Assessing the Novelty of Computer-Generated Narratives Using Empirical Metrics

**Federico Peinado · Virginia Francisco ·
Raquel Hervás · Pablo Gervás**

Abstract   Novelty is a key concept to understand creativity. Evaluating a piece of artwork or other creation in terms of novelty requires comparisons to other works and considerations about the elements that have been reused in the creative process. Human beings perform this analysis intuitively, but in order to simulate it using computers, the objects to be compared and the similarity metrics to be used should be formalized and explicitly implemented. In this paper we present a study on relevant elements for the assessment of novelty in computer-generated narratives. We focus on the domain of folk-tales, working with simple plots and basic narrative elements: events, characters, props and scenarios. Based on the empirical results of this study we propose a set of computational metrics for the automatic assessment of novelty. Although oriented to the implementation of our own story generation system, the measurement methodology we propose can be easily generalized to other creative systems.

F. Peinado (✉) · V. Francisco · R. Hervás
Departamento de Ingeniería del Software e Inteligencia Artificial, Facultad de Informática,
Universidad Complutense de Madrid, Madrid, Spain
e-mail: fpeinado@fdi.ucm.es

V. Francisco
e-mail: virginia@fdi.ucm.es

R. Hervás
e-mail: raquelhb@fdi.ucm.es

P. Gervás
Instituto de Tecnología del Conocimiento, Universidad Complutense de Madrid, Madrid, Spain
e-mail: pgervas@sip.ucm.es

Introduction

Story generation has been part of the "dream" of Artificial Intelligence since the beginning of this discipline. There are several examples, both inside and outside academia, of computational generators of narratives, showing very different degrees of success and application of scientific methodology.

As it also happens in other artistic fields, to be creative in the construction of stories involves two different challenges. The first one is making *good* stories, i.e. intelligible and interesting for the audience. The second one is making *new* stories, i.e. significantly different from others the audience already knows.

This research focuses on that second challenge, the *novelty*, a key concept to understand creativity. We consider novelty as a quality of an object *in relation to* preexisting ones. That quality presumably is inversely proportional to the similarity of the object to other ones the audience is aware of, considering not only the external resemblances but also the knowledge and the elements that have been "reused" in some way during the creative process. Human beings usually estimate this similarity informally, but in order to perform this assessment using computers, the metrics and the objects to be evaluated should be formalized and explicitly implemented.

The goal of this paper is to propose a computational approach to the assessment of novelty in computer-generated narratives. The domain is restricted to folk-tales, simplified for the purpose of our study to short narratives with simple plot structures and prototypical elements such as events, characters, props and scenarios. The approach is based on the analysis of a survey consisting on personal interviews completed by a set of voluntary evaluators and designed to obtain empirical information, both quantitative and qualitative, about the more relevant elements to consider in this type of assessment.

The results obtained indicate aspects that play a role in the determination of novelty beyond those that had been taken into account in previous works. Those aspects are related with the elements that compose a plot, the roles they play, and how they change from known plots to new ones. We present a set of arithmetic calculations over the elements of a plot that can be used as guidelines for designing the internal evaluation function of a story generator. We believe the methodology behind these formulas can be useful for the development of any creative system that requires assessment of novelty.

The paper is structured as follows. Section "Related Work" reviews work related to this research: definitions and assessment methods of creativity and novelty, as well as specific metrics for computer-generated narratives. Section "Empirical Study on Assessment of Novelty" explains the main ideas behind the design of our survey, how the study was performed and the results obtained. Section "Our Approach to the Assessment of Novelty in Stories" presents our computational approach to the assessment of the novelty of computer-generated folk-tales, formulas based on the empirical results of the previous section. Section "Conclusions" offers a summary of our conclusions about the importance of qualitative information to understand any assessment of novelty and presents some lines of future work. Finally the text of the survey employed is given in "Appendix A".

Related Work

This section reviews different definitions of creativity and novelty that can be found in the literature, and methods used for assessing novelty, i.e. computational measures and interviews with human evaluators, in well-known story generators that have been analysed in the hope of identifying the most relevant factors for the assessment of novelty.

Creativity and Novelty

Creative systems are frequently based on a model of creativity defined in two stages: *generation* and *evaluation*, referred to as "the central loop of creativity" (McGraw and Hofstadter 1993). According to this model everything the program creates is evaluated in order to filter the results that are included in the final output. Therefore, in this model the evaluation stage plays an important role but, at the same time, it is usually the hardest task.

To measure creativity the first step is to figure out what creativity means. A short definition upon which there is a consensus among the research community (Boden 1990; Ritchie 2001; Rosenman and Gero 1993) is *"the ability to produce work that is novel and appropriate"*, a clear reference to the challenges mentioned in section "Introduction".

In her seminal works, Boden (1990) argued that true creativity results from the *transformation* of conceptual space in contrast to a mere *exploration* of it. This transformation refers to a creative achievement that produces an artefact which not only is significantly different from previous ones, but also establishes new norms by which further results may be evaluated. Recently, this idea of transformation has been formally recognized as a special case of exploration, i.e. *exploration at the metalevel* according to the terminology of Wiggins (2006).

The work of Ritchie (2001, 2007) is probably the most cited reference in this topic. He looks for practical procedures to help developers of creative systems in deciding whether a particular software is or not creative using empirically observable factors. These factors are again the two basic challenges: quality, i.e. to what extent is the produced item a high quality example of its genre, and novelty, i.e. to what extent is the produced item dissimilar to existing examples of its genre, plus the addition of a third factor called *typicality*, i.e. to what extent is the produced item an example of the object class in question.

Pease et al. (2001) extend Ritchie's model emphasising the three objects susceptible to creativity measurement: the input, the output and the generation process itself. They discuss that different measures could be appropriate for different domains, generation approaches, and final applications.

Focusing on the question of novelty, Boden (1990) defines two types of creativity based on the level of novelty of the obtained result. If the result is novel to the person who has created the object, she speaks of *p-creativity*, i.e. psychological creativity. Results that have never been thought of before are due to *h-creativity*, i.e. historical creativity.

Evaluation of a creative work confronts two sets of information: the data related to the created object and the knowledge that is available to the evaluator during the evaluation. If some parts of the object were not already present in the evaluator knowledge set, then it can be said that the object is novel (Gomes et al. 1998). Macedo and Cardoso (2002) follow Boden in distinguishing between novelty and creativity, considering novelty as necessary but requiring also unpredictability, unexpectedness and surprise.

Loke (2005) concludes, after reviewing several projects, that creativity is strongly associated with novelty, and that is one reason why our efforts in this paper focus exclusively on novelty and different assessment criteria for it. Surprisingly, computational models of creativity have generally focused more on supporting the production of works that are appropriate, e.g. musical scores or poems that follow formal rules, than in supporting methods for assessing their novelty. In the majority of systems novelty is often left to chance, in the best case ruling out items (or parts of items) that are identical to those already present in the knowledge base of the system.

Pease et al. (2001) have worked on that problem, distinguishing four approaches to novelty in a creative system. Firstly, novelty relative to the body of knowledge available to the system. Secondly, novelty relative to complexity. Thirdly, novelty relative to an archetype, equivalent to the first type when the system uses archetypes as its body of knowledge. Finally, novelty as surprise, a more psychological approach depending on subjective reactions of the user.

According to Ritchie (2007), novelty is calculated by combining two separated factors: *untypicality*, i.e. how far is an item to the norm of its class of items, and *innovation*, i.e. how different is the item from those which guided its original construction in the system. Both factors have been considered in our empirical study trying to find out whether these notions are relevant for the assesment of novelty in plots or not.

Assessing the Novelty of Computer-Generated Narratives

In this section, specific methods for assessing novelty of computer-generated narratives are taken into account. We will review the methods employed in three story generation systems that follow very different approaches: Minstrel, Mexica, and ProtoPropp. These systems have been chosen as being representative of different trends on evaluation in the field of story generation. Interested readers can find more detailed discussion of these and other systems in Peinado's PhD thesis (Peinado 2008).

Novelty in Minstrel

Minstrel (Turner 1992) is a computational generator of short stories about the King Arthur and the Knights of the Round Table. It is explicitly designed as a model of human creativity, using semantic networks and case-based reasoning to create novel situations reusing a small corpus of predefined situations.

At the heart of the creative process in Minstrel are the Transform-Recall-Adapt Methods (TRAMs). These methods are a sort of "recipes" for creating the next narrative episode of a given story, imitating how other episodes stored in memory connect between them but using *similar* and not identical events, e.g. "a knight was killed after fighting with a dragon" instead of "a knight was injured after fighting with a dragon".

The evaluation of Minstrel (Turner 1992) is performed both theoretically, i.e. arguing in favor of its creative model, and empirically, i.e. using nine evaluators and three different surveys for a practical assessment. The first survey of the empirical evaluation has two parts. In the first part a story generated by Minstrel is presented to the evaluators (who do not know the origin of the text) and questions about coherence and cleverness, attention to details, use of language and overall value of the plot are asked. In the second part, three alternative endings for the story involving the suicide of the main character are presented to the evaluators that should assess effectiveness and creativeness of those "narrative solutions". The second survey has the same parts and questions as the first survey, with the only difference that the text of the story has been manually revised to improve its quality while maintaining the plot. The third survey repeats only the first part, i.e. evaluation of a given text not knowing its origin, using a text written by a real 12 year old girl. Quantitative results (ranging from 1 to 5) and qualitative information (informal comments of the evaluators) are presented, but unfortunately no deep analysis of these data is performed. Basically the author shows himself satisfied with the fact that Minstrel's creations are comparable (in terms of the surveys' results) to those of a human being.

Novelty in Mexica

Mexica (Pérez y and Sharples 2001, 2004) is a program that generates short stories about the Mexicas (the old inhabitants of México City), mostly based on the emotional relations between characters and the dramatic tension of the situations they are involved in. It is also an implementation of a model of human creativity, a cognitive account of writing called engagement-reflection (Sharples 1999).

During its *engagement* stage, Mexica generates material guided by previous content and rhetorical constraints. During its *reflection* stage the system breaks impasses (by relaxing some constraints to find actions of previous stories that can be reused to continue the story in progress), assesses the novelty and interestingness of the story in progress and verifies some coherence requirements.

To assess novelty, Mexica compares the content of the current story against stories already known to the system. The comparison is carried out by considering the stories as bags of events and measuring the similarity between any two stories as the number of events that appear in both. Mexica records the number of times that an action has been employed in previous stories, classifying actions according to their frequency of appearance. As a result of this evaluation Mexica classifies the story in progress as adequate, similar to a previous story, or a direct copy of a

previous story. Although the routine to assess the novelty of a story in progress works, Mexica does not compare parts of the story, and it does not consider the relative order of the actions that it is comparing.

The evaluation of Mexica also involved a survey. Fifty subjects from twelve different countries answered it. The subjects were asked to rate four stories created by Mexica on a 5 point scale for narrative flow and coherence, narrative structure, content, suspense, and overall quality. For comparison, two stories created by other story generation programs were included in the questionnaire, as well as a human-generated story using similar vocabulary and writing style but aiming to maximize the novelty and interest of the plot. In the questionnaires filled by the evaluators no specific questions about novelty are found. Novelty is assessed by the author, comparing the topics of several plots created with Mexica.

Novelty in ProtoPropp

ProtoPropp (Peinado and Gervás 2006; Peinado 2008) is a story generator that uses Semantic Web technologies for knowledge representation and simple combinatorial algorithms for generating the structure of new plots inspired in the Russian folktales studied by the formalist Vladimir Propp (1968).

For creating a new story, one plot of the initial corpus is modified by a random process. Thanks to the ontologies and inference engines employed, the system easily rejects stories that are not well-formed, so coherence is guaranteed. Novelty, on the other hand, is obtained only as a "subproduct" of the iterative application of changes.

In the assessment of ProtoPropp's results two methods are used. The first one is an automatic evaluation using formal metrics and the second one is an empirical evaluation using human judges and a questionnaire.

The formal metrics are based on two different formulas (Peinado 2008), one for the quality of the story and another one for the novelty. Novelty is calculated by a systematic comparison (element by element) between the structure of the generated plot and each of the plots present in the corpus used by the system. Different weights can be assigned to the comparison between different types of elements within the story (characters, locations, narrative functions...). The formula of novelty produces a real number from 0 to 1 for each element, and the average value of all these comparison is considered the representation of the p-creativity (as computed by the system) of a given story.

A questionnaire was used to ask evaluators about linguistic quality, coherence, interest and novelty of three stories generated in different ways: by ProtoPropp, randomly without constraints, and manually by the author. Results are reasonable according to the simplicity of the stories, but not specially good for novelty because the evaluators considered the random story (which is completely incoherent) more novel than the others. This particular issue, the strong but unclear relation between coherence and novelty, is a common obstacle in this type of studies.

Empirical Study on Assessment of Novelty

Our aim is to design a computational approach that measures novelty of folk-tales, having in mind to employ it for the internal evaluation of the next version of our story generation system. We started by obtaining empirical data from human volunteers. To this end, we designed a survey to be completed by a group of voluntary evaluators. Once all these judges completed the survey, their answers were analyzed and results were obtained, as shown in Section "Results and Analysis".

The Hypothesis

The study was intended to test the following hypotheses:

Hypothesis 1   Replacing an element of a plot with a different element produces a new plot which differs from the original according to the difference between the elements that have been exchanged.

Hypothesis 2   The order in which events appear in a story is significant in assessing similarity (or difference) between stories.

Hypothesis 3   Knowing the process by which a story has been generated may influence the perception of its novelty.

Hypothesis 4   Knowing the complete set of elements that are available to the generator will affect the perception of the novelty that is recognized in the result.

The Set of Plots

To test these hypotheses a set of story plots that differed from one another in controlled ways was required. Such a set was manually crafted by starting from a known folk-tale and applying a sequence of progressively more radical changes (according to our hypotheses), each set of changes resulting in a different plot.

   For the survey described in this paper, six tales were generated as follows:

- Plot F is a simplified version of the classic folk-tale "Cinderella". Changes are minimal: some events and characters have been removed (e.g. *stepsisters*), and the *shoe* has been replaced with a *dress*, for example.
- Plot E is exclusively based on Plot F changing some characters (e.g. *stepmo−ther → witch*), props (e.g.*costume → key*) and scenarios (e.g. *palace → village, . . .*).
- Plot D is exclusively based on Plot E replacing some events with very similar ones (e.g. *witch unfortunately appeared → witch appeared by a curse*).
- Plot C is exclusively based on Plot D changing the order of some of the events (e.g. *The boy wanted to know a secret of a girl* appears first in Plot C).
- Plot B is exclusively based on Plot C changing some of the events (e.g. *The boy made a promise to the spirit → An old man imposed a prohibition on the boy*).

- Plot A is a randomly generated plot to use as a baseline, not based in any of the previous plots or classic folk-tales.

The resulting six plots are presented in Table 1.

Table 1 The six plots used in the survey

Plot A	Plot B
Once upon a time... there was a baby that was very happy.The baby was kidnapped by a demon.A hunter fought the demon with a weapon.A knight fought the demon with a weapon.A prince fought the demon with a weapon.A spirit appeared by a blessing with a key in the palace.A princess rescued the baby with the key.The princess was pursued by the demon.The princess escaped from the demon.The girl gloriously returned with the baby to the village	*Once upon a time... there was a boy that was very happy.An old man imposed a prohibition on the boy.The boy violated the prohibition of the old man in the forest.The boy was murdered by a dragon.The old man publicly counteracted with a weapon.The old man was not recognized as a hero in the village.The old man fought the dragon with the weapon.The old man defeated the dragon with the weapon.The boy resurrected.The old man was transformed into a hero.The old man became a king*

Plot C	Plot D
Once upon a time... there was a boy that was very happy in a cabin.The boy wanted to know a secret of a girl.A spirit appeared by a blessing with a key in the house.A witch appeared by a curse in the cabin.The witch attempted to mock the boy.The boy resignedly accepted the help of the spirit with the key.The boy made a promise to the spirit.The boy discovered the secret of the girl in a village.The witch mocked the boy.The boy broke a promise of the spirit in the village.The girl hiddenly counteracted with the key.The girl confronted a test of courage with the key in the cabin.The girl passed the test of courage in the cabin.The girl gloriously returned with the boy to the village.The girl punished the witch for the rest of her life...	*Once upon a time... there was a boy that was very happy in a cabin.A witch appeared by a curse in the cabin.The witch attempted to mock the boy.The witch mocked the boy.The boy wanted to know a secret of a girl.A spirit appeared by a blessing with a key in the house.The boy resignedly accepted the help of the spirit with the key.The boy made a promise to the spirit.The boy discovered the secret of the girl in a village.The boy broke a promise of the spirit in the village.The girl hiddenly counteracted with the key.The girl confronted a test of courage with the key in the cabin.The girl passed the test of courage in the cabin.The girl gloriously returned with the boy to the village.The girl punished the witch for the rest of her life...*

Plot E	Plot F
Once upon a time... there was a boy subjugated to a witch in a cabin.The witch unfortunately appeared in the cabin.The witch attempted to maltreat the boy.The witch maltreated the boy.The boy desired the love of a girl.A spirit magically appeared with a key in the house.The boy gratefully accepted the help of the spirit with the key.The spirit imposed a prohibition on the boy.The boy enjoyed the love of the girl in a village.The boy violated a prohibition of the spirit in the village.The girl publicly counteracted with the key.The girl looked for the boy with the key in the cabin.The girl found the boy in the cabin.The girl returned with the boy to the village.The girl married with the boy... and they lived happy ever afterwards!	*Once upon a time... there was a girl subjugated to a stepmother in a house.The stepmother unfortunately appeared in the house.The stepmother attempted to maltreat the girl.The stepmother maltreated the girl.The girl desired the love of a prince.A fairy godmother magically appeared with a costume in the house.The girl gratefully accepted the help of the fairy godmother with the costume.The fairy godmother imposed a prohibition on the girl.The girl enjoyed the love of the prince in a palace.The girl violated a prohibition of the fairy godmother in the palace.The prince publicly counteracted with the costume.The prince looked for the girl with the costume in the house.The prince found the girl in the house.The prince returned with the girl to the palace.The prince married with the girl... and they lived happy ever afterwards!*

During the experiment, the set of plots was presented to the evaluators in a prescribed order, with the plot supposedly more distant from the original (Plot A) being presented first.

The Survey

The survey, reproduced in "Appendix A", was divided into five parts.

Parts I and V contained control questions about the evaluators and their personal data.

Part II was intended to gather information about preconceived notions about stories. It presented the list of simplified plots in the style of classic folk-tales that had to be evaluated (as described in 2). These plots had been generated manually using different approaches. Evaluators were asked to assign numerical values to the novelty of each plot, in absolute terms and in relation to the others.

Parts III and IV repeat the same questions but showing new pieces of knowledge about how the stories were created.

The idea behind the structure of this experiment was to find out how novelty values change depending on the known information about the generation process. Each group of questions had plenty of space for optional comments by the judges that were considered very valuable information for the subsequent discussion.

Survey forms were filled during personal interviews with the judges. This allowed a better control over the quality of the answers and the feedback received about the whole experiment.

Results and Analysis

The experiment was performed with a total of 13 evaluators. In the following sections we expose the results obtained with the help of the survey and our analysis and interpretation of these data.

Parts I and V: General Information

These parts of the survey ask about basic information like age and academic studies. Subjects were asked about their knowledge and experience concerning stories.

Evaluators were 70% male and 30% female, with an average age of 29. All of them have a high level of academic studies and knowledge of the English language. About 46% of the subjects stated that they remembered few folk-tales, 31% remembered between 10 and 100, and 23% remembered hundreds of them. Only 8% of the evaluators had written more than 10 stories in their life, while 46% had written a few and other 46% none.

Part II: Typical Folk-Tale, Plots and First Evaluations

The subjects were requested to provide certain background information about what a "typical folk-tale" is for them including its elements and structure. This information constitutes valuable information to guide research efforts on the

typicality of stories but it is ommited from this paper due to lack of space under the assumption that it is not directly relevant to their novelty.

Then, evaluators were shown the six plots that take part in the experiment. For each plot, evaluators should explain the topic of the story and provide an estimation of its novelty and similarity to any typical or famous folk-tale. Answers were integer values ranging from 0 ("complete disagreement" or "complete dissimilarity") to 10 ("complete agreement" or "complete similarity").

In the following questions plots were considered separately, so they are presented here in the same way. In the analysis of the overall results, the median value has been used as it is less sensitive to extreme scores than the mean.

Plot A

Evaluators assigned a median of 5 to the *novelty* of this tale. Subjects were also asked about the novel aspects they found in the tale, mentioning varied reasons.

- 40% pointed out the fact of the princess being the hero.
- 25% considered novel the presence of many characters.
- 15% emphasized the appearance of a new character in the middle of the story.

Focusing on the *similarity of the plot to a typical folk-tale*, the median obtained was 8, giving reasons about why the tale was atypical.
- 25% pointed as a lack of similarity the appearance of events that are not correctly related and have not a specific explanation. Evaluators pointed out that this resulted in a bad story.
- 15% thought that the introduction of a character in the middle of the story without any presentation is very atypical in folk-tales. Evaluators pointed out that this made the story better and more novel.

When they were asked if the presented plot was *similar to a known story*, there was almost no agreement between evaluators. 45% of them considered that it was not similar to any other story, 40% mentioned different tales but without repeating a single title (similarity median of 6), and only 15% thought that the tale was similar to the classic "Hansel and Gretel" (similarity median of only 5).

Plot B

With respect to the *novelty* of the plot, the median of the values assigned by the judges was 7.

- 30% considered the resurrection of the boy as novel.
- 25% pointed as novel the fact of having a character dying at the beginning of the story.
- 15% emphasized as something novel that the hero was an old man.

From the point of view of the *similarity with a typical folk-tale*, we obtained a median of 8. This value is quite high, so the evaluators agreed this story had the elements of a typical folk-tale, with some exceptions:
- 15% pointed to the existence of two plot lines instead of only one as something atypical. This was considered both good and novel.

- 15% considered that there might have been more explanations about the events in order to relate them, as it is usual in folk-tales. This problem was considered as bad for the story.
- 15% emphasized that the characters do not have their typical roles (the hero is an old man, the character that usually is the hero dies at the beginning, etc.). All these features made the story more novel.

When the evaluators were asked if the plot was *similar to a known story*, there was almost no agreement. 45% did not remember any similar story, other 40% mentioned different tales without repetitions (median of 5), and only 15% thought that the tale was similar to the movie "Dragon Heart" (median of 6).

Plot C

In regard to *novelty*, the obtained median is 7. The most frequent allusions to novelty were the fact of the girl being the hero (70%) and the novelty of the boy desiring to know the secret of the girl (15%).

From the point of view of the *similarity with a typical folk-tale*, we obtained a median of 7. The most mentioned reason of dissimilarity was the girl being the hero of the story (25%) which is considered novel by evaluators.

Finally, evaluators had problems for finding a *similar known story*. About 45% did not remember any similar tale, 25% mentioned some other stories (median of 6), and finally 15% thought that the story was similar to the movie "Shrek" (median of 6) and another 15% compared it again with "Hansel and Gretel" (median of 6).

Plot D

The *novelty* of this plot was considered lower than in previous tales, with a median of 6. The most frequent allusions to novelty were the fact of the girl being the hero (40%) and the girl in her role of hero not confronting the witch in the role of evil character (15%).

Taking into account the *similarity with a typical folk-tale*, evaluators considered this plot quite similar with a median of 8. The only point of the story remarked as atypical was the role of hero of the girl (40%).

When the subjects were asked whether the plot was *similar to a known folk-tale* it resulted more familiar, but there was diversity of opinions. The plot was considered similar to "Shrek" by 15% (median of 6). 15% of the evaluators pointed to "Hansel and Gretel" as similar to this plot (median of 7). Another 15% commented that the plot resembled "Cinderella" (median of 6), other 25% mentioned some other story, and 30% did not remember any similar tale.

Plot E

In regard to *novelty* in absolute terms, a median of 6 was obtained. Again, there were many subjects that stressed the hero role assigned to the girl (30%) as a sign of novelty.

Taking into account the *similarity with a typical folk-tale*, the given values had a median of 7. The most mentioned reason of dissimilarity was again the girl being the hero of the story (30%).

Finally, when the evaluators were asked if they thought the plot was *similar to any other known tale*, 40% thought that the tale was very similar to "Cinderella" (median of 6), 15% mentioned "Hansel and Gretel" (median of 7), 30% pointed to other tales, and the remaining 15% did not remember any similar stories.

Plot F

Novelty was considered significantly low for this tale, with a median of 3. Following this trend, the *similarity with a typical folk-tale* was high with a median of 8. This was caused mainly because when asked about *similar classic folk-tales*, 100% of the evaluators recognized the well-known fairy tale "Cinderella".

General Analysis

Taking into account the data obtained, there are some conclusions that can be extracted from the judgment of their novelty as expressed by our evaluators.

From the point of view of the *novelty*, it is higher in the first plots because they are not similar to any other tales. As the plots are becoming more similar to Cinderella, the novelty decreases. The elements that have been related with novelty by evaluators are the following:

- Having roles assigned to characters that usually do not play this part in tales (e.g. when the girl or the old man are the hero as happens in plot A, B, C, D and E).
- Presence of untypical assignments of characters to events (e.g. the murder of a boy as in plot B or a boy desiring to know the secret of a girl as in plot C).
- Presence of events in unusual parts of the plot (e.g. character presentation in the conflict as in plot A or having a death in the setup as in plot B).
- Unusual event (e.g. resurrection of a boy as in plot B).

Basically, it seems that novelty in a story is quite related to the novelty of the events and roles played by the characters.

In the second evaluation task that was proposed to the evaluators, they were asked to rate similarities between the plots of the previous section, taking into account that the similarity between two plots is considered independent on what plot is considered first (i.e. plot A is as similar to plot B as plot B to plot A). Evaluators were instructed to explain the criteria they were using to assign the similarity values.

The median of similarity values assigned by the evaluators for each pair of plots can be seen in Table 2.

Considering the differences between plots at the time of its creation and the ratings given by evaluators as far as similarities between plots are concerned, we can conclude the following:

- Changes of characters, props and scenarios (similarity between plot E and plot F) obtained a median similarity of 8.
- Changes of events by others very similar to the previous ones (similarity between plot D and plot E) obtained a median similarity of 8.

Table 2 Median of the similarities assigned to each pair of plots

	A	B	C	D	E	F
A						
B	3					
C	3	4				
D	3	3	9			
E	3	3	7	8		
F	2	2	5	7	8	

- Changes on the order of some of the events (similarity between plot C and plot D) obtained a median similarity of 9.
- Changes of events (similarity between plot B and plot C) obtained a median similarity of 4.
- Similarity between plot A and the rest (B, C, D, E and F) is quite low (median of 3). Evaluators mentioned that they were different stories with nothing in common.

F is the first story that generates the rest ($B \rightarrow C \rightarrow D \rightarrow E \rightarrow F$) and this is reflected in the similarity of each respect to F (source): 2 (B-F), 5 (C-F), 7 (D-F), 8 (E-F).

We can conclude that, ordering from higher to lower impact, the elements that have an influence on the similarity are:

- Complete change in the story (events and other elements).
- Complete change of the events.
- Change of the events for other similar events or changes of characters, props or scenarios.
- Change in the order of the events.

Part III: First Repetition of the Evaluation with Information About Creation of Stories

Here the evaluators were asked to re-assess the similarities between plots after being shown how each plot was created taking as base another one. In general they did not find any reason to change the similarities because of this new information. We found the following cases:

- 40% suspected the information and they had used it before. Therefore, they did not change the previous similarities.
- 30% considered the information as irrelevant and did not change the previous similarities.
- 15% changed the values previously assigned because of the new information.
- 15% changed the similarities not because of the new information, but because of taking a new look at the plots.

In the cases where the evaluators changed the similarities, the difference between values was always inside a margin of two points. So taking into account that only 30% of the evaluators changed their values and the difference was always lower than 2 we can conclude that the basic information about the tales provided in this part of the survey was not enough to influence significantly the similarity results.

Part IV: Second Repetition of the Evaluation with Information About the Elements Used to Create the Plot

Evaluators were asked again to consider the similarities between plots taking into account the limited set of elements that were used to create the plots. The idea

behind this repetition of the evaluation is to check how novelty depends on the set of elements that are used for creating new stories.

Along with these elements, the evaluators were also shown each plot with an adjacent summary of the characters, props, scenarios and narrative events (inspired in the character functions identified by Propp (1968)) that had been used in it. The median of the new values assigned by the evaluators for each pair of plots can be seen in Table 3.

These values can be compared with those calculated when the evaluators had no information about the plots (Table 2). Results are quite similar in both cases, although some of the similarities have decreased in one point when the evaluators had access to the internal structure of the plots:

- The more detailed information about events and changes with similar events (E-F) has more impact in the similarity, going from a median of 8 to a median of 7.
- With the new information, the change of order of the events (C-D) also has more impact in similarity (from 9 to 8).
- Following this trend, a complete change of events (B-C) also impacts the similarity (from 4 to 3).
- The difference of plot A with the rest of plots is also more evident with the new information (from 3 to 2).
- Similarities to Plot F (which is the first tale that generates the rest) is also more evident with the new information: 2 (B-F), 6 (C-F), 7 (D-F), 8 (E-F).

Evaluators were asked to judge the impact of every element that formed the plots:

1. 45% decided that the important elements, ordering from higher to lower impact are: events, characters, props and scenarios.
2. 15% considered that the characters are the most important elements of a plot.
3. 40% pointed out that not all the changes in events and characters have the same influence in the similarity evaluation. Changes in main events and characters have a higher impact than changes in secondary ones.
4. 40% considered that replacing a character with another one of the same type (hero, victim, evil...) has lower impact than its change by another character of different type.
5. Another 15% stated that along with events, characters, props and scenarios, they also compared the story as a whole.
6. The last 15% commented that the order of the events influences the similarity.

Table 3 Median of the similarities assigned to each pair of plots after detailed information about the plots had been presented

	A	B	C	D	E	F
A						
B	3					
C	3	3				
D	2	3	8			
E	2	4	7	7		
F	2	2	6	7	8	

We can conclude that in order to measure the novelty of a tale the most important elements are the events (1 above), followed closely by the characters (1,2). With a minor role in the global novelty would be first the props and then the scenarios (1). To calculate the novelty of events we must consider the level of the event (main or secondary) (3), the order of the events in the story (6) and the part of the story structure where they appear (5). To calculate the novelty of characters we must consider, as in the case of events, the level of the character (main or secondary) (3) and the type of the characters that have been changed (e.g. if the type of the characters are or not the same) (4).

These results are used in the next section to develop our approach to the automatic assessment of novelty for our own story generation system. Note that this preliminary experiment is more focused on getting qualitative results, to be used as a guideline for the design of our metrics, than on presenting an irrefutable set of statistically significant data.

Our Approach to the Assessment of Novelty in Stories

In order to automate the evaluation process of any story generator we propose an evaluation function that measures the novelty of a new solution against the knowledge possessed by the system. The function compares the solution story (mostly its structure) to other solutions in the system.

As discussed at the end of Section "Part II: Typical Folk-Tale, Plots and First Evaluations", in order to measure the novelty of a tale the majority of human evaluators take into account the following elements of a tale: events, characters, props and scenarios. The most important elements for human evaluators are the events that are present in a tale, so the novelty of the events will have higher weight when calculating the overall novelty of the story. Events are followed closely by characters, so the novelty of the characters will have the second highest weight. Finally, props will have a minor role in the global novelty followed closely by the scenarios which will be the elements with lower weight in calculating the total novelty. That is, the calculation of the global novelty of a story will consist of the sum of the novelty measurement of events, characters, props and scenarios, each with a different weight in the global algorithm.

The following sections explain how to calculate the novelty of each element separately and finally combine them into a single global measure.

Novelty of the Events

As stated by evaluators in section "Part IV: Second Repetition of the Evaluation with Information aboutthe Elements used to Create the Plot" it is quite different to change the main event of a story than to change a secondary one. Therefore, changes in the main events will have a higher weight in novelty assessment than changes in secondary events. The distinction between main and secondary events is not automatically determined but it is part of the input (or the knowledge representation) that is required by our metrics.

Changes that can be made on an event may be of three different types:

- Change in the order of the event in the story. For example, changing the sequence *"The boy discovered the secret. The boy broke the promise"* by *"The boy broke the promise. The boy discovered the secret"*.
- Replacement of the event with a similar one. For example, replacing the event *"The witch attempted to mock the boy"* with *"The witch attempted to maltreat the boy"*.
- Replacement of the event with a completely different one. For example, *"The girl hiddenly counteracted with the key"* with *"The old man defeated the dragon with the weapon"*.

Based on the results of the survey presented in sections "Part II: Typical Folk-Tale, Plots and First Evaluations", "Part III: First Repetition of the Evaluation with Information aboutCreation of Stories" and "Part IV: Second Repetition of the Evaluation with Information aboutthe Elements used to Create the Plot", the type of change with a higher weight in novelty measurements is the radical replacement of an event with another, followed by the replacement of an event with a similar one and finally by the change of events' orders.

So, we define the following formula for the measure of novelty:

$$EvNov = we_1 \times (we_{11} \times me_{new} + we_{12} \times me_{cho} + we_{13} \times me_{sim} + we_{14} \times me_{ch})$$
$$+ we_2 \times (we_{21} \times se_{new} + we_{22} \times se_{cho} + we_{23} \times se_{sim} + we_{24} \times se_{ch})$$

where:

- me_{new} is the number of main events that are new
- me_{cho} is the number of main events that change their order
- me_{sim} is the number of main events that are replaced with similar ones
- me_{ch} is the number of main events replaced with radically different ones
- se_{new} is the number of secondary events that are new
- se_{cho} is the number of secondary events that change their order
- se_{sim} is the number of secondary events that are replaced with similar ones
- se_{ch} is the number of secondary events replaced with radically different ones

and the following inequalities must hold for the choice of weights:

- $we_1 > we_2$
- $we_{11} < we_{12} < we_{13} < we_{14}$
- $we_{21} < we_{22} < we_{23} < we_{24}$

Appropriate values for the weights must be determined empirically as future work.[1]

If the distinction between main and secondary events is not provided as input the formula can be simplified disregarding these distinction as follows:

$$EvNovelty = we_1 \times (we_{11} \times e_{new} + we_{12} \times e_{cho} + we_{13} \times e_{sim} + we_{14} \times e_{ch})$$

---

[1] Similar considerations apply to all the weights that appear in the rest of the formulas presented in this section, though the fact will not be repeatedly mentioned to avoid redundancy.

Novelty of the Characters

Taking into account the comments made by the evaluators during the survey, not all the characters are considered equally important. It is different to change the main character of a story than to change a secondary one. Therefore, changes in the main characters will have a higher weight in novelty assessment of characters. As in the previous subsection, the information about which characters are secondary and which are central to the plot must be part of the input.

Changes that can be made on an character may be of three different types:

- Introducing a new character that did not appear in the original story. For example, the introduction of the stepsisters in plot F will be a change of this type.
- To replace a character by another one of the same type. For example, the change of the stepmother by another evil character as the witch.
- To replace a character by another one of a different type. For example, to replace a prince (which plays the role of hero in the story) with a boy (that in some tales plays the role of victim).

Based on the analysis of the comments made by the evaluators during the survey, the change with a higher weight in novelty measurement is the introduction of a new character, followed by the change of a character by another of a different type, and finally the change of a character by another of the same type.

The formula for the measure of novelty in terms of characters will be like this:

$$
\begin{aligned}
CharactersNovelty = & \ wc_1 \times (wc_{11} \times mc_{new} + wc_{12} \times mc_{chd} + wc_{13} \times mc_{chs}) \\
& + wc_2 \times (wc_{21} \times sc_{new} + wc_{22} \times sc_{chd} + wc_{23} \times sc_{chs})
\end{aligned}
$$

where

- mc_{new} is the number of main characters that are new
- mc_{chd} is the number of main characters that have been replaced by another of different type
- mc_{chs} is the number of main characters that have been replaced by another of the same type
- sc_{new} is the number of secondary characters that are new
- sc_{chd} is the number of secondary characters that have been replaced by another of different type
- sc_{chs} is the number of secondary characters that have been replaced by another of the same type

and the following inequalities must hold for the choice of weights:

- $wc_1 > wc_2$
- $wc_{11} < wc_{12} < wc_{13}$
- $wc_{21} < wc_{22} < wc_{23}$

As in the previous subsection if no distinction between main and secondary characters is provided the formula can be simplified as follows:

$$CharactersNovelty = wc_1 \times (wc_{11} \times c_{new} + wc_{12} \times c_{chd} + wc_{13} \times c_{chs})$$

Novelty of the Props

Changes that can be made on an prop may be of two different types:

- Introducing a new prop that did not appear in the original story.
- Replacing a prop with another one. For example, the replacement of the costume with the key in plot E.

Based on the analysis on the comments made by the evaluators during the survey, the type of change with a higher weight in novelty measurement is the introduction of a new prop, followed by the change of a prop by a different one.

The formula for the measure of novelty in terms of props will be like this:

$$PropsNovelty = wp_1 \times props_{new} + wp_2 \times props_{changed}$$

where:

- $props_{new}$ is the number of props that are new
- $props_{changed}$ is the number of props that have been changed

and the following inequality must hold for the assignment of weights:

- $wp_1 > wp_2$

Novelty of the Scenarios

As in the case of the props, changes that can be made on a scenario may be of two different types:

- Introducing a new scenario that did not appear in the original story.
- Changing a scenario by another one. For example, the change of the house by a cabin in plot E.

Based on the analysis on the comments made by the evaluators during the survey, the change with a higher weight in novelty measurement is the introduction of a new scenario, followed by the change of a scenario by a different one.

$$ScenariosNovelty = ws_1 \times scen_{new} + ws_2 \times scen_{changed}$$

where:

- $scen_{new}$ is the number of scenarios that are new
- $scen_{changed}$ is the number of scenarios that have been changed

and the following inequality must hold for the assignment of weights:

- $ws_1 > ws_2$

Global Measurement of Novelty

Once we have calculated the novelty of the different elements that take part in the story three more aspects must be taken into account according to the evaluators opinions discussed in section "Part III: First Repetition of the Evaluation with Information aboutCreation of Stories":

- If the story has roles assigned to characters that usually do not play this part in tales, global novelty must be increased (rol_{uch}). This requires a representation of the role that each character is playing in the story. If the input does not provide this information rol_{uch} can be omitted.
- If there are events in unusual parts of the plot the global novelty must be increased too ($event_{ups}$). This requires some kind of record of which parts of the plot are usual for each event. As in the previous case if this information is not provided $event_{ups}$ can be omitted.
- The presence of events applied to not usual characters must increase the global novelty ($event_{uch}$). This requires some kind of record of which is the likelihood of an event being applied to a character. As in the previous cases if the input does not provide this information $event_{uch}$ can be omitted.

Taking everything into account the global measure of novelty will be:

$$StoryNovelty = w_1 \times EventsNovelty + w_2 \times CharactersNovelty$$
$$+ w_3 \times PropsNovelty + w_4 \times ScenariosNovelty$$
$$+ w_5 \times rol_{uch} + w_6 \times event_{ups} + w_7 \times event_{uch}$$

with the following inequality holding for the assignment of weights:

- $w_1 > w_2 > w_3 > w_4$

Conclusions

This paper presents a computational approach for assessing the novelty of computer-generated narratives which is based on the findings of a survey performed by human evaluators on a corpus of folk-tale plots. As far as we know this is the first attempt towards the creation of novelty metrics based on empirical aspects identified in this kind of experiment with human evaluators. Although the number of evaluators is not high, the results obtained provide a clear idea of how the elements of a plot should be used when calculating its novelty. In future experiments we will focus on the most relevant aspects found in this study and we will use Internet or a similar mean for obtaining more data.

The generic method for assessing novelty involves checking for differences between the object to be measured (in this case, a story) and the objects already known by the system. This is common to both the current proposal and previous approaches. However, the results of the survey reported in this paper indicate four different aspects that play a role in the determination of novelty beyond those that

had been taken into account in previous approaches. First, the degree of novelty in stories is significantly affected by the type of elements that change from the known versions to the new. Changes in events, characters, props and scenarios of a story affect differently the value of novelty. Second, changes in the order in which events appear in a story produce significant differences in novelty. Third, changes in main elements affect the degree of novelty much more than changes in secondary elements. Finally, radical changes (exchanges between highly dissimilar elements) produce significantly higher rating of novelty than exchanges between similar elements. All these four aspects have been contemplated in the proposed metric. It must be noted that the metric we have composed is quite similar to the typical edit distance. We also take into account the number of insertions and replacements, but we do not consider deletions as a source of novelty as they have not been mentioned by any of the evaluators in the survey. In the case of event novelty, we also must consider changes of order, something that is not contemplated by edit distance.

In contrast, the various existing approaches reviewed in section "Assessing the Novelty of Computer-Generated Narratives" based the assessment of novelty only on the confrontation between new and old stories in terms of a very broad characterization of story, with no distinctions between the possible contributions to novelty of different elements, and no regard for order of presentation, which have been shown here to be relevant for the assessment of the novelty of stories.

As future work we intend to implement the formulas for novelty assessment as a fundamental part of our knowledge-based story generator. In order to provide the basic knowledge structure required to implement such a solution the story generator must employ at the very least separate representations of events, characters, props, and scenarios—in such a way as to permit the distinctions outlined above—and it must represent stories as an ordered combination of these elements. The use of an underlying ontology to represent the various elements involved in a story makes it possible to introduce the fundamental differences required to implement such a solution: clear distinctions between the different types of elements, between primary and secondary elements (as given by their role in the story), and between similar and dissimilar instances of a particular class of story element. Over such a set up the evaluation approach may be refined by empirically determining the appropriate weights for each of the formulas presented in this paper, possibly using the results of the survey presented in this paper as a seed corpora from which to extract guiding data.

Whereas the work described here has been presented from an analytical point of view, a similar approach may be applied in generative approaches to the same domain. Once we have calculated the novelty of the elements that compound the story that is being evaluated, using our ideas it is possible to propose changes that will increase its perceived novelty with a high probability, as for example: assigning roles to characters that usually do not typically play them, making characters participate in events that they do not usually participate in, or adding events in unusual order within the plot.

It remains to be seen whether the basic aspects on which this approach is based (role of particular elements in the assessed object, relative order of occurrence, relative importance of elements within the object, and degree of variation involved

in particular substitutions) may be extrapolated to domains other than storytelling. Although quality is a factor strongly dependent on the particular domain in which creativity is attempted, we believe the intuitions uncovered by the present study could help in the development of systems able to generate more novel objects, not necessarily folk-tales or stories.

Appendix A: The Survey: Subjective Evaluation of Novelty in Folk-Tales

[QUESTIONNAIRE NUMBER AND DATE]

The goal of this survey is to obtain information about how you perceive the novelty of a story. Firstly, you will read a number of simplified plots, in the style of classic folk-tales, that have been generated using different algorithmic processes. Then, you will have to value, using a numerical scale, how different they are from each other, sometimes considering additional information provided, e.g. the elements that have been used in the generation process.

A.1 Part I

Please, fill the following information about yourself.

a) *Are you male or female?*
b) *How old are you?*
c) *What is your level of academic studies?*
d) *How many folk-tales do you remember?*
e) *How many plots have you written?*
f) *What is your level of literature studies?*
g) *What is your level of English?*

A.2 Part II

First of all, spend a couple of minutes thinking about what is a "typical folk-tale". Describe your ideas in the answer box below.

a) *What kind of things appear in the "typical folk-tale" and in what order?*

[FOR EACH OF THE PLOTS, IN THE GIVEN ORDER, THE FOLLOWING BLOCK OF QUESTIONS WAS PRESENTED]

Now, read the next plot (Plot X) carefully and fill the following questions.

These questions require an answer consisting on an integer number between "0" and "10", both included. Numbers represent all the possible degrees of agreement

with what is proposed in the question, where "0" means "Complete disagreement" and "10" means "Complete agreement". When the question asks about the similarity between two plots, "0" means "Complete dissimilarity" (i.e. plots have absolutely nothing in common) and "10" means "Complete similarity" (i.e. plots are absolutely identical).

[TEXT OF THE CORRESPONDING PLOT WAS PRESENTED HERE]

a) *What is the topic of the tale?*
b) *Is this a novel (i.e. original) plot? [0–10]*
c) *How similar is this plot to the "typical folk-tale" you have described before? [0–10]*
d) *How similar is this plot to a classic folk-tale that you remember? [0–10 and the folk-tale that you remember]*

It is especially interesting for us if you explain your criteria to decide whether to write a number *lower or greater than 5*, as well as commenting how different you consider *the topics* of the folk-tales you are comparing.

Finally, fill in these questions about similarity between the precedent plots. Note that it is assumed that similarity between two plots is exactly the same no matter what plot you consider first (e.g. Plot A is as similar to Plot B as Plot B is to Plot A).

a) *How similar is Plot A to Plot B? [0–10]*
b) *How similar is Plot A to Plot C? [0–10]*
c) *...*

[EQUIVALENT QUESTIONS WERE INCLUDED FOR EACH PAIR OF PLOTS]

It is especially interesting for us if you explain your criteria to decide whether to write a number lower or greater than 5, as well as commenting what are the differences or similarities between the plots.

A.3 Part III

In this part we provide you with basic information about how the plots of Part II have been created. Then, we ask you again about the similarity between the previous plots, but this time we want you to do it from the point of view of the computational creation process, considering all the information that is revealed to you.

Basic information about the creation of the plots

Plot F is a simplified version of the classic folk-tale called "Cinderella".
Plot E is exclusively based on Plot F.
Plot D is exclusively based on Plot E.
Plot C is exclusively based on Plot D.
Plot B is exclusively based on Plot C.
Plot A has been created from scratch, i.e. not based in any other plot or classic folk-tale.

Considering this new information, fill again the questions about similarity between the plots.

[SAME QUESTIONS AS AT THE END OF PART II]

In this part it is very important, in case you change the similarity with respect to the previous part, that you comment why you made that change.

A.4 Part IV

In this part we provide you with information about the elements of the plots of Part II. Then, we ask you again about the similarity between the previous plots. You should answer *from the point of view of the computational creation process*, considering all the information that has been revealed to you.

Information about the elements of the plots *All the plots are created using the same limited set of elements. These are the narrative events, characters, props and scenarios used to create the plots* (Table 4).

Considering this new information, fill again the questions about similarity between the plots.

[SAME QUESTIONS AS AT THE END OF PART II]

It is especially interesting for us if you explain your criteria to decide the similarity and if there are aspects (characters, props, scenarios or events) more important than others for you for deciding the similarity.

Table 4 Narrative elements used to create the plots

Characters	Props	Scenarios	Narrative events
Baby	Weapon	Village	Initial situation
Boy	Costume	Forest	Prohibition and violation
Girl	Key	Cabin	Promise and break of promise
Princess		House	Investigation and discovery
Hunter		Palace	Villain apparition
Knight			Mockery
Prince			Maltreatment
Old man			Kidnapping and rescue
Fairy godmother			Murder and resurrection
Spirit			Lack and joy of love
Witch			Counteraction
Demon			Fight and victory
Stepmother			Test and success
Dragon			Look and find
			Helper apparition
			Occultation and recognition of the hero
			Pursuit and escape
			Return of the hero
			Punishment of the villain
			Wedding

A.5 Part V

Please, fill these final questions about your comprehension of the text.

a) *Have you understood the English of this questionnaire? [0–5]*
b) *Do you want to receive the results of this research project by e-mail?*

References

Boden, M. A. (1990). *The creative mind: Myths and mechanisms*. London: Weidenfeld & Nicholson.

Gomes, P., Bento, C., & Gago, P. (1998). *A function for novelty evaluation in design. Paper presented at the 4th international round-table conference on computational and cognitive models of creative design, Sydney*.

Loke, K. S. (2005). *Notes on defining novelty for computational creativity. Paper presented at the 19th international joint conference on artificial intelligence, Edinburgh*.

Macedo, L., & Cardoso, A. (2002). *Assessing creativity: The importance of unexpected novelty. Paper presented at the 2nd workshop on creative systems: Approaches to creativity in artificial intelligence and cognitive science, from the 15th European conference on artificial intelligence, Lyon*.

McGraw, G., & Hofstadter, D. R. (1993). *Letter spirit: An architecture for creativity in a microdomain. Paper presented at the 3rd congress of the Italian association for artificial intelligence, Torino*.

Pease, A., Winterstein, D., & Colton, S. (2001). *Evaluating machine creativity. Paper presented at the 1st workshop on creative systems: Approaches to creativity in artificial intelligence and cognitive science, from the 4th international conference on case-based reasoning, Washington*.

Peinado, F., & Gervás, P. (2006). Evaluation of automatic generation of basic stories. *New Generation Computing*, computational paradigms and computational intelligence, special issue: Computational creativity, *24*(3), 289–302.

Peinado, F. (2008). *Un Armazón para el Desarrollo de Aplicaciones de Narración Automática basado en Componentes Ontológicos Reutilizables*. Ph.D. Dissertation (in Spanish), Universidad Complutense de Madrid.

Pérez y Pérez, R., & Sharples, M. (2001). Mexica: A computer model of a cognitive account of creative writing. *Journal of Experimental and Theoretical Artificial Intelligence*, *13*(2), 119–139.

Pérez y Pérez, R., & Sharples, M. (2004). Three computer-based models of storytelling: Brutus, Minstrel and Mexica. *International Journal on Knowledge Based Systems*, *17*(1), 15–29.

Propp, V. (1968). *Morphology of the Folktale*. Austin: University of Texas Press.

Ritchie, G. (2001). *Assessing creativity*. Paper presented at the 3rd AISB symposium on artificial intelligence and creativity in arts and science, New York.

Ritchie, G. (2007). Some empirical criteria for attributing creativity to a computer program. *Minds & Machines*, *17*, 67–99.

Rosenman, M. A., & Gero, J. S. (1993). Creativity in design using a design prototype approach. In J. S. Gero, & M. L. Maher (Ed.), *Modeling creativity and knowledge-based creative design* (pp. 111–138). Hillsdale: Lawrence Erlbaum.

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

Turner, S. R. (1992). Minstrel: A computer model of creativity and storytelling. Ph.D. Dissertation, University of California.

Wiggins, G. A. (2006). A preliminary framework for description, analysis and comparison of creative systems. *International Journal of Knowledge Based Systems*, *19*(7), 449–458.