

# The Three Layers Evaluation Model for Computer-Generated Plots

Rafael Pérez y Pérez

Departamento de Tecnologías de la Información  
Universidad Autónoma Metropolitana, Cuajimalpa  
Av. Vasco de Quiroga 4871, Col. Santa Fe Cuajimalpa, México D. F., C.P. 05300  
[rperez@correo.cua.uam.mx](mailto:rperez@correo.cua.uam.mx), [www.rafaelperezyperez.com](http://www.rafaelperezyperez.com)

## Abstract

This paper describes a model for evaluating a computer-generated plot. The main motivation of this project is to provide MEXICA, our plot generator, with the capacity of evaluating its own outputs as well as assessing narratives generated by other agents that can be employed to enrich its knowledge base. We present a description of our computer model as well as an explanation of our first prototype. Then, we show the results of assessing three computer-generated narratives. The outcome suggests that we are in the right direction, although much more work is required.

## Introduction

The engagement-reflection (ER) computer model of writing (Pérez y Pérez and Sharples 2001) represents creativity as a constant interplay between the generation of ideas and their evaluation. As a core characteristic, such processes strongly interact and influence each other. Thus, from the ER perspective, assessment is an integral part of the creative process. In the same way, evaluation plays an essential role after the creative process has ended: i.e. following a particular criterion, it provides elements to establish the value of an agent's output. In this way, we can distinguish two different goals for the same process: 1) to contribute to the development of a story in progress; 2) to estimate if the system's output might be classified as creative. The work reported in this paper concentrates in the latter. From now onwards, we refer to a computer agent that is capable of assessing a product as evaluator. The main motivation of this project is to provide MEXICA, our plot generator, with the capacity of evaluating its own outputs as well as assessing narratives generated by other agents that can be employed to enrich its knowledge base. We can summarise it as follows: MEXICA = plot generator + evaluator.

What are the elements that need to be considered in a computer model of evaluation? In this work we present three. The following lines describe each of them.

1) A creative process generates at least two types of outputs: a final product (e.g. a solution to a problem, a poem, a story, a piece of music) and novel knowledge that expands the expertise of the creator. It is not possible to think of creativity without these two elements. Sometimes, authors engage in creative tasks with the main purpose of expanding their expertise in particular topics. For example, Picasso developed several sketches in

preparation to paint El Guernica. Based on these observations, we claim that computerised creativity (c-creativity) occurs when as a result of the creative process an agent generates knowledge that does not explicitly exist in its original knowledge-base and which plays an important role in the produced output (Pérez y Pérez and Sharples 2004); such novel knowledge becomes available within the agent's knowledge base for the generation of more original outputs (Pérez y Pérez under revision). That is, an essential aim of creativity is the generation of expertise and experience that is useful for the creative process itself. We believe that the same principle can be applied during the assessment of a narrative. A computer model of evaluation must consider if the evaluator, as a result of the assessment process, incorporates new knowledge structures into its knowledge base. This idea seems to echo the thoughts of some writers about the importance of reading. For instance, David Lodge claims that reading other authors is the best way to learn about the world and about the technical abilities required for writing (Lodge 1996). Thus, a good narrative allows discovering new perspectives in a given situation, new features that had not been seen before, novel ways of understanding a situation. In other words, it generates new knowledge in the reader.

2) The second aspect to be considered is related to the concept of story. Different authors agree that a story is defined as a sequence of actions that follow the classical Aristotelian structure: setup, conflict, complication, climax and resolution (e.g. see Claude Bremond 1996; Clayton 1996, p.p. 13-15). Usually, conflict is described as obstacles that oppose a more satisfactory state or desire. During complication, the difficulties introduced by the conflict arise incrementing the tension produced in the reader, until the climax is reached. Then, all conflicts are sorted out releasing all accumulated tensions. In other words, if one follows the Aristotelian concept of a story, a narrative must produce in the reader increments and decrements of the dramatic tension. Thus, a computer model of plot evaluation must be able to recognise if the events that comprise a narrative satisfy the Aristotelian requirements. In order to achieve this goal, one needs an agent capable of representing affective responses. (It is worth pointing out that, although in this work we adopt the Aristotelian view, there are other valid options to represent narratives).

3) The third aspect considers that an agent must be able to determine if the sequence of actions that comprise a story satisfies common sense knowledge.

In sum, a computer model of plot evaluation requires a story to be evaluated, and an agent capable of transforming the sequence of actions that comprise the story into internal representations that allows detecting novel knowledge structures (cognitive changes), its coherence (common sense knowledge) and representing increments and decrements of the dramatic tension of the tale (affective responses). In the same way, it is necessary to determine how these components influence each other.

This type of model requires an agent's knowledge-base that represents the experience of the evaluator: a structure is novel when it does not previously exist in its knowledge-base; the information necessary to evaluate the coherence and the story's tension resides within this repository. Thus, different agents with different knowledge and beliefs should produce different evaluations of the same product. Even the same agent, if its knowledge base is modified, might produce different evaluations of the same product. The following lines describe a computer model for plot evaluation that subscribes to these ideas. It is built on top of the results we obtained from previous research on this topic.

## Related Work

Ritchie (2007) suggests criteria for evaluating the products of a creative process (the process is not taken into consideration); in general terms such criteria evaluate how typical and how valuable the product is. The goal is, using existing evaluations of typicality (and atypicality) and value, to construct more complex criteria. Colton (2008) considers that skill, imagination and appreciation are characteristics that a computer model needs to be perceived to have (see also Pease et al. 2001). Jordanous (2012) employs a group of human experts to develop criteria for evaluation of a computer generated product. It includes characteristics like Spontaneity and Subconscious Processing, Value, Intention and Emotional Involvement, and so on. All these are interesting ideas, although some are too general and difficult to implement (e.g. see Pereira et al. 2005). Some work has been done in evaluation of plot generation. Peinado et al. (2010) also have worked in evaluation of stories, although they work was oriented to assess novelty. I am not aware of any model of plot generation that includes the characteristics of the present work.

In your review of related work, Ritchie's criteria aren't merely evaluating how typical/valuable products are, but using existing evaluations of typicality (and atypicality) and value to construct more complex criteria. Also, although Jordanous's case study example uses human expert evaluations to evaluate different criteria, she does not insist that her criteria are measured by human experts - quantitative/automated tests could also be used.

## Our Plot Generator

Our research in generation and evaluation of narratives is based on the MEXICA agent (Pérez y Pérez and Sharples 2001; Pérez y Pérez 2007). We claim that, as a

result of engagement-reflection cycles, our storyteller produces plots that are novel, coherent and interesting. MEXICA employs a dictionary of story-actions and a set of Previous Stories, both defined by the user as text files, to construct its knowledge base. Story-actions have associated a set of preconditions and post conditions that represent common sense knowledge. For example, the precondition of the action character A heals character B is that B is injured or ill. Otherwise, the action does not make sense.

In MEXICA, a story is defined as a sequence of actions that follows the next format: character performing the action, description of the action, object of the action (another character); for instance, the jaguar knight attacked the enemy. The format allows some variations, e.g. only one character performing an action; for instance, the princes went to the forest. We refer to this way of organising a narrative as MEXICA's format. The Previous Stories represent well-constructed narratives and provide information about how the story-world works. They represent the experience and knowledge of the agent. Any new story generated by MEXICA can be added to the Previous Stories.

The Contextual Structures are the main representation of knowledge within the system. They associate emotional links and tensions between characters with logical actions to perform. For instance, a Contextual Structure might register that when a character A is in love with a character B (an emotional link between two characters) something logical to do is that A buy flowers to B, or that A serenades B, and so on. Contextual Structures are built from the set of Previous Stories; later, they are employed to generate new outputs during plot generation. Employing the same process, knowledge structures can be built from any new story created by the system or by any other agent (as long as the story follows the MEXICA's format).

Tensions represent conflicts between characters. When the number of conflicts grows the value of the tension rises; when the number of conflicts decreases the value of the tensions goes down; when the tension is equal to zero all conflicts have been solved. Thus, the storyteller keeps a record of the dramatic tension in the story. The following are examples of situations that trigger tensions: when the life of a character is at risk; when the health of a character is at risk; when a character is made a prisoner; and so on. Every tension is assigned a value. So, each time an action is performed by a character the system calculates and records the value of all active tensions. With this information the storyteller is able to graph the curve of tension of the story. Such a curve is referred to as the Tensional Representation.

## Description of the Model

The work reported in this paper employs and extends the results we obtained in previous efforts to understand automatic plot evaluation. The approach we have followed is to break this complex problem into relatively simpler sub problems. Thus, we developed a computer model for assessing novelty (Pérez y Pérez et al. 2011) and a computer model for assessing interestingness (Pérez y Pérez and Ortiz 2013) as first steps before building

the integral model of evaluation (we did not publish the result of our model for assessing coherence). Based on those results, we came out with a general model that I present here. The following lines provide a general view of this work.

We exploit the infrastructure built for MEXICA. Thus, a dictionary of story-actions and a set of Previous Stories, both defined by the user as text files, are used to construct the evaluator's knowledge base. It is interesting to notice that our agent employs the same information to generate a plot and to evaluate a plot.

We have been successful in developing tools that are capable of transforming a sequence of actions (i.e. a story in MEXICA's text format) into internal structures that our computer agent can manipulate. Employing such tools, it is possible to perform an analysis of the dramatic tension of the story under evaluation and of the changes that such a plot produces into the agent's knowledge structures. I refer to the process of transforming a sequence of actions into structures that represent knowledge and affective reactions as *Interpretation* (see figure 1).

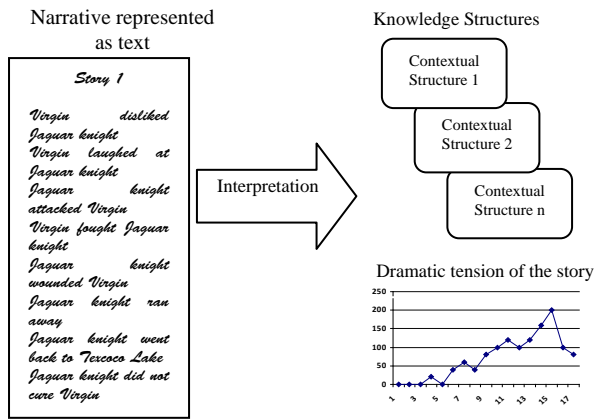


Figure 1. The Interpretation Process transforms a sequence of actions in a text format into a set of knowledge structures and affective reactions (dramatic tension).

Once the interpretation has been performed the agent has the necessary information to analyse the attributes of the story under assessment. Based on our previous work, we have selected a set of eight features, known as the story-characteristics, which are useful for evaluating a plot: opening, closure, climax, reintroducing complications, satisfaction of preconditions, repetition of sequences of actions and two types of novel knowledge structures. Typically, they have a value ranging from zero to one, where one is the most desirable value. They represent knowledge structures and affective reactions. Details of the story-characteristics are given some lines ahead.

In order to implement the model for assessing the novelty it was necessary to choose a set of story-characteristics that were associated to the production of original plots; the same applies for the model of evaluation of interestingness and coherence. Some story-characteristics are used in more than one of those systems. For instance, sorting out all the problems that characters have at the end of the story (correct closure of the narrative) is important for both, the model of coherence

and the model of interestingness; the generation of unusual situations (new knowledge structures) is important for the model of interestingness and the model of novelty; and so on.

It is possible to employ the three models mentioned above to obtain a global evaluation of a story. That is, given a plot, we can run the system that evaluates novelty, then the system that evaluates interestingness and lastly the system that evaluates coherence; finally, we can calculate the average result. However, this procedure has some flaws. As mentioned earlier, some story-characteristics are employed in more than one model. As a result, they might be overrepresented in the overall calculus distorting the final value. In the same way, story-characteristics might be linked in ways that individual models cannot represent. For instance, one story might get a high score in novelty but a low score in coherence. However, it does not make sense to claim that a story is very original when it is unintelligible. A famous example of a similar situation is the sentence "Colorless green ideas sleep furiously" (Chomsky, 1957); this sentence does not seem mean anything coherent but sound like an English sentence. Thus, it seems sensible to have one model for a general evaluation, where all story-characteristics can interact, rather than three individual ones.

Some of the story-characteristics, although useful, are not essential for a good plot. So, if they are present they help to enhance the story; if not, the story still can be a good narrative. We referred to such characteristics as Enhancers. For instance, if the problems of a character seem to be solved and out of the blue new conflicts arise (reintroducing complications) the plot might be considered as more exciting. This characteristic is not required to develop a good plot but its presence helps. So, Enhancers add extra points to the evaluation. The use of Enhancers might be conditioned to the good results of other characteristics. For instance, if a given story is unoriginal it does not make sense to consider it more interesting only because there is a reintroduction of complications. Following the same logic, the model contemplates the use of Debaserers, i.e. story-characteristics that, when they are missing, they decrement in some points the global evaluation of a plot.

In our previous models the relationships of the story-characteristics are defined by expressions like the following:

$$E = C1W1 + C2W2 + C3W3 + \dots CnWn$$

where E represents the result of the evaluation, C one of the characteristics to be assessed and W its weight. However, this expression lacks flexibility. For example, it is not possible to represent conditioned Enhancers or Debaserers. In the same way, some characteristics might play a more relevant role during one stage of the assessment than during others. For example, a story must be lucid; otherwise, it is not worth evaluating the plot. So, at this point those characteristics associated to coherence have a high priority for the evaluation process. However, once this requirement is satisfied, other characteristics start to

take precedence. To illustrate this situation the reader can picture a logic story that is boring, i.e. it lacks increments and decrements of tension. In this case, those characteristics associated to interestingness became more relevant for the evaluation process. As a result, the global assessment probably would produce a low value even if the coherence is pretty good.

The model also considers what we refer to as the compensation effect. In the overall evaluation, characteristics highly rated might compensate those with lower grades by adjusting their weights. For example, picture a story that shows exceptional original situations; even if the plot suffers for some coherence problems, the overall rate might still be pretty high.

## Description of the Story-Characteristics

The following lines describe the story-characteristics that I employ in this work and how to calculate their values.

**Opening:** We consider that a story has a correct opening when at the beginning there are no active dramatic tensions in the tale and then the tension starts to grow. If at the beginning of the story the value of the tension is zero, then Opening is set to one; if at the beginning of the story the value of the tension is equal to the main peak (the climax), then Opening is set to zero; otherwise, Opening is set to a proportional value between zero and one.

$$\text{Opening} = 1 - (\text{Tension at the first action} / \text{Peak})$$

**Closure:** We consider that a story has a correct closure if all the dramatic tensions in the story are solved when the last action is performed. That is, following Pérez y Pérez and Sharples, a story “should display an overall integrity and closure, for example with a problem posed in an early part of the text being resolved by the conclusion” (Pérez y Pérez and Sharples 2004). If at the end of the story the value of the tension is equal to the main peak (the climax) then Closure is set to zero; If at the end of the story the value of the tension is equal to zero (all problems are solved), then Closure is set to 1; otherwise, Closure is set to a proportional value between zero and one.

$$\text{Closure} = 1 - (\text{Tension at last action} / \text{Peak})$$

**Climax:** All stories should include a climax. In the graphic of tensions the climax is represented by highest peak. However, it is not the same a story with an incipient peak that a story with a clear elevated crest. In order to evaluate the peak, MEXICA calculates the average value of all Previous Stories’ climax and employs it as a reference. Thus, if the peak’s value is equal or major than the reference, then Climax is set to 1; if there is no peak, then Climax is set to zero; otherwise, it is set to a proportional value between zero and one.

$$\text{Climax} = (\text{Current climax} / \text{Reference value climax})$$

If Climax > 1 then Climax = 1

**Reintroducing Complications:** We refer to the situation where a narrative has a resolution and then tensions start to rise again as reintroducing-complications. In this work, we appreciate narratives that seem to end and then new problems for the characters emerge, i.e. where all tensions are solved and then they rise again. This formula can be observed in several examples of narratives like films, television-series and novels. MEXICA calculates the average number of complications that are reintroduced in the Previous Stories and employs it as a reference. Thus, if the number of times that the current story reintroduce complications is equal or major than the reference, then Reintroducing Complications is set to 1; if there is no reintroduction of complications, then Reintroducing Complications is set to zero; otherwise, it is set to a proportional value between zero and one.

**Novel Contextual Structures:** In this work a new story generates new knowledge when it generates structures that did not exist previously in the knowledge base of the system and that can be employed to build novel narratives. Each action within a plot has the potential of introducing an unknown context for the agent. So, if all actions that comprise the story under evaluation generate unknown contexts, then Novel Contextual Structures is set to one; if none of the actions produce an unknown context, then Novel Contextual Structures is set to zero; otherwise, Novel Contextual Structures is set to a proportional value between zero and one.

**Original Value:** Besides calculating the number of novel contextual structures, it is necessary to determine how original they are with respect to the information that already exists in the knowledge base. With this purpose we define a parameter known as the Limit of Similitude (LS) that represents the maximum percentage of likeness allowed between two knowledge structures. If the percentage of similitude between a given Contextual Structure and all structures in the knowledge base is minor to LS, we refer to such Contextual Structure as original. In this way, we can distinguish between novel situations and really original ones. Thus, the Original Value is equal to the ratio between the total number of original structures and the total number of contexts produced by the tale.

**Preconditions:** All actions have associated preconditions that represent common sense knowledge. If the preconditions of all story actions are fulfilled, then Preconditions is set to one; if none of the preconditions of all story actions are fulfilled, then Preconditions is set to zero; otherwise, it is set to a proportional value between zero and one.

**Repetition of Sequences:** There are some attributes that contribute to the lack of coherence in a plot. The repetition of sequences of actions performed by the same characters illustrates this situation. We include this feature to show some of the problems that computer generated narratives might suffer. Thus, in this implementation, Repetition of Sequences is set to one when there are no repetitions; otherwise, it is set to zero.

## The Three-Layers

The model described in this paper represents evaluation as a process organized in three layers (see figure 2).

Layer-0 includes those characteristics that a plot must satisfy in order to be considered for evaluation. These characteristics do not add points to the evaluation; they are requirements that need be satisfied in order to proceed to evaluate the plot. Otherwise, the process is ended. They are known as the *required-characteristics*.

Layer-1 includes what I refer to as the *core-characteristics*. They are the backbone of the evaluation process and represent those essential features that form a plot.

Layer-2 includes what I refer to as the *Enhancers* and the *Debasers*. Enhancers are characteristics that add extra points to the result obtained from the previous layer. Debasers represent features that decrement the result obtained from Layer-1. Their use might be conditioned to the result of other story-characteristics.

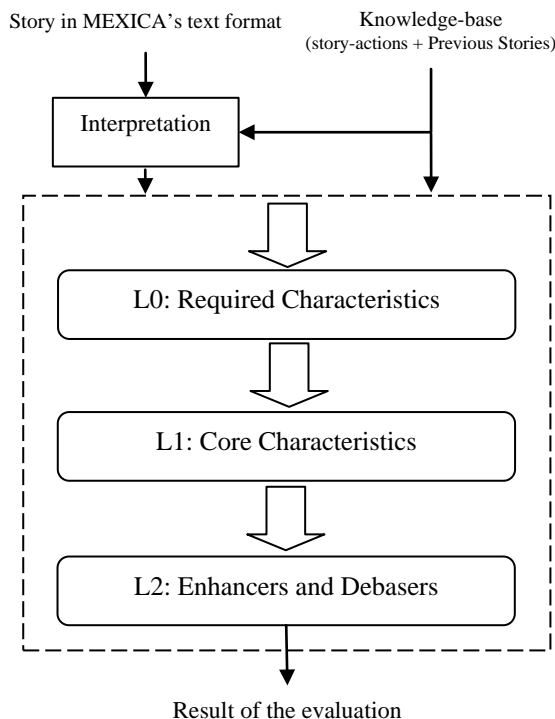


Figure 2. The three layers evaluation model.

A story-characteristic can be employed in more than one layer. Actions' preconditions illustrate this situation: it is not worth to evaluate an unintelligible story (Preconditions in Layer-0); however, a mainly sounded story with few inconsistencies might only be penalized with some negative points (Preconditions in Layer-2).

The following lines provide details about the implementation.

**Layer-0:** In the current implementation, the number of Fulfilled Preconditions and the number of Novel Contextual Structures are selected as the Required-Characteristics. If most actions within a story have unfulfilled preconditions or the story under evaluation is too

similar to any of the previous stories, then the systems considers that is not worth evaluating the plot. The user provides the minimum rates that the story-characteristics Fulfilled Preconditions and Novel Contextual Structures must reach to continue with the evaluation process.

**Layer-1:** In the current implementation, the following elements have been selected as the core-characteristics: Climax, Closure and Novel Contextual Structures. All they have been assigned the same weight. These characteristics have been chosen because: a narrative without climax is not a story; Closure is important to keep the coherence and interestingness of the tale; novelty is an essential feature of any story. The result of the evaluation in Layer-1 is the average value of the three core-characteristics.

**Layer-2:** In the current implementation, Preconditions and Repeated Sequences have been chosen as Debasers. They represent features that we take for granted; however, if they are missing within a narrative we immediately notice them. Thus, if they have a value lower than a reference provided by the user, the result of the evaluation obtained in Layer-1 is decremented by  $n$  units, where  $n$  is a parameter defined by the user.

```

IF Preconditions < Reference-Preconditions THEN
    Decrement-Result-Evaluation-1
IF Repetition-Sequences < Reference-RS THEN
    Decrement-Result-Evaluation-1

```

The following characteristics have been chosen as Enhancers: Opening, Reintroducing Complications and Original Value. Thus, if they have a value higher than a reference provided by the user, then the result of the evaluation obtained in Layer-1 is incremented by  $m$  units, where  $m$  is a parameter defined by the user. Enhancers are only employed when there are not repetition of sequences of actions, the evaluation in Layer-1 and the Closure reach a minimum value defined by the ser.

```

IF (Repetition-sequences = 1) and (Result-Layer-1 > Reference-L1) and (Closure > Reference-Closure) THEN
    BEGIN
        IF Opening > Reference-Opening THEN
            Increment-Result-Evaluation-1;
        IF Reintroducing-Complications > Reference-RC THEN
            Increment-Result-Evaluation-2;
        IF Original-Value > Reference-OV THEN
            Increment-Result-Evaluation-3;
    END

```

As a final step, the evaluator generates a report to explain the criteria employed during the process of evaluation. The report is divided in four sections: section one includes a general comment about the whole narrative; section two provides observations about the story's coherence; section three incorporates notes about the story's interestingness; and section four offers comments about the narrative's novelty.

The report is generated by matching the value of some of the story-characteristics with predefined texts. In general, there are at least five possible options that can be employed for each of such story-characteristic.

```

IF Value-Story-Characteristic > 0.9 THEN
  Employ-Text-1
ELSE IF Value -Story-Characteristic > 0.8 THEN
  Employ-Text-2
ELSE IF Value -Story-Characteristic > 0.7 THEN
  Employ-Text-3
ELSE IF Value -Story-Characteristic > 0.6 THEN
  Employ-Text-4
ELSE Employ-Text-5;

```

The following lines describe the way each section is built.

**Section one.** The system employs the final result of the evaluation process (output of Layer 2) to select the right text.

**Section two.** The coherence section includes three types of comments: one associated to the satisfaction of preconditions, one related to the right closure and the last one connected to the repetition of sequences of actions. The first two types of comments are always printed; the last type of comment is omitted when the tale does not include repeated sequences of actions. Thus, the system employs the story-characteristics Preconditions, Closure and Repetition of sequences to generate the text.

**Section three.** The interestingness section includes five types of comments, each one related to the following story-characteristics: Opening, Climax, Reintroducing complications, Closure and Original value. The first two comments are always included in the report while the last three comments are only printed when some requirements are satisfied. The next lines explain the conditions that need to be satisfied in order to incorporate the last three remarks into the report. If the story-characteristic Climax  $\geq 0.7$  then the system adds comments about the closure. This makes sense because the climax represents the conflicts in the story and the closure indicates how those conflicts are sorted out.

If the story-characteristic Closure  $\geq 0.7$  then comments regarding the original value are inserted in the report. That is, the system only includes comments about singular features of the plot when it has an adequate ending. That is, in the current implementation originality loses importance when the story has a bad finale.

If the story-characteristic Closure  $\geq 0.7$  and the Reintroduction of complications  $\geq 0.75$  then the system inserts some comments about the reintroduction of complications in the report. In this case, besides considering the closure, the system requires that the story includes a clear instance of the reintroduction of complications. Otherwise, it is no point to make comments about this feature.

All these parameters can be modified by the user.

**Section four.** The novelty section includes comments about the originality of the story. The system selects the appropriate text depending on the value of the story-characteristic Novel contextual structures.

## Testing the Model

To test the model we evaluated three stories: two generated by MEXICA and one generated by another story teller.

In Layer-0 we established the following conditions to continue with the evaluation process: Preconditions  $> 0.7$  and Novel Contextual Structures  $> 0.35$ .

In Layer-2 we established the following requirements for the Debaser:

```

IF Preconditions < 0.7 THEN
  Decrement-Result-Evaluation-in-2points;
IF Repetition-Sequences < Reference-RS THEN
  Decrement-Result-Evaluation-in-3points;

```

In Layer-2 we established the following requirements for the Enhancers:

```

IF (Repetition-sequences = 1) and (Result-Layer-1  $\geq 0.7$ ) and
(Closure > 0.75) THEN
BEGIN
  IF Opening = 1 THEN
    Increment-Result-Evaluation-in-0.5points;
  IF Reintroducing-Complications > 0.8 THEN
    Increment-Result-Evaluation-in-1point;
  IF Original-Value > 0.5 THEN
    Increment-Result-Evaluation-in-1.5points;
END

```

The values of the parameters are the result of several tests we have performed.

Story 1.

This story was developed by MEXICA-impro and reported in (Pérez y Pérez et al. 2010).

Jaguar knight is introduced in the story  
 Princess is introduced in the story  
 Hunter is introduced in the story  
 Hunter tried to hug and kiss Jaguar knight  
 Jaguar knight decided to exile Hunter  
 Hunter went back to Texcoco Lake  
 Hunter wounded Jaguar knight  
 Princess cured jaguar knight  
 Enemy kidnapped Princess  
 Enemy got intensely jealous of Princess  
 Enemy attacked Princess  
 Jaguar knight looked for and found Enemy  
 Jaguar knight had an accident  
 Enemy decided to sacrifice Jaguar knight  
 Hunter found by accident Jaguar knight  
 Hunter killed Jaguar knight  
 Hunter committed suicide

The following lines show the values of the story-characteristics:

```

Preconditions: 1
Opening: 1
Closure: 0.6
Climax: 1
Novel Contextual Structures: 0.71

```

Original Value: 0.71  
 Repeated Sequences: 1  
 Reintroducing Complications: 0  
 Result-Layer-1: 0.77

Figure 3 shows the graphic of tension of story 1. Because the Closure did not reach the value of 0.75 the Evaluator decided not to employ the Enhancers.

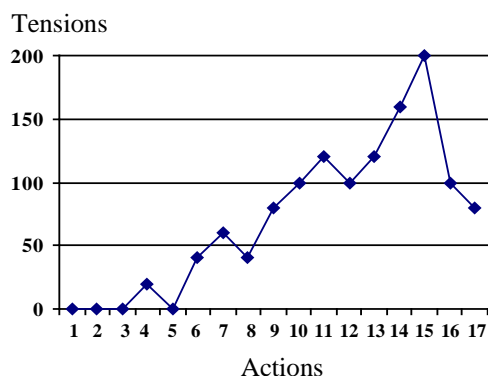


Figure 3. Tensional Representation of story 1.

The following lines produced by the agent provide the reasons of the final result:

#### EVALUATION OF THE STORY

This is a good effort. With more practice you will be able to create nice plots. Here are some comments about your work that I hope will be a useful feedback.

#### COHERENCE

The story is very logical; all actions are nicely integrated and form a coherent unit. It requires that all complications that characters faced are sorted out by the end of the last part. You need to pay more attention to this aspect.

#### INTERESTINGNESS

The text has a good introduction. The story reaches a nice climax with a good amount of tension. This is an important characteristic of a good narrative. Great! Sadly, the bad closure damages the interestingness of a story.

#### NOVELTY

The plot is kind of inventive.

My evaluation of your story is ->77/100

#### Story 2.

This story was produced by MEXICA for this paper.

Virgin disliked Jaguar knight  
 Virgin laughed at Jaguar knight  
 Jaguar knight attacked Virgin  
 Virgin fought Jaguar knight  
 Jaguar knight wounded Virgin  
 Jaguar knight ran away  
 Jaguar knight went back to Texcoco Lake  
 Jaguar knight did not cure Virgin

Tlatoani was an inhabitant of the Great Tenochtitlán  
 Tlatoani and Jaguar knight were rivals  
 Tlatoani fought Jaguar knight  
 Jaguar knight ran away  
 Jaguar knight went back to Texcoco Lake  
 Jaguar knight did not cure Virgin

The following lines show the values of the story-characteristics:

Preconditions: 1  
 Opening: 0.8  
 Closure: 0.28  
 Climax: 1  
 Novel Contextual Structures: 0.86  
 Original Value: 0.86  
 Repeated Sequences: 0  
 Reintroducing Complications: 1  
 Result-Layer-1: 0.71

Figure 4 shows the graphic of tension of story 2. The story has a really bad Closure; however, the good Climax and the relatively good result of Contextual Novel Structures push the result in Layer-1. However, Repeated Sequences are highly punished (the succession of actions 6, 7 and 8 is repeated at the end of the tale) and therefore the evaluator decrements in 3 point the final result.

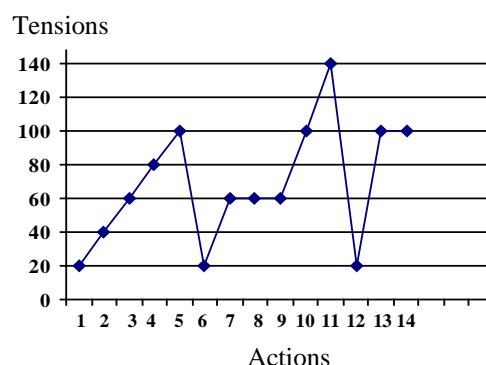


Figure 4. Tensional Representation of story 2.

The following lines show the report explaining the evaluation process.

#### EVALUATION OF THE STORY

Sorry, but this story is not good.

Here are some comments about your work that I hope will be a useful feedback.

#### COHERENCE

The story is very logical; all actions are nicely integrated and form a coherent unit. Unfortunately, there are several loose ends that need to be worked out (it reminds me of the really bad end of the TV show "Lost"). As a result the plot lacks an adequate conclusion, an important characteristic of a good narrative. You are repeating sequences of actions; as a consequence the plot is confusing!



### INTERESTINGNESS

The plot starts with some tension. The story reaches a nice climax with a good amount of tension. This is an important characteristic of a good narrative. Great! Sadly, the bad closure damages the interestingness of a story.

### NOVELTY

I find this story pretty original! I love it!

My evaluation of your story is ->41/100

Notice the last sentence in the report. Because the Original Value got a high rate the evaluator includes this sentence. It is necessary to correct this problem.

### Story 3.

This story was produced by MINSTREL (Turner 1993, p. 622). The original tale narrates the story of a knight, known as Lancelot, how was hot tempered. Andrea was a lady of the court and one day she went to the woods to pick berries. By accident, Lancelot found Andrea in the woods and he fell in love with her. Sometime later, Lancelot found again Andrea in the woods, and he saw that she was kissing another knight known as Frederik. So, Lancelot thought Andrea was in love with Frederik and got really jealous; so, he killed Frederik. Andrea told Lancelot that Frederik was her brother. Lancelot hated himself and became a hermit; Frederik was buried in the woods and Andrea became a nun. In the following lines we show the same narrative but as a MEXICA plot:

Lady and Eagle Knight were brothers  
Lady went to Chapultepec Forest  
Jaguar knight found by accident Lady  
Jaguar knight was very impressed by Lady  
Jaguar knight fell in love Lady  
Lady went to Tlatelolco Market with Eagle Knight  
Jaguar knight found by accident Lady  
Jaguar knight got intensely jealous of Eagle knight  
Jaguar knight attacked Eagle knight  
Jaguar knight killed Eagle knight  
Jaguar knight realised that Lady and Eagle Knight were brothers  
Jaguar knight hated Jaguar Knight  
Jaguar knight exiled Jaguar knight

We transformed this narrative by trying to find similar actions in MEXICA's dictionary to those described in the original tale. The following lines show the values of the story-characteristics:

Preconditions: 1  
Opening: 1  
Closure: 0.75  
Climax: 0.8  
Novel Contextual Structures: 0.54  
Original Value (surprise): 0.54  
Repeated Sequences: 1  
Reintroducing Complications: 0  
Result-Layer-1: 0.70

Figure 5 shows the graphic of tension of story 3. In this case, it is possible to employ the Enhancers and as a result the evaluation reaches the value 0.9. This happens

because the opening and the original value contribute with two points.

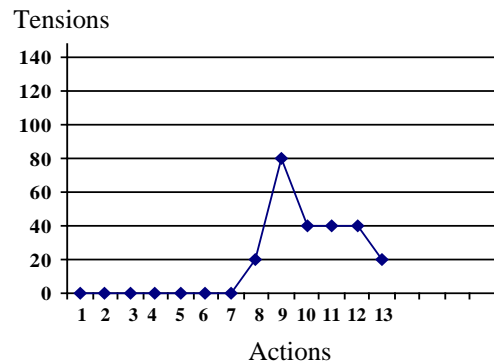


Figure 5. Tensional Representation of story 3.

The following lines show the report explaining the evaluation process.

### EVALUATION OF THE STORY

This is a good story. Great! Soon you will become a real writer. Here are some comments about your work that I hope will be a useful feedback.

### COHERENCE

The story is very logical; all actions are nicely integrated and form a coherent unit. At the end there are still some tensions that are not solved; it would help to the coherence and interest of the narrative if characters worked them out by the conclusion. I recommend you to avoid repeating actions (e.g. Jaguar knight Found by accident the Lady).

### INTERESTINGNESS

The text has a good introduction. The climax of the story is good, although for my taste I would prefer a little extra tension. A better end would contribute to have a more interesting tale. There are surprising events that make the story appealing. I enjoyed that!

### NOVELTY

The plot is kind of inventive.

My evaluation of your story is ->90/100

## Discussion and Conclusions

This paper reports a computer model for plot evaluation. The model is based on the idea that affective reactions and the generation of new knowledge are important characteristics of plot evaluation. It requires a story and a process that allows transforming a sequence of actions into structures that the agent can manage. In this way, it is possible to evaluate any story produced by any agent, as long as the narrative fulfils the constraints of the format.

I refer to the process of transforming a sequence of actions into structures that represent knowledge and affective reactions as Interpretation. This work shows the im-



portance of interpretation and its role during evaluation. If a group of agents share similar interpretations, and similar knowledge structures and beliefs (knowledge bases), they probably will produce similar evaluations. Otherwise, they will generate different outputs, maybe even contradictory ones.

The three layers provide a flexible way to work with the story-characteristics. It allows giving different weights to some features during one stage of the assessment than during others; employing what we refer to as the compensation effect; conditioning the use of the Enhancers and Debasers; and so on.

The work reported in this paper is based on an Aristotelian view of what a story is. Under this framework, the model proposes a way to understand how the evaluation process might work. However, it is well known that there are other valid approaches to build, and therefore to assess, interesting narratives. Unfortunately, it is not possible yet to develop a model that comprises all of them. Evaluation is a very complex task and we are far to understand it. So, it makes sense to develop achievable programs and then start to build on top them. Hopefully, in few years we will be able to incorporate different approaches in our system.

In the current model there are several aspects that need to be revised. For instance, it is necessary to represent features like suspense, flashbacks, and so on. Similarly, it is necessary to incorporate mechanisms that allow the system to manipulate in more creative ways the structures that are already represented; e.g. we would like to provide the evaluator with the capacity of explicitly leaving unsolved conflicts as part of an interesting closure within a narrative (when this resource is properly employed it has very positive effects on the reader). So, there is much work left to be done.

Some colleagues seem to be concerned about some characteristics of this work. Their main objection has to do with the fact that “The implementation of the used metrics is based on features certainly not present in all plot generation systems” (anonymous reviewer). There is a misunderstanding here. Our model evaluates plots; we do not necessarily care about the characteristics of the storyteller. That is, the system assesses the features present in the narrative, not in the program that generated it. So, we do not see a problem here. Nevertheless, clearly this research has been developed around our storyteller.

The main goal of this project is to provide MEXICA with the capacity of evaluating its own outputs. As explained earlier, the system can also evaluate a plot produced by any other agent as long as it is represented as text with the following format: character performing the action, description of the action, object of the action (another character). (It is also necessary that all story actions employed in the plot are declared in the dictionary of the system). That is the scope of our model.

It is necessary to consider that some plot-generators might produce outputs in the MEXICA's format that include features that cannot be interpreted by our system and therefore cannot be included as part of the assess-

ment (e.g. suspense). So, in these cases the evaluation performed by our model might be considered as incomplete.

Can this model be employed in other domains? We believe that the answer is yes. The model requires a product to be evaluated and a way to interpret such a product, i.e. a mechanism to perceive its relevant characteristics. The three layers provide a flexible method to organise and analyse such characteristics. As a result of the evaluation process the agent incorporates new structures into its knowledge base and represents affective responses. We believe that all these essential features of our model apply in other areas like, for instance, visual composition. Hopefully, this document will encourage some researchers to test the model in novel areas.

## Acknowledgements

This research was sponsored by the National Council of Science and Technology in México (CONACYT), project number: 181561.

## References

- Bremond, C. 1996. ‘La lógica de los posibles narrativos’ (trad.) In *Análisis Estructural del Relato*, pp. 99-121. México, D.F: Ediciones Coyoacán.
- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.
- Clayton, J. J. 1996. ‘Introduction: on fiction’. In *The heath introduction to fiction*, pp. 1-32. USA: D.C. Heath and company.
- Colton, S. 2008. Creativity versus the perception of creativity in computational systems. *Creative Intelligent Systems: Papers from the AAAI Spring Symposium*. 14–20.
- Deckers, L. 2005. *Motivation Biological, Psychological, and Enviromental*. Pearson.
- Jordanous, A. 2012. A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative. *Cognitive Computation*, 4(3): 246-279
- Lodge, D. 1996. *The practice of writing: essays, lectures, reviews and a diary*. London: Secker & Warbug.
- Pease, A.; Winterstein, D.; and Colton, S. 2001. Evaluating machine creativity. In Weber, R. and von Wangenheim, C. G., eds., *Case-based reasoning: Papers from the workshop programme at ICCBR 01Vancouver*. Canada 129–137.
- Peinado, F.; Francisco, V.; Hervás R. and Gervás, P. 2010. Assessing the Novelty of Computer-Generated Narratives Using Empirical Metrics. *Minds and Machines*. 20(4):565-588.
- Pereira, F. C.; Mendes, M.; Gervás, P., and Cardoso, A. 2005. Experiments with assessment of creative systems: An application of Ritchie’s criteria. In Gervás, P. Veale, T. and Pease, A., eds., *Proceedings of the workshop on computational creativity*, 19th international joint conference on artificial intelligence.
- Pérez y Pérez, R. (under review). Computer-based Model for Collaborative Narrative Generation.

- Pérez y Pérez, R. 2007. Employing Emotions to Drive Plot Generation in a Computer-Based Storyteller. *Cognitive Systems Research* 8(2): 89-109.
- Pérez y Pérez R. and Ortiz, O. 2013. A Model for Evaluating Interestingness in a Computer-Generated Plot. In *Proceedings of the Fourth International Conference on Computational Creativity*, Sydney, Australia, pp.131-138.
- Perez y Perez, R., Ortiz, O., Luna, W. A., Negrete, S., Peñaloza, E., Castellanos, V., and Ávila, R. 2011. A System for Evaluating Novelty in Computer Generated Narratives. In *Proceedings of the Second International Conference on Computational Creativity*, México City, México, pp. 63-68.
- Perez y Perez, R., Negrete, S., Peñaloza, E., Castellanos, V., Ávila, R. and Lemaitre, C. 2010. MEXICA-Impro: A Computational Model for Narrative Improvisation. In *Proceedings of the international conference on computational creativity*, Lisbon, Portugal, pp. 90-99.
- Pérez y Pérez, R. and Sharples, M. 2004. Three Computer-Based Models of Storytelling: BRUTUS, MINSTREL and MEXICA. *Knowledge Based Systems Journal*. 17(1):15-2.
- Pérez y Pérez, R. and Sharples, M. 2001 MEXICA: a computer model of a cognitive account of creative writing. *Journal of Experimental and Theoretical Artificial Intelligence* 13(2):119-139.
- Ritchie, G. 2007. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines* 17:76-99.
- Turner, S. R. 1993. *MINSTREL: A computer model of creativity and storytelling*, PhD Dissertation, University of California LA, 1993