A computer model to evaluate the coherence of characters' behaviour based on their emotional relationships

Rafael Pérez y Pérez

Departamento de Tecnologías de la Información Universidad Autónoma Metropolitana, Cuajimalpa, Ciudad de México <u>rperez@cua.uam.mx</u>

Abstract

This paper describes a model that evaluates the emotional coherence (EC model) in a narrative. A story can be defined as a sequence of actions. Each time an action is executed, new emotional relationships or conflicts between characters are established, or existing ones are modified. This interaction between characters generates what I refer to as an Emotional Trace (ET), i.e., a record of the emotional relationships and conflicts between characters as the plot progresses. The model described in this paper offers a methodology based on the analysis of the ET to determine the emotional coherence of a story. A prototype of the EC model was implemented in MEXICA and used to evaluate three scenes. Through an Internet questionnaire, a group of subjects were asked to evaluate the coherence of the same three scenes. The results of the study reported in this paper suggest that the EC model represents a solid first step to provide MEXICA, and other automatic storytellers, with the ability to evaluate the emotional coherence of its characters.

Introduction

The automatic generation of narratives is a relevant area of research in computational creativity. Through the years researchers have employed diverse techniques to generate stories, e.g., problem solving, planning, rule-based systems, genetic algorithms, and recently deep neural networks (Sharples and Pérez y Pérez 2022). These approaches have been useful to understand the role of characters' goals, authors' goals, combinatorial approaches, language models, and so on, in automatic narrative generation. But a story is not only a sequence of words and phrases, or a set of goals to be achieved by a character. It also comprises emotions, conflicts, fears, hopes, traditions, ways of interpreting the world, ways of communicating, ways of challenging our most entrenched ideas, among others. Thus, research in the automatic generation of narratives requires to explore mechanisms to include, at least, some of these social, emotional and cognitive dimensions.

Our research group has focused for several years on the study of the use of emotional relationships and conflicts

between characters as a means to develop stories. The use of emotions in plot generators is not new. For example, TALE-SPIN (Meehan 1976) and MINSTREL (Turner 1993) use variables that characterize the emotional states of characters, and those variables are used as part of the conditions necessary to activate some goals. DAYDREAMER (Mueller 1987) goes further, by employing these types of variables to control the flow of the program, that is, to activate and deactivate goals during the execution of the program (for details on how these systems work see Pérez y Pérez and Sharples, forthcoming). However, I do not know of any system that works with emotional relationships and conflict between characters as a mechanism to progress a story action by action.

For many years, one of the main challenges in this area of knowledge has been the creation of systems capable of producing coherent narratives. One popular solution is to establish mechanisms that allow the programmer to ensure that the output that the system produces is consistent. A typical example is the use of human-designed storystructures, which provide a logical and fluent way to progress a plot. In this case, the production of a story consists in instantiating such story-structures. More recent programs, such as those based on deep neural networks, use statistical relationships to generate texts. The first versions of this type of systems produced sequences of words that soon lost coherence. Their main problem was the "lack of memory" of what had happened earlier in the plot. Therefore, as the text progressed, the sentences lost connections with each other. The development of Transformers (Vaswani et al. 2017) significantly helped to reduce this problem. But it is still necessary to continue working on this limitation. In our case, our system, known as MEXICA (Pérez y Pérez and Sharples 2001; Pérez y Pérez 2015), employs emotional links and conflicts between characters to build a story context that allows progressing the plot, action by action, in a coherent way, avoiding the use of predefined story-structures. This approach offers an alternative research path to those mentioned earlier for the production of coherent texts.

The mechanisms involved in the way we humans write harmonious narratives are much more complex than those represented by our systems. The study of what I refer to as *Automatic Story Coherence* (ASCO), i.e., the study of how to represent in computational terms those methods and knowledge structures necessary for the production of texts that humans classify as coherent, is relevant for computational creativity. In this paper I present a model that aims to contribute to understand better how to develop coherent narratives that avoids using predefined storystructures.

How did the idea of building the EC model come about?

As Montford and Pérez y Pérez explains (forthcoming), the construction of our storyteller MEXICA followed four main steps:

1. Development of a cognitive account of creative writing.

2. Transformation of that cognitive account into a computer model.

3. A detailed study of how each of the elements in the computer model interact, and how they manipulate and transform information.

4. Evaluation of the outputs produced by the systems and analysis of the relation between the output's features and the elements and parameters of the model.

Step one consists of carrying out a study on different theories of the writing process. In his cognitive account of creative writing, Mike Sharples (1999) consolidates and expands the work of researchers in creativity and writing. These are the cores ideas of his cognitive account. As we write, we are constantly switching between two states known as engagement and reflection. During engagement, the writer automatically generates sequences of actions that progress the plot. The typical example is when we daydream. During reflection, the writer evaluates the material produced so far and, if necessary, he modifies it. For example, if an action generated during engagement is not fully justified in the plot, during reflection new events that give meaning to such an action are introduced. In this way, the writing process is a constant cycle between engagement and reflection.

Step two consists of figuring out how to represent in computer terms the ideas expressed by Sharples, and how to complement his cognitive account. The result of step two is the computer model of engagement and reflection (ER model) (Pérez y Pérez 1999). One of the most interesting contributions of the ER model to Sharples' account is the representation and use of emotional links and conflicts between characters as a way to progress a plot (Pérez y Pérez 2007). Steps three and four are essential to evaluate the theoretical aspects of the ER model and its functionality. The evaluation of the narratives produced, as well as the study of the relationship between the characteristics of such narratives and the parameters of the system, are critical to continue developing this project.

While working in step four I detected the need to develop the model of emotional coherence (EC model). The following illustrates how during the developing of a plot MEXICA produces situations that require to assess the emotional coherence of characters. The system starts and during engagement generates the following sequence of actions: The princess cured the knight's wounds; the knight was grateful towards the princess; the princess and the knight fell in love. Then, the program switches to Reflection. MEXICA begins the analysis of the story in progress and realizes that the plot requires justifying why the princess heals the knight. Using its knowledge base, the program discovers that if in a previous action someone hates the knight and then hurts him, then the princess's action is justified. In this way, MEXICA inserts the action 'Someone hated the knight and then hurt him' at the beginning of the narrative. The next step is to instantiate the unknown character. The program has several routines to perform that task. The most important of them is inspired by the study of human improvisation, where it is stated that whenever possible a character should be reintroduced in a story in progress (Johnstone 1989). In this way, MEXICA reintroduces the princess (who is the only option), which leads to the following sequence: the princess hated the knight and then hurt him; the princess cured the knight's wounds; the knight was grateful towards the princess; the princess and the knight fall in love.

This description feels wrong. If the princess hates the knight, it does not make sense that then she heals him and falls in love with him. That is, based on common-sense knowledge, there is a contradiction between the behaviour of the princess and the emotions she feels towards the knight. I refer to this situation as *emotionally incoherent*. Thus, it is evident the necessity of designing mechanisms that allows the system to detect this type of sequences of actions. None of the theories about creativity and writing studied to build the ER model contemplated a similar circumstance. In this article I report the solution that I designed to extend the ER model and, in this way, provide MEXICA with the ability to evaluate the coherence of the emotional behaviour of the characters.

What is an Emotional Trace?

As explained earlier, MEXICA develops a story action by action. An action has associated a set of preconditions and consequences, in terms of emotional relations and conflicts between characters. Emotional relations can have a positive or negative valence. The precondition of the action Character-A cured Character-B is that B must be ill or wounded (a conflict); the consequence is that B is very grateful towards A (an emotional link with a positive valence). Each time an action is executed, new emotional relationships or conflicts between characters are established, or existing ones are modified. This interaction between characters generates what I refer to as an *Emotional Trace* (ET). Thus, an Emotional Trace is defined as the record of the emotional relationships and conflicts between characters as the plot progresses. This work claims that the ET is important to evaluate the coherence of a narrative.

Let me elaborate this idea. Imagine a story with three characters, Carmen, Julia and Maria. As the story progress, the first interaction between characters establishes the initial state of their emotional trace. If the consequence of the first action is that Carmen is fond of Julia (an emotional link with a positive valence), one expects that in the following events that positive relation between them continues or strengths. In this way, as the story progress, it makes sense that Carmen helps Julia to solve a problem, or that they become best friends. I refer to this situation as a positive emotional trace between Carmen and Julia. Similarly, if the consequence of the first interaction between characters triggers an aversion from Carmen to Maria (an emotional link with a negative valence), one expects that in the following events that negative relation between them continues or strengths. In this way, it makes sense that Carmen and Maria end being enemies, or that they sabotage each other's goals. I refer to this situation as a negative emotional trace between Carmen and Maria.

Thus, given a partial story where two characters have developed a positive emotional trace, the coherence of a narrative is disrupted when an action with negative consequences between these two characters is added to the tale. For instance, if Carmen and Julia are best friends, it does not make sense that out of the blue Carmen betrays Julia. Similarly, a partial story where two characters have developed a negative emotional trace, the coherence of a narrative is disrupted when an action with positive consequences between these two characters is added to the tale. For instance, if Carmen and Maria are enemies, it does not make sense that in the next action Carmen asks Maria to be the godmother of her son. I refer to this type of situations as disruption of the emotional trace. Thus, in this work, a narrative is emotionally coherent when there are not disruptions in any of its ETs.

However, good stories are full of descriptions where a character betrays his brother, or falls in love with an enemy, and so on. These scenarios are not accidents; the change in the emotional trace between characters have a narrative purpose. To keep the coherence of the tale, this change in the emotional relations between characters must be justified. For instance, imagine a story where Carmen and Paul are rivals, i.e., they have a negative ET. But one day Carmen realises that Paul risked his life to safe her young nephews in danger. As a result, now Carmen is fond of Paul. In this case, the action where Paul saves the children has the

purpose of changing the emotional relation from Carmen to Paul. Now, it makes sense that the plot continues with a scene where Carmen is friendly towards Paul. I refer to this type of actions as *transitional actions*, because they help to make a coherent transition from a negative emotional trace towards a positive one, or vice versa.

The Emotional Coherence Model (EC model)

In this model, all characters who participate in an action have associated an attribute that can be set to one of two values: Proactive (P) or Reactive (R). A proactive character represents an actor that performs an action with the aim of provoking an emotional reaction or conflict in himself or in the other character. By contrast, a reactive character represents an actor that reacts to the action executed by a proactive character or, in some occasions, that reacts to the designs of fate (for example, in the event of an accident). Given an action that involves two characters, e.g., Character-A ACTION Character-B, there are four possible configurations of proactive and reactive characters:

• Character-A is proactive and character-B is reactive, represented as (ApBr).

• Character-A is reactive and Character-B is proactive, represented as (ArBp).

• Character-A is proactive and Character-B is proactive, represented as (ApBp).

• Character-A is reactive and Character-B is reactive, represented as (ArBr).

In the following, I provide details about how the EC model works. For the sake of clarity, I assume that an action only includes two characters, A and B. So, the consequences of an action can trigger emotional links and/or conflicts, either in one of the characters, or in both characters.

Case when the consequences of an action trigger emotional links and conflicts in one of the characters. For this analysis I assume that A is proactive and B is reactive (ApBr). In this case, there are two possible scenarios:

(1) When the post conditions of an action only trigger emotional links and conflicts from the proactive Character A towards the reactive Character B. That is, the consequences of the action performed by the proactive character only produces an emotional reaction in itself; character B does not react towards A. This instance is known as consequences on Character A (CA). An example is "A was jealous of B", whose consequences triggers a negative emotional link from A towards B, represented as -CA. In the same way, the consequence of the action "A admired B" is that character A establishes a positive emotional link towards B, represented as +CA.

(2) When the post conditions of an action only trigger emotional links and conflicts from the reactive Character B towards the proactive Character A. That is, the consequences of the action performed by the proactive character only produces a reaction in the reactive actor, so A does not react towards B. This instance is known as consequences on Character B (CB). An example is "A insulted B" whose consequence triggers a negative emotional link from B towards A, represented as -CB. A second example is "A cured B", whose consequence triggers a positive emotional link from B towards A, represented as +CB.

Given these two scenarios, the EC model includes three rules for a coherent emotional trace.

Rule 1. Given an Emotional Trace (ET) and an action to be used to continue the story (ACT_{t+1}), when the consequences of the action ACT_{t+1} are negative (-CA or -CB), then the ET cannot include positive emotional relations from the Proactive Character A towards the reactive Character B (see figure 1). Otherwise, the story is classified as emotionally incoherent.



Figure 1. Representation of Rule 1.

Let me elaborate this idea. To perform the action A wounded B, that triggers a negative emotional link from B towards A (-CB), makes sense if A and B are rivals, i.e., if the ET between A and B includes a negative emotional relation from A (proactive) towards B (reactive). By contrast, the same action A wounded B does not make sense if A and B are best friends, i.e., if the ET between A and B includes a positive emotional relation from A (proactive) towards B (reactive) towards B (reactive).

Rule 2. Given an Emotional Trace (ET) and an action to be used to continue the story (ACT_{t+1}), when the consequences of the action ACT_{t+1} are positive (+CA or +CB), then the ET cannot include negative emotional relations from the Proactive Character A towards the reactive Character B (see figure 2). Otherwise, the story is classified as emotionally incoherent.



Figure 2. Representation of Rule 2.

Thus, the action A buys a birthday present for B, whose consequence triggers a positive link from B towards A (+CB) makes a lot of sense if they are best friends (i.e., if

their ET includes positive emotional links between the characters); on the other hand, if character A and character B are rivals (i.e., if their ET includes negative links), this action is illogical.

Rule 3. Given an Emotional Trace (ET) that includes both, a positive and negative relations from Character A towards Character B (this situation is known as clashing emotions), the consequences of the action ACT_{t+1} can be either positive (+CA or +CB) or negative (-CA or -CB). If the consequences are positive, the negative emotional relations from the Proactive Character A towards the reactive Character B in the ET are not considered anymore (it is as if the system eliminated them). If the consequences are negative, the positive emotional relations from the Proactive Character B in the ET are not considered anymore (it is as if the system eliminated them). If the consequences are negative, the positive emotional relations from the Proactive Character A towards the reactive Character B in the ET are not considered anymore.



Figure 3. Representation of Rule 3.

So, if character A develops clashing emotions towards character B, then the behaviour of character A can be driven either by their positive emotions towards B, or by its negative emotions. But when a decision is made, it cannot be changed. During the rest of the tale character A must behave in the same way.

These three rules also work when A is the reactive character and B is the proactive character (ArBp). The difference is that the emotional relations that need to be checked in the ET are those from B (proactive) towards A (reactive).

Case when the consequences of an action trigger emotional links and/or conflicts in both characters. Some actions might produce a reaction in both characters, i.e. as a consequence of performing the action, both characters develop emotional relations and/or conflicts towards each other. This case is known as consequences on A and B (CAB). CABs can have four combinations of proactive and reactive characters: ArBr, ApBr, ArBp, ApBp, and the consequences might be positive, negative, or a mixture of them (see table 1). Let me examine each situation.

(i) The consequences of an action are always logical when both characters are reactive (ArBr). In other words, in this case there cannot be an emotional incongruity. An example is "A and B had an accident", where nobody has the intention to harm the other.

(ii) In a situation where A is proactive and B is reactive (ApBr), the action's consequences are initially evaluated as a CA and then as a CB. That is, first the coherence of all

emotions from A towards B are evaluated, considering that A is the proactive actor. Then, the coherence of all emotions from B towards A are evaluated, considering that A is the proactive actor. Here there is an example. The consequences of the action "While healing his wounds, Character A falls in love with Character B" are that Character A develops a strong love emotion towards B (+CA) while Character B reacts with gratitude towards A (+CB). Clearly, B has no intention whatsoever in this action. So, this action only makes sense if character A has positive emotional links towards B, no matter what B's emotional links towards A are.

(iii) The situation where ArBp is a mirror case of the previous one. So, the proactive and reactive characters are inverted, but the process is the same.

(iv) For the situation where both characters are Proactive (ApBp), the process works as follows. The coherence of all emotions from A towards B, and those from B towards A, are evaluated, considering that A is the proactive actor. Next, the coherence of all emotions from A towards B, and those from B towards A, are evaluated, considering that B is the proactive actor. In other words, (ApBp) can be picture as first processing the consequences as (ApBr) and then processing them as (ArBp). For instance, "A and B became best friends". In this example, both characters have the purpose of strength their relationship. The action only makes sense if both characters A and B do not have any negative emotional relations or conflicts between them. Similarly, "A and B insulted each other" only makes sense if characters A and B had negative emotional links or conflicts between them.

Proactive	Description
and reactive characters	
ArBr	The consequences of an action are always logical.
ApBr	Considering that A is the proactive actor, first the coherence of all emotions from A towards B are evaluated; then, the coherence of all emotions from B towards A are evaluated.
ArBp	This is a mirror case of the ApBr.
ApBp	The coherence of all emotions from A towards B, and those from B towards A, are evaluated, considering that A is the proactive actor. Next, the coherence of all emotions from A towards B, and those from B towards A, are evaluated, considering that B is the proactive actor.

Table 1. The four cases of proactive and reactive characters when the action has consequences on A and B (CAB).

Testing the EC model

A prototype of the EC model was developed to evaluate how these ideas work. Given an initial sequence of actions, we asked the program to verify if different endings have or not an emotional coherence. The initial sequence introduces a king, a princess and a knight (for the sake of clarity, in these examples I employ friendly texts rather that the raw inputs that the system uses):

The king was the proud father of the princess. For many years, the king and the knight had hated each other. However, the knight and the young princess fell in love...

This description establishes the following emotional links: a strong positive emotional relation between the king and his daughter; a strong hated-base negative relation between the king and the knight; and a strong positive love relation between the princess and the knight. Next, the system evaluates three scenes that share the same initial sequence but had different endings. For the sake of clarity, the endings are marked in bold.

Scene 1:

The king was the proud father of the princess. For many years, the king and the knight had hated each other. However, the knight and the young princess fell in love. The king killed the princess; then, he killed the jaguar knight.

This scene ends with the king killing the lovers. So, the king is the proactive actor and the lovers are the reactive actors. After analysing this story, the system generates the following report:

Result of the analysis of illogical actions.

The story includes the following 1 illogical action(s):

Action 7 -> King killed Princess

Explanation:

7 King killed Princess -> Earlier KING had established a positive relationship with PRINCESS.

The system reports that the king and the princess have a positive relationship, so her murder does not make sense (the system only reports uncoherent actions). To keep this story coherent, the solution is to include a transitional action, e.g., The king felt betrayed by the princess. The king can only feel betrayed if he and the princess have a strong positive emotional relation. The consequence of this action is that the king develops negative emotions towards the princess. That will justify the murder.

In the following example, the king feels betrayed by the princess (this is a transitional action) and then he kills the knight (the king is the proactive character and the knight is the reactive character). But this time, the princess (proactive character) wounds the king (reactive character), and then she cures him:

Scene 2:

The king was the proud father of the princess. For many years, the king and the knight had hated each other. However, the knight and the young princess fell in love. The king felt betrayed by his daughter. So, he killed the knight. The princess wounded the king. Later, the princess cured the king.

Result of the analysis of illogical actions.

The story includes the following 1 illogical action(s):

Action 10 -> Princess cured King Explanation:

10 Princess cured King -> The consequences of this action are not in concordance with the emotional trace from PRINCESS towards KING. Earlier PRINCESS had established a negative relation with KING.

At the beginning of the story the king and the princess had a positive relationship. The murder of her lover produces that the princess establishes a negative relation with the king. So, she has clashing emotions towards her father. Then, the princess decides to wound her father. So, the behaviour of the princess establishes that the negative relation in their ET is dominant over the positive one. Now, she needs to be congruent with that behaviour. But she is not (she cures her father). That is why the system reports that curing the king does not make sense.

In the last example, the king feels betrayed by the princess (this is a transitional action) and then he attacks the knight (the king is the proactive character and the knight is the reactive character). As a reaction the knight (proactive character) wounds the king (reactive character). Then, the princess (proactive character) kills the knight (reactive character) and heals the king (reactive character):

Scene 3:

The king was the proud father of the princess. For many years, the king and the knight had hated each other. However, the knight and the young princess fell in love. The king felt betrayed by his daughter. So, he attacked the knight. The knight wounded the king. The princess killed the knight and then she cured the king.

In this case, the princess has strong positive and negative emotional relations towards the king (because he attacked her lover) and also towards the knight (because he wounded her father). The princess reacts negatively towards the knight and positively towards her father. Both reactions follow the rule 3 of coherence, so the system does not report any problem.

What do people think about these endings?

Through an Internet questionnaire, a group of subjects were asked to evaluate the coherence of the same three scenes introduced in the previous section. The aim was to study if the criteria followed by the EC model resembles the criteria employed by a group of human judges when performing the task of analysing the consistency of a sequence of actions.

The questionnaire was in Spanish and was divided into four sections. The first part explained the objective of the study and requested the age, gender, and the last academic degree of the participant. In the three remaining sections, each of the scenes from the previous section were presented, and the participants were asked two questions to evaluate the coherence of the behaviour of the characters that participated in the closing of the scene. The possible answers to each question were a numerical value between 1 and 5, where 1 represented "very little coherent" and 5 represented "very coherent". In addition, for each question, participants were asked to explain why they granted that grade.

The questions for scene 1 were:

- Does it seem coherent to you that the king kills the princess?

- Does it seem coherent to you that the king then kills the knight?

The questions for scene 2 were:

Does it seem coherent to you that the king kills the knight?Does it seem coherent to you that the princess first wounds the king and then heals him?

The questions for scene 3 were:

- Does it seem coherent to you that the king attacks the knight, and in response to the attack the knight injures the king?

- Does it seem coherent to you that the princess then kills the knight and cures the king?

39 subjects answered the questionnaire. 53.7% of them identified themselves as males, 43.7% as females and 2.6% as nonbinary. The range of ages covered from 19 to 76 years. 51.3% had technical degree or were undergraduate students, 20.5% had a bachelor degree, and 28.2% had a postgraduate degree. In the following analysis I include descriptions made by the participants about why they decided to grant a specific coherence value. Those descriptions were translated from Spanish to English by this author with the help of an automatic translator. For reasons of space, I only include a small sample of these comments.



Figure 4. Answers to the first question of scene 1.

Results of scene 1. Figure 4 shows the results to the first question of scene 1. 66.7% of the subjects ranked the action where the king killed the princess with a value of 1 or 2, i.e., most people thought that it did not make sense. 20.5% of the subjects ranked the action with a value of 3, i.e., they were unsure about the coherence of the action. 12.9% of the subjects ranked the action with a value of 4 or 5, i.e., they classified this action as coherent.

Some participants that ranked the action as incoherent (i.e., they assigned values of 1 or 2) explained their reasons to give such values as follows:

- "Because the king loved the princess, she was his daughter"

- "Because if you love your daughter you will never hurt her"

- "If he is proud of her, it means that there is also an affective bond that would make it impossible for him to hurt her, why would he kill her then?"

- "If the father was proud of her daughter it doesn't make sense for him to suddenly kill her."

- "It seems to me that the king's love for his daughter, the princess, surpasses his hatred for the knight, and if the princess is happy with the knight, no matter how much the king hates the knight, it does not seem to me that there is any justification for killing the princess."

These comments suggests that the positive emotional relationship described in the text between the king and the princess is the main reason why the action where the king kills the princess seems incoherent. This view matches rule 1 of the EC model that establishes that if the king (the proactive character) has positive emotional relations towards the princess (the reactive character) and then he performs an action with negative consequences for her (e.g., killing her), such an action is classified as incoherent. Thus, the evaluation generated by our computer model coincides with the opinion of the majority of the subjects.



Figure 5. Answers to the second question of scene 1.

Figure 5 shows the results to the second question of scene 1. 48.7% of the subjects ranked the action where the king killed the knight with a value of 4 or 5. That is, most people thought that it did make sense. 28.2% of the subjects ranked the action with a value of 3, i.e., they were unsure about the coherence of the action. 23.1% of the subjects ranked the action with a value of 1 or 2, i.e., they classified this action as uncoherent.

Some participants that ranked the action as coherent (i.e., they assigned values of 4 or 5) explained their reasons to give such values as follows:

- "Yes, it is consistent because it was mentioned at the beginning of the story that the king and the knight hated each other, then, due to that hatred, it can be deduced that he killed him."

- "the king and the knight hated each other"

- "because he hates him"

These comments seem to be consistent with rule 1 of the EC model. In this case, the king (the proactive character) has negative emotional relations towards the knight (the reactive character) and then the king performs an action with negative consequences for the knight. As a result, the action is classified as coherent. Thus, the evaluation generated by our computer model coincides with the opinion of the majority of the subjects.

Results of scene 2. Figure 6 shows the results to the first question of scene 2. 58.9% of the subjects ranked the action where the king killed the knight with a value of 4 or 5, i.e., most people thought that it did make sense. 28.2% of the subjects ranked the action with a value of 3, i.e., they were unsure about the coherence of the action. 12.8% of the subjects ranked the action with a value of 1, i.e., they classified this action as uncoherent. None ranked the action with a value of 2.

Some participants that ranked the action as coherent (i.e., they assigned values of 4 or 5) explained their reasons to give such values as follows:

- "He hates the knight and he killed him because he felt his daughter betrayed him. It makes sense."

- "Because of his hatred towards him"

- "Because he hated it. He could have killed him out of jealousy or out of hate itself."

"For many years the king and the knight hated each other.""Because of the unbridled hatred that the King felt for the knight."



Figure 6. Answers to the first question of scene 2.

Like in the previous question, these comments seem to be consistent with rule 1 of the EC model. In this case, the king (the proactive character) has negative emotional relations towards the knight (the reactive character) and then the king performs an action with negative consequences for the knight. As a result, the action is classified as coherent. Thus, the evaluation generated by our computer model coincides with the opinion of the majority of the subjects.

Figure 7 shows the results to the second question of scene 2. 48.7% of the subjects ranked the behaviour of the princess with a value of 3, i.e., most people were uncertain about the coherence of her actions. 33.3% of the subjects ranked the situation with a value of 4 or 5, i.e., they thought that the actions made sense. 17.9% of the subjects ranked the actions with a value of 1 or 2, i.e., they classified the princess' behaviour as incoherent.



Figure 7. Answers to the second question of scene 2.

Some participants that were uncertain about the coherence of the action (i.e., they assigned a value of 3) explained their reasons to give such a value as follows:

- "The intermediate events that justify the change of posture of the princess are missing. One can imagine them, but the story does not deliver them..."

- "It could be because he was her father and she could have hurt him by chance."

- "It seems possible to me, since I imagine a scenario in which the princess reacts impulsively but then her love for her father and her possible guilt make her cure him. However, it seems unlikely to me."

- "Both things can happen without contradicting each other, if the affective bond between the king and the princess is strong, any situation can be overcome. However, that would detract from the love story between the knight and the princess."

-"Yes and no, this would depend on how the character of the princess has been presented to us throughout the tale, since that would be very important to define what she would have done, since after injuring him she could flee or remain in pain by the death of her knight but repentant for hurting his father"

Scene 2 shows a scenario where the princess develops conflictive feelings towards her father; she hates him and

she loves him. The clashing emotions appears to cause confusion to the participants. In their comments, they acknowledge that the princess has reasons to act in either way. Rule 3 of the EC model states that if the princess develops clashing emotions towards the king, the behaviour of the princess can be driven either by their positive emotions towards the king, or by its negative emotions. But when a decision is made, it cannot be changed i.e., during the rest of the tale the princess must behave in the same negative way towards the king. Situations where the princess regrets from her previous actions, as suggested by some of the subjects, are not contemplated in the current version of the model. These results suggest that the basis of rule 3 are correct but this rule might require to be updated (see the discussion section).

Results of scene 3. Figure 8 shows the results to the first question of scene 3. 83.9% of the subjects ranked the interaction between the king and the knight with a value of 4 or 5, i.e., most people thought that it did make sense. 12.8% of the subjects ranked the action with a value of 3, i.e., they were unsure about the coherence of the situation. 7.7% of the subjects ranked the action with a value of 1 or 2, i.e., they classified this action as uncoherent.



Figure 8. Answers to the first question of scene 3.

Some participants that ranked the action as coherent (i.e., they assigned values of 4 or 5) explained their reasons to give such values as follows:

- "This sequence is very logical and very coherent. If the two hated each other, it means the king's response to the knight and the consequence of his attack (he was wounded by the knight)."

- "they were enemies"

- "Yes, he could hurt him as a defence to the attack"

- "They are rivals, the fact of being attacked by your rival gives you an excuse to hit back."

- "The king was carried away by emotions and attacked, the knight defended himself."

Again, these comments seem to be in accordance with rule 1 of the EC model. Thus, the evaluation generated by our computer model coincides with the opinion of the majority of the subjects. Figure 9 shows the results to the second question of scene 3. 48.8% of the subjects ranked the princess' behaviour with a value of 1 or 2, i.e., most people thought that it did not make sense. 28.2% of the subjects ranked this situation with a value of 3, i.e., they were unsure about the coherence of the actions. 23% of the subjects ranked the action with a value of 4 or 5, i.e., they classified these actions as coherent.



Figure 9. Answers to the second question of scene 3.

Some participants that ranked the action as incoherent (i.e., they assigned values of 1 or 2) explained their reasons to give such values as follows:

- "If she is in love with the knight she would not kill him. It makes sense to me that she heals the king"

- "Because she loved the knight"

- "No, the knight defended himself from an attack; if the princess really loved him she wouldn't have killed him. At the most she would have turned him into a toad."

- "Well, the love she felt for the knight would not cause her to have killed her lover"

- "Normally the princess, being supposedly 'in love' with the knight, would have made him escape, while she would stay to heal the king"

These comments suggest that the fact that the princess reacts negatively towards the knight is not necessarily a problem; rather, the strong consequences of her reaction (killing her lover) is what makes the princess's actions being perceived as illogical. The fact that the princess is in love seems to constrain in the eyes of the evaluators what constitutes an adequate negative reaction. None of the comments argues against the princess healing the king.

In scene 3, the princess experiences two different clashing emotions. The princess was in love with the knight and, at the same time, she hated the knight because he injured her father. Simultaneously, the princess loved her father, but also hated him because he attacked her lover. Thus, the princess had strong positive and negative emotional relations towards the king and also towards the knight. Rule 3 of the EC model states that if the princess develops clashing emotions towards the king and towards the knight, the behaviour of the princess can be driven either by her positive or negative emotions towards each of those characters. But when a decision is made, it cannot be changed. Scene 3 satisfies this rule; the princess decides to act nasty towards the knight and compassionate towards the king. So, the model classifies this scene as coherent. What the model does not evaluate is whether the princess's reaction is excessive, which seems to be the main reason why subjects evaluated the princess' actions as uncoherent.

Discussion

This paper describes a model of emotional coherence for narrative generation that has been instantiated as a computer program. One of the main challenges in research on automatic narrative generation is how to produce coherent texts. There are several characteristics to consider when one evaluates the consistency of a tale. For example, the adequate structure of the narrative, the congruence between the behaviour of the characters and their goals, their personalities, their roles in the story, the social norms represented in the story, among many others. This work focuses on a model capable of evaluating the coherence of the characters' behaviour based on their emotional relationships. This type of analysis is necessary because the development of a story necessarily implies the construction of emotional relationships between its characters that change over time. Such changes must be consistent, otherwise the story loses cohesion.

The results of the study reported in this paper suggest that the EC model has a solid foundation. Most of the explanations given by the study's participants about why they evaluated characters' actions in a certain way seem to coincide with the fundamentals that drive the model. Comments made by the participants in the clashing emotions conditions suggest the need to consider the following situations originally not contemplated:

(i) In scenes like the one where the princess hurts the king and then heals him, it is necessary to relax the rigidity of rule 3 in order to allow including situations such as when the princess regrets her previous actions and therefore it makes sense that later she heals her father.

(ii) In scenes like the one where the princess decides to kill the knight and then cure her father, it is necessary to consider whether the consequences of her actions towards the knight are proportional to the emotional relationship between them.

(iii) The model should also evaluate if the emotional links between characters that are important to understand characters' behaviours are clearly shown in the story. This will help to prevent readers' confusions.

The EC model works in contexts where emotional relationships between characters, such as friendship or rivalry, are explicitly represented in the system and shape the behaviour of the characters. That is, negative emotions between characters foster actions with negative consequences, while positive emotions between characters foster actions with positive consequences. Typically, this behaviour is stable and only changes through transitional actions. For other narrative contexts, for example, stories about a serial killer where the assassin has no emotions towards his victims, or at least not the type of emotions that most of us would feel in a similar circumstance, the EC model requires to be expanded.

The work described in this paper has been incorporated into MEXICA, our storyteller, where it is used during engagement, where the story is progressed, during reflection, where the story is analysed and modified, and during the evaluation phase, where the system evaluates its own output. The system also includes a module where the user can type and evaluate any story, as long as the text follows the rigid format the MEXICA employs. The EC model also has other potential applications. For example, it can be used as part of the fitness function in systems based on genetic algorithms. It can also be used in videogames or interactive storytelling.

Narratology distinguishes between the story level (or content) and the narrative discourse (or expression). The model of emotional coherence works at the story level. However, once the system has evaluated the consistency of the story, it can use resources such as ellipsis, at the narrative discourse level, to produce more interesting narratives.

The results of this work suggests that MEXICA is useful to test, modify and expand the original concepts and theories employed as framework to build the computer model of engagement and reflection. There are many interesting aspects related to story generation which have not been represented computationally. The vast limitations of all existing systems make evident the need to explore new mechanisms. It is clear the necessity of finding new ways to produce narratives automatically, that go beyond the way we commonly use techniques such as problem solving, genetic algorithms, and recently deep neural networks. The incorporation of perspectives, methodologies and knowledge arising from the humanities and the social sciences will undoubtedly revitalize the area. Automatic narrative generation, and computational creativity in general, will greatly benefit from it.

References

Johnstone, K. 1989. *IMPRO Improvisation and the Theatre*. Methuen Drama.

Meehan, J. R. (1976). The Metanovel: Writing Stories by Computer. PhD Dissertation, Yale University.

Montfort, N. and Pérez y Pérez, R. (forthcoming). Computational Models for Understanding Narrative. Revista de Comunicação e Linguagens. https://rcl.fcsh.unl.pt/index.php/rcl

Mueller, E. T. (1987). Daydreaming and computation: A computer model of everyday creativity, learning, and emotions in the human stream of thought (Technical Report

UCLA-AI-87-8). Doctoral dissertation, Computer Science Department, University of California, Los Angeles, CA.

Pérez y Pérez, R. (1999). MEXICA: a computer model of creativity in writing. Doctoral dissertation, University of Sussex, UK.

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.

Pérez y Pérez, R. (2015). A Computer-based Model for Collaborative Narrative Generation. *Cognitive Systems Research*, 36-37, p. 30-48.

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*. Volume 13, number 2, pp. 119-139.

Pérez y Pérez, R. and Sharples, M. (forthcoming). An Introduction to Narrative Generators: How Computers Create Works of Fiction. Oxford: Oxford University Press.

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

Sharples, M. and Pérez y Pérez, R. (2022). *Story Machines. How Computers Have Become Creative Writers*. London: Routledge.

Turner S. R. (1993). *MINSTREL: A computer model of creativity and storytelling*, PhD Dissertation, University of California LA.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, December 2017, pp. 6000-6010.