to screenplay, the concise and bare-bone style of which is more suitable for the capabilities of our system at this stage than the more elaborate prose fiction style. All the stories used in this study are composed of two phases and have simple narrative structures. Some stories have a strong FD structure, whereas others are FD neutral (i.e., where there is no clear agonist or antagonist). When used as a source story, a story is presented in its entirety. Otherwise, as a target story, only the first phase is included. The average length of a source story is 73.25 words, and 38.0 for target stories. As mentioned above, we changed the name of the protagonist to a different one in each story in order not to imply any surface connection. SAM's generated stories were minimally edited to eliminate grammatical errors. These revisions correct obvious low level grammatical mistakes, such as capitalization and missing determinants. Changes requiring a significant modification in the sentences were not made.

The study is organized both by task and by content. It contains four main tasks, each of which evaluates one of our system's main generative steps. In each task, participants answer the same set of questions for four different source/target (S/T) pairs. Each S/T pair contains a complete source story and an incomplete target story. The four S/T pairs in the study were selected to represent different degrees of surface and FD similarity (Table I). They also represent different degrees of success in SAM's performance.

1) Task 1 (Story Elements Mapping): This task is designed to evaluate to what extent our system can identify mappings between source and target domains in ways similar to human participants. For each S/T pair, a participant sees a source story, a target story, and two lists of entities (i.e., characters and objects) and relations (e.g., "Herman is at the booth") included in the source and target stories, respectively. We only list entities and relations explicitly mentioned in the stories in order to minimize our particular interpretations. Each participant is asked to identify as many analogical connections between the two lists as possible.

2) Task 2 (Story Similarity): This task allows us to compare the stories that the participants find the most similar to a target story to Riu's results. For each of the four S/T pairs, the participant is asked to rank four potential matching source stories based on their respective similarities to a target story.

3) Task 3 (Analogical Projection): This task aims at evaluating the quality of SAM's analogical projection. For each of the four S/T pairs, the participant is presented with a complete source story and an incomplete target story. We first ask her to continue the incomplete story by writing at most three relatively simple sentences in English free-text. This method is based on what is known as the "story continuation test" from the empirical literary studies field [39]. As we intend to "blackbox" Riu's analogical projection process from the participant, the following description is provided as guidance: "If a new story is very similar to a known story, we can sometimes predict what happens by drawing analogy from the known story. Read the complete story A, and a similar but incomplete story B. Continue story B in ways you think are the most similar to story A." Next, we present the participant with a continuation generated by SAM and ask her to rate its overall quality on a five-point Likert scale. In this task, the participant's free writing offers insight into what is the most "natural" and relatively unconstrained continuation to a human reader, and the rating provides a quantitative evaluation of SAM's output.

4) Task 4 (Overall Story): This task evaluates the quality of the complete stories generated by SAM. In addition to the four story continuations generated by SAM (also used in task 3), we added two more as benchmarks. One of them is a poorly constructed story, created by manually copy-pasting the second phase of a story after the first phase of another. It represents what we believe is a low-quality ASG story. The other benchmark is, unknown to the participants, written completely by the human author who created the story world. These two additions are intended to set a baseline for the range of scores. The order in which the six stories are arranged is randomized. Each story is rated on a five-point Likert scale along three dimensions: plot coherency, character believability, and overall quality. Story generation has multiple challenges; some (e.g., character believability) arguably more central to the narrative content than others (e.g., grammar). In this task, we intend to separate the different aspects of readers' satisfaction. Finally, we ask the participants for any additional feedback.

B. Results

In response to our e-mail recruitment, 31 people completed the survey. Among them, 27 were male, three were female, and one was undisclosed. Their age range was between 18 and 49, with a mean between 26 and 27. Below are results for each task.

1) Task 1 (Story Elements Mapping): This task collects data on what the participants regard as the appropriate analogical mapping between the story pairs. In order to assess the contribution of FD, we compared the mappings identified by human participants with those generated at random, and with those generated by our system using four different settings of domain knowledge. The settings are a) SAM-fd: the standard setting only with FD; b) SAM-bare: a bare setting where we removed the FD annotations from the standard setting; c) SAM-wn: where we replaced the FD in the standard setting with domain knowledge of categories automatically extracted from the "hypern" database in WordNet; and d) SAM-wnfd, including both FD and WordNet knowledge.

Under the SAM-wn setting, for instance, the entity "fish" is supplemented with WordNet properties such as "aquatic-vertebrate," "vertebrate," and "animal." This helps SME match entities from two stories according to how many properties they share. However, too much domain knowledge will significantly increase the computational complexity of finding an analogical mapping using SME. Hence, for each entity, we set the upper limit of six additional properties in the order WordNet returns them, as six is the maximum number with which the experiments would run under reasonable time bounds (1 h).

TABLE II
PROBABILITY THAT A PARTICIPANT IDENTIFIES AN ANALOGICAL CONNECTION GENERATED BY EACH CONFIGURATION OF SAM (HIGHER IS BETTER), AND THE NUMBER OF CONNECTIONS FOUND (SIZE)

	Rnd.	Hmn.	SAM-fd	SAM-bare	SAM-wn	SAM-wnfd
S/T 1	0.04	0.46	0.48	0.35	0.35	0.48
S/T 2	0.06	0.61	0.76	0.81	0.81	0.76
S/T 3	0.05	0.57	0.48	0.47	0.48	0.48
S/T 4	0.07	0.75	0.80	0.80	0.80	0.80
Avg.	0.05	0.60	0.63	0.61	0.61	0.63
size	-	7.14	4.50	3.50	4.00	4.25

For each setting of domain knowledge, we computed the probability that given a connection identified by SAM, a given participant also identified that connection. Again, a connection is a particular one-to-one correspondence between two entities or relations in source and target domains, whereas a mapping is the whole set of connections between the two domains. In other words, if SAM identifies an analogical connection that no participant reported, this probability is 0. If all participants reported the same connection as SAM, the probability is 1. We refer to this probability as the connection score.

The results of the different settings (Table II) show that the connection score of randomly generated connections is very low (0.05 on average). By contrast, the connection score of the participants is 0.60 (this means that given two participants and a connection identified by one of them, the probability that the other participant also identified it, as evaluated using a standard leave-one-out procedure, is 0.60). This number is closely matched by SAM's connection scores using all four versions of domain knowledge. This means that the connections identified by SAM are indistinguishable from those identified by a human participant.

The key difference between human participants and SAM is the size of the mappings (i.e., the number of analogical connections) they each find. The participants found an average of 7.14 connections, whereas SAM found significantly fewer.[2] SAM-bare only finds an average of 3.50 analogical connections. With the FD annotation, this number rises to 4.50 connections. SAM-wn finds an average of 4.00 connections and SAM-wnfd finds an average of 4.25 connections. This indicates that domain knowledge helps SAM identify more connections, and the knowledge provided by FD is the most effective in this respect. In particular, SAM with FD could find the largest amount of connections. It signals that FD annotations provide compact and useful domain knowledge.

We believe the reason why SAM finds more connections with FD than with WordNet, interestingly, is the following. Note that the knowledge added by FD corresponds precisely to the high-level structure of the scenes, and the knowledge added by WordNet corresponds to specific properties of the individual entities (and not to their relations). Thus, having in mind that structure-mapping theory, used by SME, has a bias toward high-level relations, a possible explanation of the results is that SME is more influenced by the FD knowledge than by that of WordNet.

[2]For each S/T pair, we only count the number of connections between entities and relations that are explicitly mentioned in the stories. Connections of the implicit domain knowledge used by SAM, such as FD or categories from WordNet, are not counted toward this size measure. As described in Section IV-A, the participants receive the exact sets of entities and relations without implicit domain knowledge.

TABLE III
KENDALL τ CORRELATION INDEX BETWEEN THE GROUND TRUTH AND 1) SEVERAL CONFIGURATIONS OF RIU, 2) A RANDOM ORDERING, AND 3) A RANDOM PARTICIPANT. (LOWER τ MEANS MORE CORRELATION)

	Riu-fd	*Riu*-bare	*Riu*-wn	*Riu*-wnfd
Structural Similarity	0.08	0.13	0.33	0.21
Surface Similarity	0.33	0.33	0.29	0.33
Random Ordering	0.50			
Random Participant	0.14			

In terms of particular S/T pairs, S/T 1 in Table II displays significantly lower connection scores in all three domain knowledge settings compared to other S/T pairs. This is because S/T 1—the S/T pair with low FD similarity and low surface similarity—does not support a clear analogy. As a result, participants tend to disagree in their mappings. By contrast, S/T 4—the pair with both high FD and surface similarities—exhibits higher scores. In this pair, most participants find exactly the same analogical mapping, consisting of seven analogical connections. The mapping found by SAM-fd is very similar; it contains four connections, all of which are among the seven found by most participants.

2) Task 2 (Story Similarity): We evaluated how much Riu's retrieval results align with the participants' intuitive notion of analogical similarity. The participants' rankings of the potential matching stories were aggregated using the standard Borda count [40]. The aggregated participants' ordering, which we refer to as the ground truth, is compared with the ranking generated by Riu's memory retrieval component. We do so by using the Kendall τ ranking correlation index [41], which is 0 for two identical orderings, 1 for opposite orderings, and expected to be around 0.5 for random orderings.

As in task 1, we compared the ground truth with: a) a random ordering; b) the ordering given by a random participant in our study; and c) the ordering Riu generated with four different domain knowledge settings: with only FD, without FD or WordNet, with only WordNet, and with both. In each domain knowledge setting, we tested Riu in two conditions: a) only using a basic surface similarity measure (based on the percentage of keywords shared between the two stories); and b) using both surface and structural similarity measures, as actually used in Riu. Results are summarized in Table III.

The ordering generated by Riu with FD is almost identical to the ground truth, except for two out of 24 order relations (six order relations for each of the four S/T pairs), yielding a very low τ distance, 0.08. It confirms that Riu's retrieval component aligns with the participants' intuitive notion of similarity in the short stories. The orderings generated using knowledge from WordNet are less similar to the ground truth. We believe that this is because, when using WordNet, the retrieval component focuses too much on the surface similarity between the entities in the stories, rather than the story structure. This result shows that the knowledge provided by FD contributes better to this similarity assessment than the domain knowledge provided by WordNet, since the former contains structural relations between the entities/relations rather than their surface similarity captured

TABLE IV
A SAMPLE CLUSTERING OF PARTICIPANTS' FREE-WRITING CONTINUATIONS FOR TASK 3 USING S/T 4

Cluster	Size	Cluster Description	Representative Example
1	19	Julian paddles back	*"Julian had to paddle the motorboat back to the dock."*
2	5	Julian swims back	*"Julian was forced to swim to shore and explain what happened to his father."*
3	5	They call for help	*"Herman had to hire a bigger boat to drag the small boat back to the dock."*

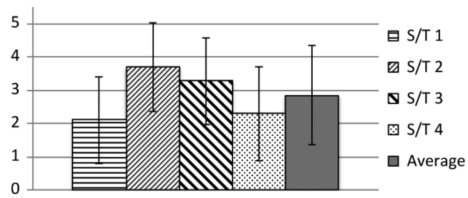

Fig. 7. Average scores of SAM-generated story continuations in task 3. Error bars indicate standard deviation.

TABLE V
ANALYSIS OF THE PARTICIPANTS' FREE-WRITING CONTINUATIONS. "*" MARKS WHERE SAM'S SEMANTIC MISTAKE OCCURRED

		Analogous to the Source		Not Analogous
	Clusters	= SAM	$\neq$ SAM	
S/T 1	6	0	19	11
S/T 2	5	11	5	14
S/T 3	7	3	20	6
S/T 4	3	19*	5	5

in the latter. In addition, all the orderings generated using surface similarity are significantly different from the ground truth. This justifies Riu's use of the more computationally expensive structural similarity.

Curiously, the ordering generated by Riu with FD is closer to the ground truth (τ distance of 0.08) than that of a random participant (τ distance of 0.14) in our study. We believe this is due to the fact that different participants pay attention to different factors in the stories when assessing their similarities.

3) Task 3 (Analogical Projection): Let us first look at how the participants rated the quality of the continuations generated by SAM (standard version with FD) on a five-point Likert scale. Fig. 7 summarizes the results, showing the means and standard deviations of the ratings. Among the four SAM-generated story continuations, S/T 2 and S/T 3 were considered relatively high quality by the participants (3.70 and 3.27), while the other two received lower ratings. The continuation rated the lowest was from S/T 1, quoted below:

> [Source:] "In the carnival, Jacob played at a booth where it had twenty narrow fishbowls with goldfishes inside. He tossed the blue ball and watched it ricochet off the rims of the bowls. Jacob cried. Jacob's father, Andrew, snatched one of the bowls and gave it to Jacob. The attendant was intimidated by Andrew's size and let it slide. Andrew told Jacob that life wasn't about fun and games. The goldfish died a week later and Andrew made Jacob flush it down the toilet."

> [Target:] "As a child, Eva would often sniff the honeysuckle in the backyard. Unbeknownst to her, there was a bee's nest by the honeysuckle."

> [SAM-Generated Continuation:] "Eva cried."

As S/T 1 contains low surface and low FD similarity, it is difficult for SAM to come up with an analogical projection. The only knowledge it was able to transfer from the source domain is that the agonist cried. As illustrated below, the participants faced the same difficulty in their own free-writing.

S/T 4 with high surface and high FD similarity also received a lower rating, even though SAM transferred a lot of content from the source story. Based on the qualitative feedback (discussed below), we believe that this is due to a semantic mistake made by SAM. In the source story, the protagonist forgot to fill up the tank of her car. After realizing her mistake, she had to turn around before reaching her destination. In the target story, a boat ran out of gas. And yet in the continuation generated by SAM, the protagonist still drove the boat back, which clearly violates common sense.

In addition, the participants' own free-writing story continuations provided us with useful information to contextualize their ratings of the ones generated by SAM. For each S/T pair, we first clustered these continuations using grounded theory methods. For example, for S/T 4, the target story was the following (the source is not included due to space limits):

> "Herman, Julian's father, owned a small motorboat. One night Julian snagged the keys and took the boat from the dock to a nearby island without permission. The boat ran out of gasoline in the middle of the bay."

The participants' continuations for this particular story were grouped into three clusters, based on their analogical projection. One example of each cluster is shown in Table IV.

Our further analysis of the results is summarized in Table V. The "clusters" column shows the number of clusters obtained. In S/T pairs with low surface similarity (i.e., S/T 1 and 3), participants came up with a more varied set of continuations. By contrast, in S/T 4 (strong FD and surface similarity), most participants converge in how to continue the story.

All participants' continuations are also divided into two groups based on whether they are analogous to the source. Among those that are analogically related, we further differentiate those containing the same analogies as the ones SAM found ("= SAM" column) from the others ("$\neq$ SAM" column). We categorize two analogies as the same when two continuations depict the same narrative events, even though they are depicted through different text. For story pairs with high surface similarity (S/T 2 and 4), a significant number of participant-authored continuations are similar to SAM's. In particular, for S/T 4, 19 out of all 29 participant-authored continuations have almost identical content: the protagonist has to row back to a dock. SAM generated a very similar continuation, where the protagonist brings the boat back to a dock. However, due to the semantic mistake discussed above,

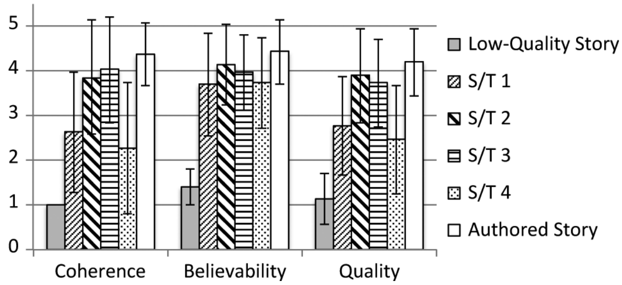

Fig. 8. Average scores of SAM-generated stories in task 4, compared to two benchmark stories. Error bars indicate standard deviations.

SAM did not understand that when a boat is out of gas, it cannot be brought back differently. In cases where there are no strong FD similarities (S/T 1 and 2), the participants' continuations often do not contain clear analogical connections to the source story. It illustrates that the participants encountered similar difficulties as SAM did in this situation.

4) Task 4 (Overall Story): In this final task, the participants rated the quality (i.e., story coherence, character believability, and overall quality) of six complete stories: four generated by SAM (standard with FD), one poorly constructed story, and one written by a human author. Results, summarized in Fig. 8, show that the ratings for the low-quality story and the human-authored story define the two extremes and set the context for the rest. The scores obtained by SAM are closer to those of the human-written story than to the low-quality one. Specially in terms of character believability, SAM's score was relatively high (3.88 on average out of five compared to 4.43). Certain generated stories, 2 and 3, obtain much higher scores than the average. Comparing Figs. 7 and 8, we can see that the overall scores obtained by the stories in task 4 are highly correlated with those in task 3, as expected.

Although there is a strong correlation with identical ordering of the four stories (Pearson correlation coefficient of 0.95), the same story is always rated higher in task 4 than in task 3. This is an interesting phenomenon because the stories have not been changed. A possible reason is that the ratings of task 3 were collected right after the participants wrote their own continuation. Thus, they may have higher expectations for the quality of the continuations provided to them at that time. Another possibility is that in task 4, the overall ratings are collected after rating of specific narrative dimensions. In other words, this number is influenced by the participants' immediately prior rankings for coherence and character believability.

5) Users Feedback: The feedback provided some further insights into the obtained results. For example, several participants complained that they identified additional analogies between the S/T pair, but were not able to specify them using the given list of entities and relations in task 1. This means that the CUD and/or the ontology we authored could be improved, potentially leading to better results.

A number of participants mentioned that some stories had grammatical mistakes and thus had rated them lower. Remember that although we fixed some obvious low-level grammatical mistakes in stories generated by SAM (e.g., capitalization, missing determinants), changes requiring a significant modification in the sentences were not made. These comments indicate that the quality of the text used to represent our stories had an impact on the user ratings. We believe that text generation is an integral part of computational narrative that contributes to a user's engagement. We have already started working toward this goal in our current work [42].

Several participants wondered which stories were created by computers and which by humans. Some even asked whether some of the semantic mistakes, such as the "boat running out of gas," were introduced purposefully or were errors.

Finally, some feedback may explain the very large number of analogies identified by participants in task 1. One comment says: "A few analogy-matching parts seemed like a stretch and I provided answers where in everyday circumstances I would normally say there was no analogy." It shows that some participants may have gone to great lengths to identify analogies that they normally would not have identified.

C. Discussion

Summarizing, our user study has helped us answer our three research questions (Section IV).

1) Analogies found by our system align with analogies found by our participants. Also, the retrieval mechanism aligns with the participants' intuitive notion of similarity in the short stories. These are specially important facts given the result of task 3, where we see that identifying the appropriate source stories is crucial to the success of ASG. When the source was not sufficiently similar to the target, most of our participants did not continue the target by analogy. Instead, they invented a continuation that is mostly unrelated to the source.
2) A common theme across all our results is that FD helps SME, the internal algorithm used by our system, to find analogies, for better retrieval and projection results. We believe that this is because FD aligns particularly well with structure-mapping theory by providing plot-level information about the relations between the narrative elements in the stories.
3) Although the quality of stories generated by SAM is still not on a par with a human-authored story, participants rated some of SAM's stories relatively highly, specially in terms of character believability.

V. Related Work

As we have described SAM and Riu in detail, this section will offer targeted comparison with related work. One of the most related algorithms to SAM is the Story Translator [34]. It uses analogy as the main generative method and planning to fill in the gaps in the analogy-generated content. Its input is a story represented as a plan and two domain models. The domain models contain the set of objects and planning operators available in each domain. The CAB algorithm [25] is used to find a mapping between the two domain models, and this mapping is then used to translate the input story from one domain to the other, filling the gaps using planning in case the mapping is not complete. The main difference between SAM and the Story Translator is that the Story Translator only computes analogies between domain models, whereas SAM also operates with specific narrative events. For example, SAM can find an analogical connection concerning the specific event "Eva would often sniff the honeysuckle," whereas the Story Translator would focus on analogies concerning the general definition of the planning operator *sniff*. Notice that SAM can also operate at this level, if as part of the domain knowledge, the properties of *sniff* are specified. Also, SAM directly generates text; the Story Translator returns its generated stories in the form of plans. This allows SAM not to be confined by the expressive restrictions of plans, and its output can be directly presented to an average user.

Computational analogy was also used by Li and Riedl [43] for story generation. However, in the approach of Li and Riedl, analogy was used to create new types of gadgets to be used inside a planning-based story generation system, rather than to generate actual story events.

The related technique of case-based reasoning (CBR) has been used in story generation systems [33], [44]. For example, MINSTREL [33] is a general model of creativity that generates stories by executing transform–recall–adapt methods (TRAMs). Some of those TRAMs, like the "cross-domain-solution" TRAM, use computational analogy. In particular, given a problem (an incomplete story) in a domain D_1, the TRAM finds another domain D_2 and an analogical mapping between D_1 and D_2 (similar to the Story Translator). Then, it maps the story from D_1 to D_2, solves the problem, and then maps back the result from D_2 to D_1.

By contrast, MEXICA [45] generates stories by adding one action at a time to a given story. In order to select the next action to add, MEXICA retrieves, from a story repository, a past story that is the most similar to the current state of the story. This process of comparison can be seen as trying to find an analogical mapping between the current story state, and the past stories. Whereas the Story Translator and MINSTREL find analogies at the domain definition level, MEXICA finds them between specific story states [called story-world contexts (SWCs)], which are equivalent to the phases in our system. In contrast with those systems, Riu uses both domain definitions (domain knowledge) and specific narrative events to find analogies between the target and source stories.

Other work that uses analogy-related techniques includes the PRINCE system [46]. It enriches a story by generating metaphors about its story elements in the domain T using their equivalents in domain S. Also, the GRIOT system [12] implements the conceptual blending theory [47] and uses it to generate affective blends for interactive poetry.

Overall, comparing ASG to planning-based systems (such as Tale-spin [8] and Fabulist [10]), the latter have the advantage in that the author can specify the initial and ending states of a story, and thus may have more direct control of the generated story. In ASG systems, such control is exerted by providing different source stories. In general, a source story in ASG can influence both the content and the discourse of the generated story. Stories generated by planning tend to focus on actions and change, as indicated by the planning operators. ASG systems, however, do not have this bias. If the source and target stories are action based, then the resulting story will also be more likely so, but if the source and target stories are very descriptive, the same will be passed on to the resulting stories. In terms of knowledge engineering needed, planning systems require a domain model with a complete planning operator definition, while ASG systems require a collection of source stories. Both approaches depend significantly on the labor-intensive knowledge engineering process.

VI. Conclusion and Future Work

In this paper, we have presented and evaluated an approach to story generation based on computational analogy. The key characteristics of our technical approach are: 1) the use of a dual representation formalism for stories having a CUD and a HUD; 2) the use of FD to enhance the story representation especially at the plot level; and 3) the internal use of structure-mapping theory to find analogies between stories by means of the SME algorithm.

In this paper, we have placed our focus on the empirical evaluation of the ASG components of our Riu system: a story retrieval component and a story generation component (SAM). The study confirmed our three main hypotheses: 1) similarity and analogical mapping in our system aligns with the participants' intuitive perception of them; 2) FD significantly enhances story generation in Riu; and 3) stories of relatively high quality can be generated using computational analogy.

The evaluation provided us with a significant amount of insight for future work. As already reported in the evaluation of MINSTREL [33] and confirmed in our study, participants are influenced by the quality of the final text. Therefore, we plan to extend SAM's text generation capabilities, as explored in our current work [42]. Additionally, semantic mistakes can also drastically affect readers' reaction to the stories. We intend to exploit additional domain knowledge about the entities and relations to minimize such mistakes. For example, we intend to explore commonsense knowledge-bases such as CyC or Open Mind, or the incorporation of action definition knowledge, as used in planning-based story generation systems.

As media theorists and art historians have repeatedly pointed out, technological invention by itself does not automatically deliver the birth of a medium. After the first public screening in the Grand Café in 1895, for instance, it took filmmakers and inventors decades to invent the medium by developing the major elements of filmic storytelling, including the closeup, the chase scene, and the standard feature length [3]. In this process, a key component was to expand the range of expressions a film could convey. Informed by the history of traditional media, our long-term goal is to broaden the expressive power of computational narrative so that it can encompass a wider variety of human experiences and conditions. This paper presents one of our first steps toward this goal by exploring analogy-based story generation and evaluating these stories from the vantage point of the readers.

References

[1] J. D. Bolter and R. Grusin, *Remediation: Understanding New Media*. Cambridge, MA, USA: MIT Press, 1999.

[2] B. Laurel, *Computers as Theatre*. Menlo Park, CA, USA: Addison-Wesley, 1991.

[3] J. H. Murray, *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Cambridge, MA, USA: MIT Press, 1998.

[4] M. Mateas and A. Stern, "A behavior language for story-based believable agents," *IEEE Intell. Syst.*, vol. 17, no. 4, pp. 39–47, Jul./Aug. 2002.

[5] C. Crawford, *Chris Crawford on Interactive Storytelling*. Berkeley, CA, USA: New Riders, 2004.
[6] G. Landow, *Hypertext 2.0: The Convergence of Contemporary Critical Theory and Technology*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1997.
[7] M.-L. Ryan, *Avatars of Story*. Minneapolis, MN: Univ. Minnesota Press, 2006.
[8] J. Meehan, "The metanovel: Writing stories by computer," Ph.D. dissertation, Yale University, New Haven, CT, USA, 1976.
[9] M. Lebowitz, "Creating characters in a story-telling universe," *Poetics*, vol. 13, pp. 171–194, 1984.
[10] M. O. Riedl and R. M. Young, "An intent-driven planner for multi-agent story generation," *Proc. 3rd Int. Joint Conf. Autonom. Agents Multiagent Syst.*, pp. 186–193, 2004.
[11] M. Theune, E. Faas, A. Nijholt, and D. Heylen, "The virtual storyteller: Story creation by intelligent agents," *Proc. Technol. Interactive Digit. Storytelling Entertain. Conf.*, pp. 204–215, 2003.
[12] D. F. Harrell, "Walking blues changes undersea: Imaginative narrative in interactive poetry generation with the Griot system," in *Proc. Workshop Comput. Aesthetics, Artif. Intell. Approaches to Happiness and Beauty*, 2006, pp. 61–69.
[13] N. Montfort, "Generating narrative variation in interactive fiction," Ph.D. dissertation, Univ. Pennsylvania, Philadelphia, PA, USA, 2007.
[14] J. Zhu and S. Ontañón, "Towards analogy-based story generation," in *Proc. 1st Int. Conf. Comput. Creativity*, Lisbon, Portugal, 2010, pp. 75–84.
[15] L. Aronson, *Screenwriting Updated: New (And Conventional) Ways of Writing for the Screen*. Los Angeles, CA, USA: Silman-James, 2000.
[16] M. Turner, "Double-scope stories," in *Narrative Theory and the Cognitive Science*, D. Herman, Ed. Stanford, CA, USA: CSLI Publications, 2003.
[17] H. Murakami, *Hard-Boiled Wonderland and the End of the World*. New York, NY, USA: Kodansha, 1991.
[18] S. Ontañón and J. Zhu, "Story and text generation through computational analogy in the RIU system," in *Proc. AAAI Conf. Artif. Intell. Interactive Digit. Entertain.*, 2010, pp. 51–56.
[19] R. P. Hall, "Computational approaches to analogical reasoning: a comparative analysis," *Artif. Intell.*, vol. 39, no. 1, pp. 39–120, 1989.
[20] R. M. French, "The computational modeling of analogy-making," *Trends Cogn. Sci.*, vol. 6, no. 5, pp. 200–205, 2002.
[21] T. G. Evans, "A program for the solution of a class of geometric-analogy intelligence-test questions," in *Semantic Information Processing*. Cambridge, MA, USA: MIT Press, 1968.
[22] B. Falkenhainer, K. D. Forbus, and D. Gentner, "The structure-mapping engine: Algorithm and examples," *Artif. Intell.*, vol. 41, pp. 1–63, 1989.
[23] K. J. Holyoak and P. Thagard, "Analogical mapping by constraint satisfaction," *Cogn. Sci.*, vol. 13, pp. 295–355, 1989.
[24] J. E. Hummel and K. J. Holyoak, "Distributed representations of structure: A theory of analogical access and mapping," *Psychol. Rev.*, vol. 104, pp. 427–466, 1997.
[25] L. B. Larkey and B. C. Love, "Cab: Connectionist analogy builder," *Cogn. Sci.*, vol. 27, no. 5, pp. 781–794, 2003.
[26] M. Mitchell, "Copycat: A computer model of high-level perception and conceptual slippage in analogy-making," Ph.D., Univ. Michigan, Ann Arbor, MI, USA, 1990.
[27] R. M. French, *The Subtlety of Sameness: A Theory and Computer Model of Analogy-Making*. Cambridge, MA, USA: MIT Press, 1995.
[28] G. E. McGraw, Jr., "Letter spirit (part one): Emergent high-level perception of letters using fluid concepts," Ph.D. dissertation, Dept. Comput. Sci., Indiana Univ., Bloomington, IN, USA, 1995.
[29] D. Gentner, "Structure-mapping: A theoretical framework for analogy," in *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*. San Mateo, CA, USA: Morgan Kaufmann, 1988, pp. 303–310.
[30] D. R. Hofstadter and M. Mitchell, "The Copycat project: A model of mental fluidity and analogy-making," in *Advances in Connectionist and Neural Computation Theory: Analogical Connections*. Norwood, NJ: Ablex, 1994, vol. 2, pp. 31–112.
[31] L. Talmy, "Force dynamics in language and cognition," *Cogn. Sci.*, vol. 12, no. 1, pp. 49–100, 1988.
[32] S. Ontañón and J. Zhu, "Story representation in analogy-based story generation in Riu," in *Proc. IEEE Symp. Comput. Intell. Games*, 2010, pp. 435–442.
[33] S. R. Turner, "Minstrel: A computer model of creativity and storytelling," Ph.D. dissertation, Univ. California at Los Angeles, Los Angeles, CA, USA, 1993.
[34] M. Riedl and C. León, "Generating story analogues," in *Proc. AAAI Conf. Artif. Intell. Interactive Digit. Entertain.*, 2009, pp. 161–166.
[35] S. Ontañón and J. Zhu, "On the role of domain knowledge in analogy-based story generation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2011, pp. 1717–1722.
[36] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976, vol. 1.
[37] D. Gentner and K. D. Forbus, "MAC/FAC: A model of similarity-based retrieval," *Cogn. Sci.*, vol. 19, pp. 141–205, 1991.
[38] S. Ontañón and J. Zhu, "The SAM algorithm for analogy-based story generation," in *Proc. AAAI Conf. Artif. Intell. Interactive Digit. Entertain.*, 2011, pp. 67–72.
[39] C. Emmott, A. Sanford, and L. Morrow, "Capturing the attention of readers? stylistic and psychological perspectives on the use and effect of text fragmentation in narratives," *J. Literary Semantics*, vol. 35, no. 1, pp. 1–30, 2006.
[40] T. W. Sandholm, "Distributed rational decision making," in *Multiagent Systems*, G. Weiss, Ed. Cambridge, MA, USA: MIT Press, 1999, pp. 201–258.
[41] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.
[42] J. Valls and S. Ontañón, "Natural language generation through case-based text modification," in *Case-Based Reasoning Research and Development*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 2012, vol. 7466, pp. 443–457.
[43] B. Li and M. Riedl, "A phone that cures your flu: Generating imaginary gadgets in fictions with planning and analogies," in *Proc. 4th Workshop Intell. Narrative Technol.*, Stanford, CA, 2011, pp. 41–48.
[44] P. G. B. Díaz-Agudo, F. Peinado, and R. Hervás, "Story plot generation based on CBR," *J. Knowl.-Based Syst.*, vol. 18, no. 4-5, pp. 235–242, 2005.
[45] R. Pérez and M. Sharples, "Mexica: A computer model of a cognitive account of creative writing," *J. Exp. Theor. Artif. Intell.*, vol. 13, no. 2, pp. 119–139, 2001.
[46] R. Hervás, R. P. Costa, H. Costa, P. Gervás, and F. C. Pereira, "Enrichment of automatically generated texts using metaphor," in *MICAI 2007: Advances in Artificial Intelligence*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 2007, vol. 4827, pp. 944–954.
[47] G. Fauconnier and M. Turner, *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York, NY, USA: Basic Books, 2002.

Jichen Zhu received the M.S. degree in computer science and the Master of Entertainment Technology degree from Carnegie Mellon University, Pittsburgh, PA, USA and the Ph.D. degree in digital media from Georgia Institute of Technology, Atlanta, GA, USA.

She is an Assistant Professor in the Digital Media Program of the Antoinette Westphal College of Media Arts, Drexel University, Philadelphia, PA, USA. Her research focuses on the intersection of AI, human–computer interaction, creative expression, and critical/media theory. Her current research interests include interactive narrative, serious games, computational creativity, and digital humanities.

Santiago Ontañón received the Ph.D. degree from the Autonomous University of Barcelona (UAB), Barcelona, Spain.

He is an Assistant Professor in the Computer Science Department, Drexel University, Philadelphia, PA, USA. His main research interests are game AI, case-based reasoning, and machine learning, fields in which he has published more than 90 peer-reviewed papers. Before joining Drexel University, he held postdoctoral research positions at the Artificial Intelligence Research Institute, Barcelona, Spain; at Georgia Institute of Technology, Atlanta, GA, USA; and at the University of Barcelona, Barcelona, Spain.