Integrating a Plot Generator and an Automatic Narrator to Create and Tell Stories

Nick Montfort[1] and Rafael Pérez y Pérez[2]

[1] Program in Writing and Humanistic Studies
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139 USA
nickm@nickm,com
[2] Departamento de Tecnologías de la Información
Universidad Autónoma Metropolitana (Cuajimalpa)
Av. Constituyentes 1054, México D. F. 11950, México
rpyp@rafaelperezyperez.com

Abstract. We describe the integration of MEXICA, a well-developed story generation system which produces high-level plots, with the nn Narrator, an existing system for automatically narrating based on a simulated world and a plan with narrative specifications. In our initial integration of these systems in a pipeline, we did not seek to model the creative process explicitly. The initial integration nevertheless exposed issues of story representation and of the distinction between content and expression that are relevant to creativity. We describe these and discuss plans for a blackboard system that connects plot generation and narration more deeply.

Key words: Story generation, plot generation, narrative variation, narrating, system integration, story representation, action, event representation.

1 MEXICA-nn: Plot Generation plus Narrative Variation

Computer systems to generate stories are a fairly recent innovation, with the most significant early example (James Meehan's Tale-Spin) dating from the mid-1970s. Given this short history, such systems have been used for quite some time to understand creativity. Scott Turner's Minstrel, developed in the mid-to-late 1980s, was the first system to explicitly model creativity; it would move on to a new topic after a certain number of repetitions [1]. One system discussed in this paper, Rafael Pérez y Pérez's MEXICA, goes further and uses an explicit model of the creative process to reason about which plot elements will be included.

"Story generators," as they are often called, have been strongly focused on the content level (what happens in the story) rather than on the level of expression (how the story is told). For this reason, they can be more precisely characterized as *plot generators*. Some work has been done on shaping the expression of a story: Pauline

by Eduard Hovy [2] is an important early example. The other system discussed in this paper, Nick Montfort's nn, focuses more closely on variations in discourse that are specific to narrating. Some systems can work to some extent at both the story and discourse level, but we know of no previous efforts to integrate a well-developed story generation system and a well-developed system for narrating. This seems to be a significant oversight, given that literary study and narrative theory have elaborated for decades on how important both the story level and discourse level are for interesting storytelling.

In this paper, we describe the initial integration of MEXICA and the nn Narrator, systems that are each specialized to deal with one of the two essential levels of story. The two systems have been incorporated as subsystems of MEXICA-nn in a pipelined architecture. A plot representation originates from MEXICA, is handed to a middleware component that adds some information and transforms it into the type of representation required by nn, and, with the use of a manually specified plan for narrating, the story is finally told in a specific way by the nn Narrator.

We intend this as a first step toward effective and aesthetically pleasing automatic story generation which employs automatic narration. However, neither of us believes that the current pipelined scheme is really a good representation of the creative process. We believe it is very rare for fiction writers and other creative storytellers to develop everything at the story level first, in complete isolation from the language of the discourse level, and to determine how the telling will be done only at a later step — never revising, never reimagining. Our ultimate goal is to integrate these two systems in a different way, using a blackboard architecture. This would allow the plot representation and plan for narrating to be updated by either MEXICA or nn. Although MEXICA would primarily work at the story or plot level and nn at the discourse or expression level, both subsystems could suggest changes and collaboratively develop the final narrative. While we have learned some interesting things in developing our initial pipelined system, we hope that it will mainly be useful because it will inform this future work.

2 The Plot Generator MEXICA

The subsystem that deals with plot generation is MEXICA [3]. It was inspired by the engagement-reflection (E-R) cognitive account of writing as creative design [4]. MEXICA produces plots of stories about the Mexicas, the inhabitants, in centuries past, of what is now México City. (These people are sometimes inaccurately called the Aztecs.) During engagement the system generates sequences of actions guided by rhetorical and content constraints; during reflection, the system breaks impasses, evaluates, and, if necessary, modifies the material generated so far. Then, the system switches back to engagement and the cycle continues until the narrative is finished.

Creativity is one of the most astonishing characteristics of humans beings. It is a research area in the intersection of disciplines including cognitive science, artificial intelligence, psychology, and design. The core motivation behind the engagement-reflection model of computational creativity that informs MEXICA is to contribute to the study of the creative process, in particular in the area of plot generation.

One of the main problems in plot generation is the production of coherent sequences of actions. Traditionally, storytellers have dealt with this by employing explicit goals of characters and predefined story structures. These approaches helped to reduce the coherence problem. But they required that core parts of the stories be defined by the designers beforehand rather than being generated by the system. As a result, these programs and their outputs suffered from rigidity. MEXICA is an answer to the necessity of more flexible plot generators.

MEXICA is formed by two main blocks: the construction of knowledge structures (the K-S module) and the generation of plots through engagement-reflection cycles (the E-R module). The K-S module takes as input two text files defined by the user: a dictionary of valid story-actions and a set of stories known as the Previous Stories. The dictionary of story-actions includes the names of all actions that can be performed by a character within a narrative along with a list of preconditions and postconditions for each. In MEXICA all preconditions and postconditions involve emotional links between characters (e.g., jaguar knight hates the enemy). The Previous Stories are sequences of story actions that represent well-formed narratives. With this information the system builds its knowledge base. The E-R module takes two main inputs: the knowledge-base and an initial story-action (e. g., the princess heals jaguar knight) that sets in motion the E-R cycle. Once the E-R cycle ends the system perform the final analysis, a process that touches up the generated plot.

MEXICA is implemented in Pascal/Delphi. The system includes a text editor where the user defines in text files the dictionary of valid story-actions and the group of previous stories. To produce a new plot, the system provides an interface where the user introduces an initial action that triggers the engagement-reflection cycle. MEXICA creates two reports: 1) a detailed trace of the engagement-reflection cycle, which includes the final plot; 2) a representation of knowledge structures. Both reports can be visualized within the program. The implementation includes 21 modifiable parameters that control different aspects of the E-R cycle, allowing the user to experiment with the model.

MEXICA shows that the E-R model of computational creativity allows the generation of novel and interesting plots [5]. The program uses a representation of emotions to associate possible actions and continue the story in progress. In this way, MEXICA contributes to the study of the role of emotions during the creative process [6]. The program includes a process to evaluate the interestingness of the story it creates. As far as we know, no other plot generator is capable of this type of evaluation.

3 The nn System and Its Narrator Module

The subsystem that deals with the expression level is nn, a system originally created to allow the implementation of interactive fiction (that is, text adventure games) with narrative variation [7]. The components of nn that are specific to interactive fiction are not used in MEXICA-nn; the module that is used is one that has been central to research in narrative variation: the Narrator.

Enabling new types of creative writing and programming is the main purpose of nn. Specifically, nn is designed as an interactive fiction system, to enable general, simple manipulation of the discourse level. It is already possible to place characters and objects in various locations in an interactive fiction worlds, to animate them and have them react. The eventual release of nn will hopefully provide author/ programmers of interactive fiction with discourse-level facilities that are similar to the existing story-level capabilities of state-of-the-art systems such as TADS 3 and Inform 7. While much work in IF has dealt with creating intelligent or emotional characters, nn seeks to do something different: to create more interesting narrators.

For the purposes of integrating MEXICA and nn, many modules of nn have been left out. The ones dealing with tokenizing and parsing input, for instance, have been omitted, as there is no user input in MEXICA-nn. The one that simulates the fictional world is also not necessary, because in MEXICA-nn the events of the story are simply provided by MEXICA and do not need to be simulated within an interactive fiction framework. Others dealing with specialized functions that do not apply to the narrating task (such as the "meta" operations of *restart*, *undo*, *quit*, and so on) have also been omitted. What is left is the core module of the nn research project, the Narrator, and the two representations that are essential to narrating: the focalizer worlds and the plan for narrating.

The Narrator module has an internal architecture. It is structured as a three-stage pipeline, using a model that is typical of natural language generation systems. In the reply planner, the high-level structure of the reply (or, in the case of MEXICA-nn, of the entire output document) is determined. The narrative parameters and the reply structure determined in this first stage are sent to the microplanner, which determines the grammatical specifics based on those inputs. Finally, the detailed specification for output is converted into text by the realizer, which does lexicalization.

nn is implemented in Python 2.5. To create a work of interactive fiction, an author/ programer writes a game file which is simply Python code that makes use of pre-defined classes and assigns values to certain expected variables. The existents in the fictional world are represented as a dictionary of objects (each of class Actor, Room, or Thing) and the actions as a list of Action objects, each of which contains one or more lower-level Event objects.

Several important types of narrative variation have been modeled in the system. Specifically, one significant aspect of narrative variation has been implemented within each of the five major categories described in *Narrative Discourse* [8]. New

theoretical ideas have been developed as a result of these implementations. One is that in generating text, order and time of narrating must be considered jointly even though Genette has placed them in separate categories. Another is that simple sequences are inadequate to represent order, while an ordered tree provides a good representation. Other results pertain to narrative distance (which may be best seen as a composition of other effects) and the worlds representing perception by focalizers (which are fundamentally different from the simulated world).

4 A First Pass at Connecting the Systems

A MEXICA-nn narrative is generated by running a modified MEXICA to generate XML as defined by the MEXICA plot schema. An invocation of the middleware component of the system then takes this plot representation and elaborates it, transforming it to a different XML representation of existents and actions that conforms to the nn schema. Finally, nn is invoked with a specific plan for narrating to produce a narrative.

The middleware maps high-level plot elements, as output by MEXICA, to more specific sets of actions, each of which must include a sequence of one or more events. For instance, the MEXICA action FAKED_STAB_INSTEAD_HURT_HIMSELF is mapped to two actions in nn; the first one has FEINT as its verb and the second one has INJURE as its verb. The FEINT action has a MISC event within it. This is the type of event used to represent something that does not have a physical consequence or involve the application of force. The INJURE action is best represented with a IMPEL event (representing the character's application of force to himself) followed by a MODIFY event (representing that the character's health was affected). These detailed events can come into play if the speed of narration is slowed down and every detail of what happens is being represented by the narrator.

In one run, the modified version of MEXICA used as a subsystem of MEXICA-nn produced the following XML representation of a plot:

```
<story>
 <existents>
  <location name="Lake"/>
  <location name="City"/>
  <character name="HUNTER" gender="male"  location="Lake" />
  <character name="JAGUAR_KNIGHT" gender="male"  location="Lake" />
  <character name="VIRGIN" gender="female"  location="Lake" />
 </existents>
 <actions>
  <action verb="ACTOR" agent="JAGUAR_KNIGHT" time="1" />
  <action verb="ACTOR" agent="VIRGIN" time="2" />
  <action verb="ACTOR" agent="HUNTER" time="3" />
  <action verb="WERE_RIVALS" agent="JAGUAR_KNIGHT" object="HUNTER"
   time="4" />
  <action verb="FAKED_STAB_INSTEAD_HURT_HIMSELF"
   agent="JAGUAR_KNIGHT" object="HUNTER" time="5" />
  <action verb="CURED" agent="VIRGIN" object="JAGUAR_KNIGHT"
```

```
   time="6" />
  <action verb="MUGGED" agent="JAGUAR_KNIGHT" object="VIRGIN"
   time="7" />
  <action verb="WENT_TENOCHTITLAN_CITY" agent="JAGUAR_KNIGHT"
   NewLocation="City" time="8" />
  <action verb="HAD_AN_ACCIDENT" agent="JAGUAR_KNIGHT" time="9" />
  <action verb="DIED_BY_INJURIES" agent="JAGUAR_KNIGHT" time="10" />
 </actions>
</story>
```

This representation was translated by the middleware into an elaborated XML representation that used the nn schema and specified finer-grained events. The modified version of the nn Narrator that was used was then able to produce a variety of different narrations. This one was accomplished using chronological order, setting the temporal position of the narrator to a point before the events (hence, the narration takes place in the future tense), and adding a filter that expresses surprise:

> The jaguar knight will oppose the hunter!
>
> Meanwhile, the hunter will oppose the jaguar knight!
>
> Then, the jaguar knight will feint at the hunter, dude!
>
> Then, the jaguar knight will strike himself!
>
> Then, the virgin will heal the jaguar knight! Awesome!
>
> Then, the jaguar knight will threaten the virgin, man! Awesome!
>
> Then, the jaguar knight will rob the virgin!
>
> Whoa, then, the jaguar knight will go to city!
>
> Then, the jaguar knight will slip!
>
> Then, the jaguar knight will fall!
>
> Then, the jaguar knight will die!

The next narration was produced using settings for retrograde order (so that the most recent event is narrated first) and with time of narrating set to be simultaneous with the last event. Although this is a simple rearrangement out of chronological order, it shows the generation of different tenses that are necessary for more nuanced orderings, such as occasional flashbacks. The final action, told first, is narrated in the present tense and the ones leading up to it are told in the past tense:

> The jaguar knight dies.
>
> Before that, the jaguar knight fell.
>
> Before that, the jaguar knight slipped.
>
> Before that, the jaguar knight went to city.
>
> Before that, the jaguar knight robbed the virgin.
>
> Before that, the jaguar knight threatened the virgin.
>
> Before that, the virgin healed the jaguar knight.
>
> Before that, the jaguar knight struck himself.
>
> Before that, the jaguar knight feinted at the hunter.
>
> Before that, the hunter opposed the jaguar knight.
>
> Meanwhile, the jaguar knight opposed the hunter.

A straightforward present-tense narration can be generated where the narrator speaks hesitantly (a stylistic variation that isn't actually specific to narrative) and where all the details of each action are told as slowly as possible:

> The jaguar knight, um, starts, er, to rival the hunter. The jaguar knight, um, opposes the hunter.
>
> Meanwhile, the hunter starts to rival the jaguar knight. The hunter opposes the jaguar knight.
>
> Then, the jaguar knight starts to feint the hunter. The jaguar knight, um, feints at the hunter.
>
> Then, the jaguar knight starts to injure himself. The jaguar knight strikes himself.
>
> Then, the virgin starts to, er, cure the jaguar knight. The virgin heals the jaguar knight.
>
> Then, uh, the jaguar knight, um, starts to threaten, uh, the virgin. Uh, the jaguar knight, er, threatens the virgin.
>
> Then, the jaguar knight starts to rob the virgin. Um, the jaguar knight robs the virgin.
>
> Then, the jaguar knight starts to go. The jaguar knight goes to, um, city.
>
> Then, the jaguar knight starts to, uh, slip. The jaguar knight slips.
>
> Then, the jaguar knight starts to fall. The jaguar knight falls.
>
> Then, the jaguar knight starts to, uh, die. The jaguar knight dies.

5 Differences in Representation of Action and Story

Some important differences between MEXICA's representation and the nn Narrator's were revealed by integrating the two systems. nn combines a representation of actions and existents with a plan for narrating, while MEXICA, focused on the plot level, lacks the latter type of representation. But nn's representation of actions and events is also quite different from MEXICA's. One mismatch is that no Things (the class used in nn for inert items or props) are represented explicitly by MEXICA; it is simply implicit in an action involving stabbing that the character undertaking the action has an edged weapon of some sort by the time an action is carried out. The character may possess that weapon to begin with or acquire it in the course of the story. Which of these happens and how exactly it happens needs to be added during elaboration for nn's Narrator to able to narrate in the best possible way.

Additionally, the granularity of action is different in the two systems. A single MEXICA plot element may correspond to multiple high-level actions in nn. For instance, a reflexive MEXICA action involving A and B, such as FIGHT_EACH_OTHER, corresponds to two simultaneous actions in nn, one with A as agent and one with B as agent. A plot element that corresponds to a compound action maps to a sequence of actions that occur one after the other in nn. Each of the low-

level events within these actions must also be specified to indicate each blow of the battle. nn's more numerous and finer-grained actions are difficult to specify, but when they are in place, they are very useful for creating narrative variation of different sorts, since each event may be included or excluded and, if told at all, may be told more or less rapidly. Some of the work of mapping MEXICA's plot representation to an nn story representation is straightforward; other work involves creativity and authorship. The more creative work involves determining what specific sequence of blows occurs in a fight, what weapon a character uses to attempt to stab another character, and how a particular weapon or other object is acquired by a character. In the future, we plan to develop MEXICA-nn to accomplish these things not with fixed mappings but rather in a way that is consistent with MEXICA's model of the creative process.

On the other hand, the nn Narrator's representation is not more detailed in every way. MEXICA uses a model of character's emotional affinities for one another in determining the plot. This model could be used in narrating, too. To provide a simple example: a narrative could be produced in which actions are elided or passed over very quickly when they are undertaken by characters that jaguar knight dislikes. The resulting narrative would favor jaguar knight by emphasizing the actions of his friends, or, more precisely, the people who he likes at that point in time. MEXICA also uses particular techniques to generate interesting plots. Unlikely plot elements are proposed by the system and, using abductive reasoning, the system tries to find explanations for them in the form of other plot elements. For instance, if the princess loves jaguar knight at one point and then later intentionally kills him, it may be because she discovered that he killed her father. If information about the plot generation process was available to the Narrator, the narration of the story could follow the trace of the creative process, describing the unlikely event first and then, in a flashback, explaining why it came to pass.

As we have discussed, the type of representation used is substantially different for the nn Narrator and for MEXICA. To some extent, one system might benefit from having a finer-grained representation that is motivated by the other. But this is unlikely to be the best answer in every case. In many ways, MEXICA takes a higher-level view of actions and existents for a good reason. A complete computer model of story generation, from plot generation to narrating, is likely to require different types of knowledge structures at different levels of abstraction (or, knowledge representations that can be easily manipulated at different levels of abstraction) interacting together in order to generate its final output.

MEXICA employs three types of knowledge representations at three different levels of abstraction: the Tensional representation (graphics of the dramatic tension of the story in progress), the Atoms (groups of emotional links and tensions between characters) and the Concrete representation (a copy of the story actions and the previous stories). The Concrete representation includes more precise information than the Atoms; and the Atoms include more precise information than the Tensional

representation. The engagement-reflection cycle employs these three knowledge structures during plot generation.

The nn Narrator, on the other hand, deals with a representation of the telling that is absent from MEXICA (the plan for narrating), a representation of characters' perceptions (the focalizer worlds), and a representation of the "real world" as far as the fiction is concerned (the world model, with actions and existents). The last of these is the point of overlap with MEXICA, but even then, as we have described, the way actions are represented is not the same and the same existents are not present in both systems.

We find it appealing to develop a system that goes upwards and downwards through concrete and abstract knowledge representations during the weaving and telling of a story. A more general representation of an action (e.g., jaguar knight and the enemy fight each other) is suitable during initial enplotment, when a general framework for the narrative is being put together; a detailed description of a fight (e.g., types of weapons used, particular blows) allows the system to very the telling to a greater extent, since there are more choices of what to omit and what to include and the actions can be arranged in a wider variety of ways. As we continue development, we will look to develop a system that employs representation at both levels to generate a plot and a particular telling. The middleware described in this paper is the first step in this phase of the work.

6 Implications and Future Work

There are other steps to be taken beyond modification of knowledge representations and a better pipelined integration. Currently, when MEXICA-nn is invoked the user specifies a particular plan for narrating and the pre-defined parameters of that plan are simply applied to the story. A better approximation of the creative process would involve having MEXICA-nn choose a plan that is appropriate to the story material. For instance, a story of a character's steady downfall might be narrated chronologically, while a story with many reversals might be told using flashbacks to connect the most interesting actions with each other.

We hope that plans for narrating can eventually be developed at a deeper level, co-evolving with the plot. In a blackboard system, nn could propose a plan for narrating of a specific sort — for instance, sylleptic narration in which everything that happens in the city is told first, then everything that happens in the forest, and then everything that happens by the lake. nn could request that a larger number of interesting actions occur in the forest. MEXICA could then work to satisfy this request, either by changing some plot elements so that they occur in different location or by adding plot elements. Similarly, MEXICA could register requests regarding the plan for narration, asking that plot elements be told in a certain order or that certain existents or plot elements be emphasized.

This paper has described the initial integration of MEXICA and the nn Narrator to produce a complete computer based storyteller, MEXICA-nn. This work has led us to reflect on the importance of manipulating represented knowledge at different levels of abstraction when developing computer models of creativity. We hope this paper encourages people to integrate systems. In the short term, it is a source of important information about meaningful differences between systems that might be difficult to obtain in another way. In the long term, we believe integrating well-developed systems will be essential in tackling complex problems such as the invention and telling of stories, which require specialized creative processes that must also interact and depend upon one another.

References

[1] Wardrip-Fruin, N. (2009). *Expressive Processing*. Cambridge: MIT Press.

[2] Hovy, E. (1988) *Generating Natural Language under Pragmatic Constraints*. Hillsdale, N.J.: Lawrence Erlbaum, 1988.

[3] Pérez y Pérez, R. & Sharples, M. (2001) MEXICA: a computer model of a cognitive account of creative writing. Journal of Experimental and Theoretical Artificial Intelligence. Volume 13, number 2, pp. 119-139

[4] Sharples, M. (1999). *How we write: Writing as creative design*. London: Routledge.

[5] Pérez y Pérez, R. & Sharples, M. (2004) Three Computer-Based Models of Storytelling: BRUTUS, MINSTREL and MEXICA. Knowledge Based Systems Journal. Vol. 17, number 1, pp. 15-29.

[6] Pérez y Pérez, R. (2007). Employing Emotions to Drive Plot Generation in a Computer-Based Storyteller. Cognitive Systems Research. Vol. 8, number 2, pp. 89-109.

[7] Montfort, N. (2007). "Generating Narrative Variation in Interactive Fiction." PhD dissertation, University of Pennsylvania. <http://nickm.com/if/Generating_Narrative_Variation_in_Interactive_Fiction.pdf>

[8] Genette, G. (1980). *Narrative Discourse: An Essay in Method*. Trans. J. E. Lewin. Ithaca, NY: Cornell.