IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 28, 2019, accepted August 11, 2019, date of publication August 23, 2019, date of current version September 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936919

Empirical Insights Into Short Story Draft Construction

CARLOS LEÓN[1], PABLO GERVÁS[1], AND PABLO DELATORRE[2]

[1]Department of Software Engineering and Artificial Intelligence, Institute for Knowledge Technologies, Complutense University of Madrid, 28040 Madrid, Spain
[2]Department of Computer Science, University of Cadiz, 11001 Cádiz, Spain

Corresponding author: Carlos León (cleon@ucm.es)

This work was supported in part by the IDiLyCo Project funded by the Spanish Ministry of Economy, Industry and Competitiveness under Grant TIN2015-66655-R, in part by the Project FEI INVITAR-IA of the University Complutense of Madrid, and in part by the Andalusian Government under the University of Cadiz Programme for Researching and Innovation in Education 2015/2016 under Grant SOL-201500054211-TRA.

ABSTRACT Existing cognitive models of narrative creation provide accounts for story invention that, while useful, are too high level to be directly applied to formal systems like computational models of narrative generation. Inversely, existing automatic story generation systems that try to implement cognitive models can only rely on approximations to the general concepts these models describe. In order to provide insight to fill the gap between these two approaches, we have conducted a study in which human participants would invent and write short stories while reflecting on their thoughts out loud. The sessions and the analysis of the recordings was designed so that we could observe which specific modifications the participants apply to their story drafts, with the intention to inform the process of creating computational systems based on cognitive descriptions of the narrative creation process. After running the experiments, annotating the videos and analysing the output, we have concluded that there are a number of common modifications that humans tend to apply to a newly created draft, and that this information can be used to the development of storytelling systems.

INDEX TERMS Narrative construction, draft revision, empirical observation, computational storytelling, cognitive models.

I. INTRODUCTION

Narrative composition is a common creative activity. From the simple description of a short event to the complex task of writing a novel, humans must carry out certain processes that are usually deemed creative. As such, narrative composition has been studied from several perspectives like narratology, cognitive science, artificial intelligence and computational creativity. Each discipline has typically contributed from its particular perspective, focusing on the task at hand and mostly interested on those features most prominently related to the field.

In particular, each field has typically considered narrative creation from a specific abstraction level, making assumptions regarding the features on which they can safely build a model. Nevertheless, recent studies show an increasing interest in converging towards an account for narrative processing that overcomes this division and provides a standard explanation of narrative as a cognitive phenomenon [1]–[6].

However, there is still a gap between what we know about narrative as a narrotological or cognitive phenomenon and the current formalization of the fundamental narrative processes. While some of the existing formalizations actually try to adhere to cognitive models, the plethora of different approaches to computational models of narrative (see Section II-C) evidences the lack of consensus on what cognitive aspects can be robustly applied to their computational counterparts. Moreover, looking at lower level descriptions of the cognitive processes involved in narrative is not only relevant for designing implementations, but also for providing experimentally backed descriptions useful for understanding how knowledge constituents (entities and relations) are represented in the human mind.

One of the most interesting lenses under which to explore how to fill this gap between models of narrative creation and formalizable models is to study narrative generation as a construction of the creative object (in the case of this

The associate editor coordinating the review of this article and approving it for publication was Weiping Ding.

study, the narrative draft) and the writer as the creative agent. By doing so, it is possible to define narrative creation as a constant refinement of an evolving draft, which is initially empty.

Under this perspective, the object of study is not the narrative itself, but the modifications made to the draft until it becomes an accepted narrative, starting from its conception. These operations are potentially closer to a procedural model of the task of narrative creation, which can be of value if a computational cognitive model of narrative is to be built.

This paper reports on a study in which a group of human participants were asked to compose and refine a short narrative in several stages. During the process, the participants are required to voice their reflections on the process out loud. The experiments took place individually, and the process was recorded and analysed in order to identify the modifications that the subjects apply to the draft in order to refine and create a finished story. The experiments were run in a time-constrained setting, in order for the participants to write short plots. While this restricts the result of the experiment to a subset of narrative creation, it makes it possible to perform a detailed analysis at a lower level.

The experiment has been designed as an exploratory analysis task. Our main objective is to discover the most relevant modifications applied by the user, in the hope that this information can drive the design of formal and computational models of the task of narrative creation. This exploration has a strong component of qualitative analysis, but a strong effort for quantitative data has been made.

The rest of this paper is structured as follows. Section II reviews the relevant cognitive models of writing. Section III describes how this work addresses modelling narrative composition from a more low level lens that tries to follow the observed cognitive phenomena. In order to run an experimental process and obtain this low level model, the set of narrative constituents is detailed in Section III-A. The experiments are described in Section IV, and the corresponding results and analysis are reported in Section V. The relative value of the outcome can be found in Section VI, and the overall conclusions are finally summarized in Section VII.

II. REVIEW OF PREVIOUS WORK

In this section we review both the cognitive models of writing (Section II-A) and the computational approaches to writing modelled after cognitive features (Sections II-B and II-C).

A. EXISTING COGNITIVE MODELS OF WRITING

The cognitive model of writing proposed by [7] identifies three basic process: planning a set of ideas, translating these ideas into a draft of the text, and reviewing the resulting draft to progressively improve it. The three processes occur in alternating fashion, governed by a fourth monitoring process that switches control from one to the other as needed. The planning process is concerned with actively producing new material, both in terms of generating actual new ideas but also establishing goals to be addressed by other processes. The translating process addresses the task of putting ideas into words, taking into account the restrictions and resources imposed by the particular language under consideration. The reviewing process evaluates the text that has been produced by the earlier processes and revises it based on the result of the evaluation. This model is designed to capture the characteristics of communicative composition (of the kind involved when writing essays or functional texts). As such, it could be seen to be applicable to literary creativity only to a restricted extent. However, it seems reasonable to assume that a sensible computational model of literary creativity should build upon basic procedures employed for composition of texts of general nature. In terms of response to limitations imposed from external sources, Flower and Hayes' model is understood to obey constraints defined by ''the rhetorical problem'', which is understood to refer jointly to the rhetorical situation, the audience and the writer's goals.

A different theoretical account of writing is presented by [8], who understands the task as a problem-solving process which involves processes of both creative thinking and design, over the text and its content. For Sharples, the conceptual space on which a writer operates is a subset of all possible texts identified by a set of constraints which describe what kind of outputs are expected. Sharples explains that limiting the search space in this way reduces the burden placed on searches over long term memory triggered by the task. In operational terms, Sharples breaks down his description of how writers operate into a cyclic process that alternates between two different phases: engagement and reflection. The engagement phase is mainly a productive stage in which new material is generated by searching the conceptual under the given constraints. The reflection phase is a more concerned with contemplation of the results, together with an optional stage of revising the constraints in view of the results obtained to that point. Structurally, Sharples his concept of the reflection phase is described as a three-step process of reviewing, contemplating and planning the result. The reviewing step involves reading the result, possibly carrying out minor edits of the draft, but most importantly it concerns the consideration of the driving constraints, both in terms of formulating them explicitly and considering possible modifications of them. The following step of contemplation operates on these explicit formulations of the constraints, and leads to the planning step, in which plans or intentions that will drive the following phase of engagement are formulated.

It is clear from the models described so far that the role of reading what has been written to check that the inferences to be drawn are indeed the ones that the author expects plays a significant role in processes of revision. Cognitive of the narrative comprehension task also exist, and it is relevant to consider how they may influence the study being described. The work of [9] describes narrative comprehension as involving ''progressive enrichment of the mental representation of a text beyond its surface form by adding information obtained via inference, until a situation model (representation of the fragment of the world that the story is about) is constructed''.

The set of inferences to be considered as part of this process has been described in the work of [10], who describe comprehension of narrative in terms of the construction of a causal network corresponding to the narrative This causal network would be built up by the progressive identification of causal relations between the different events of a story. The network representation established in this fashion establishes the overall unity and coherence of the story.

Although it was never intended as a cognitive model, but more as a computational decomposition of the tasks that might be involved in a computational solution to the construction of narrative, the ICTIVS model [11] is another theoretical model of the writing task that includes a revision-oriented stage of interpretation of drafts in the process of building a narrative. The model originates from the analysis of the task of story construction as an instance of a basic situation in which two agents communicate. In this analysis, communication is understood as the successful exchange of a linear sequence of text, in such a way that the text encodes a complex set of data known to the sender and (a reasonably similar version of) these data become available to the receiver. This exchange requires a process whereby the sender first condenses the initial set of data into a message and then the receiver builds from the message a representation of the data as close as possible to the original. The ICTIVS model arises from the hypothesis that, in order to estimate the probability of success of his communication attempt, the sender of a narrative needs mirror images of the processes that will be applied by the receiver, so she can perform a tentative interpretation of her own message before validating it for transmission.

In order to capture separately the tasks of coming up with the content to be transmitted, encoding it into the discourse that will form the message, and consider the process of tentative interpretation implied in its validation, the model includes five stages: *INVENTION* – creating or establishing the content for the message; *COMPOSITION* – constructing the discourse that conveys the message; *INTERPRETATION* – applying mirror processes to those that will be applied by the receiver to estimate what will be understood; *VALIDATION* – predicting the impact that the message as sent, via the interpreted constructed from it, may have on the receiver; and *TRANSMISSION* – actual operation of sending the message to the receiver. The sender may cycle over the first four stages until satisfied that the interpretation and the impact predicted by his own mirror processes correspond to the reaction she expects from the receiver, and only then will she decide to transmit the message. This potential cycle though the stages is represented in Figure 1. The version depicted here corresponds to the original model in which feedback and social interaction are not addressed. Extensions of the ICTIVS model have been written for dealing with the task of reviewing the driving constraints as an integral part of the revision process [12] and the role of the purpose of author in the reviewing purpose [13].

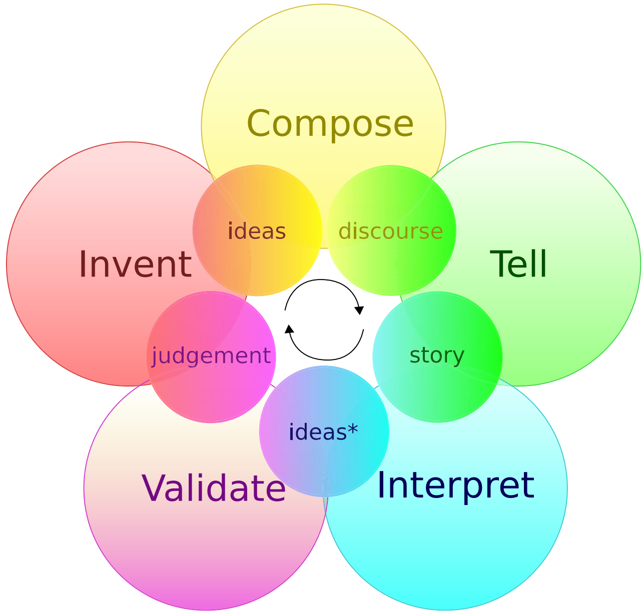

FIGURE 1. Graphical depiction of the ICTIVS model.

B. COMPUTATIONAL MODELS OF NARRATIVE CONSTRUCTION THAT RELY ON COGNITIVE MODELLING

There have been a number of efforts to develop computational models of the task of narrative construction that rely on models of specific cognitive abilities [14]–[19]. These can be classified into two different categories: those that model abilities of the author and those that model abilities of the reader. In order to ease the review, author-focused models are described first, and reader-focused models are described second.

The systems that model abilities of the author have focused very largely on planning. Planning is a well identified task involved in narrative construction as carried out by humans (as described in the models presented in Section II-A). Several systems for narrative construction based on planning technologies have been developed. Most of these systems rely on establishing an analogy between a story and a formal plan, in terms of both having an initial situation and a final goal, and requiring the identification of a path from the initial situation to the final goal. This analogy provides a powerful methodology for setting existing planning tools to the task of generating narratives, but, to the best of our knowledge, none of the authors proposing this type of approach consider it a plausible model of the cognitive operation of humans when constructing narratives.

The leading example of planning-based story generators was the Fabulist system [15]. Fabulist relied on the intent-driven partial order causal link (IPOCL) planning algorithm to build a plan that acts as a story by leading from a given initial situation to the required final outcome (defined as a planning goal). The IPOCL algorithm ensures the resulting narrative sequences are causally coherent (in the sense that character actions are motivated in some way by the context) and character behaviour is believable. However, the goals considered in the Fabulist system are goals of the characters, and the planning being applied is, in truth, analogous to the type of planning that the characters may carry out in deciding on their actions, rather than the planning carried out by an author.

This dichotomy between character planning and author planning (phrased in terms of character goals and author goals) was a central contribution of the work of [14]. Dehn set out to model the role of memory in the process of story writing, and developed the Author system as a computational implementation of that model. For Dehn, the work of an author is driven by a set of goals that constrain the narrative that is to be built. These goals take the form of requiring the story be consistent and plausible, that the behaviour of characters is believable, or that the story hold reader attention throughout its length. Some of these goals sometimes decompose into lower level goals that drive the characters into situations that the author considers will fulfill her higher level goals. The final form of the story is reached when the full complex set of author goals has been achieved. Author goals determine the structure of the story, and they drive the process of building the story, but they are not normally explicitly apparent in the surface form of the resulting story.

In spite of this distinction, subsequent efforts that apply planning technologies to model the narrative construction task remain close to the analogy between a plan and a story, rather than considering the option of modelling the planning tasks actually carried out by an author. Some of the systems resulting from this effort include further models of cognitive abilities.

Author-based computational models of narrative are usually able to produce complete short narratives, but they tend to have the limitation of omitting the impact of the produced plot on the reader. They assume an implicit evaluation process in which a human reader understands and accepts the story, but this is not carried out explicitly.

Regarding reader-focused models, Suspenser [16] was a system designed for constructing a narrative structure such that a given story world is presented to the reader in way that induces a given level of suspense. Suspenser takes three elements as input: a content to be told, a particular moment in the story plan in which the suspense induced in the reader will be measured, and the expected length of the desired story. The system then generates a discourse that conveys the input story content achieving a suspense effect at the required point. The Suspenser system operates under the following assumptions: a) the suspense experienced by the reader is related to the number of plans that the protagonist may consider solving his problems at that point of the story, and b) the way in which the reader builds her understanding of the story is determined by the structure of the discourse presented to her.

This approach of modelling suspense in terms of the number of anticipated solutions that the reader builds was further extended in the Dramatis system [17], which was a computational human behavior model of suspense. In this model, readers compute the set of possible escape plans that the protagonist may apply to avoid some negative outcome that is threatening him. The reader experiences an increase in perceived suspense whenever a possible escape stops being available or its probability of success is somehow reduced. The Dramatis system included a number of elements that can be understood as computational models of cognitive abilities involved in story understanding: a model of salience of elements in the discourse as perceived by the reader, and an algorithm for predicting which escape plan for the protagonist would be perceived by the reader as the one most likely to succeed.

Another system that relied on planning technologies to model certain cognitive abilities of a reader to guide a story generation process was Prevoyant [18]. Prevoyant was a computational model of story construction that generated stories whose structure had been optimised to maximise the surprise induced in the readers mind. The optimisation applied involved artful use of flashback and foreshadowing. To achieve this, Prevoyant relied on a reader model representing the reader's conception of a story world as constructed during reading.

Finally, [19] defined a computational model of how readers understand stories as they read them. The model was based on predicting narrative focus and inferences made. This model was used to support a narrative discourse generation system which selected content from a partial plan that represented the facts, objects, and events holding in a story world to create discourses designed to be easy to understand by a reader. In general, the reader-based models proposed in the literature tend to put the focus on the cognitive aspects of story interpretation, and usually assume or simplify a depth exploration of plot generation.

These systems, although they have rich computational models of human cognitive abilities involved in the processing of narrative, do not include computational modelling of the revision procedure as described in the cognitive models reviewed in Section II-A.

C. COMPUTATIONAL MODELS OF THE FEEDBACK LOOP IN NARRATIVE CONSTRUCTION

MEXICA [20] was a program designed to generate short stories. Because it was explicitly based on Sharples' model of the cycle of engagement and reflection [21], it is a clear example of computational system for narrative composition that includes a revision loop in its operation. The fundamental units of representation for stories in MEXICA are *story actions* (defined in terms of preconditions and post-conditions) which are combined into *previous stories*. MEXICA would read and interpret the set of previous stories to build a collection of schemas, known as *Story-World Contexts (SWC)*, that is used as basic elements for generating new stories. Story-World Contexts encode abstractions of the configurations of emotional links and tensions between the characters that held at the point of appearance of a given action in prior stories. Because of this, they can be used by the system to select when a certain action can be added to an ongoing story draft. MEXICA actively adds actions to the draft based on Story World Context during its engagement phase. The reflection phase is in charge of revising the plot to validate the ongoing draft under criteria for coherence and novelty. Validation in terms of novelty involves comparison with the set of previous stories. Identification of stories too similar to prior stories triggers the creation of guidelines that will force the system to undertake corrections during the next engagement phase.

A number of computational creativity systems have been developed that follow partially the guidelines established by the ICTIVS model of narrative composition [11]: the WASP poetry generator, the STellA story generator, and the RACONTEUR storyteller for chess games.

The WASP poetry generator [22], [23] combined basic computational models of a poet's ability to scan verse and identify stanzas with the ability to string together words into fluent text. It applied these abilities in an evolutionary approach that iterated over a population of drafts, modifying them to better match the required constraints, and filtering them based on fitness functions that measured metrical conformance. In terms of the technologies it applies, it combines n-gram modelling and evolutionary approaches. It operates by generating drafts in the form of flows of text (ICTIVS stage of INVENTION), converting these flows of text into poem drafts in given strophic forms (COMPOSITION), evaluating different aspects of these poem drafts (VALIDATION), and applying modifications to the drafts based on the results of their evaluation and aimed to correct identified shortcomings. Although the WASP system does not include a process to be aligned with the ICTIVS stage of INTERPRETATION, it does implement a cycle of progressive revision of ongoing drafts informed the results of validating the intermediate outputs. It also implements the idea of a final stage in which the drafts are deemed ready to be communicated to the target audience.

In STellA (Story Telling Algorithm) [24], [25], generation is carried out by means of combining two stages alternatively: in the first stage, a free non-deterministic simulation produces a set of alternative scenarios. The second stage (the narrative layer) selects the most promising scenarios among the candidates. Once the sequence of states is interpreted as a satisfactory narrative, the generation stops.

A subset of the ICTIVS model is implemented in STellA. The unconstrained generation that is carried out in the generative stage corresponds to the INVENTION stage in ICTIVS. Since STellA is not focused on discourse generation, COMPOSITION is addressed in a straightforward way: redundant information is filtered, and facts are laid out in increasing time order. The narrative layer performs INTERPRETATION by analysing the partial sequence at some specific state of the generation. The interpretation is carried out by comparing the current state against user-defined constraints, objectives and curves. In particular, input curves are compared against the emotional state of the partial story. The curves are used to represent narrative arches corresponding to several emotional dimensions. This interpretation is responsible for discarding or giving preference to candidate simulations. Since STellA does not address text rendering, TRANSMISSION is only applied to valid, finished stories by means of template-based text generation.

The RACONTEUR system composes discourses to communicate (a selection of) the set of facts in a chess game [26]. The composition process operates as self-evaluating cycle, in which the discourse that has been constructed at each point is decoded into a description of the facts it should communicate, and its quality measured in terms of how the interpreted version of the facts compares with the original. This self-evaluating cycle can be understood as a baseline implementation of the reviewing stage of the writing task (as understood by [7]), and also as an implementation of the sequencing of the COMPOSITION, INTERPRETATION and VALIDATION stages of the ICTIVS model.

III. MODELLING THE COGNITIVE PROCESS OF WRITING BY FOCUSING ON THE DRAFT

The background presented in Section II reveals that existing cognitive models of writing do not provide an explanation sufficiently grounded as to create formal or computational systems addressing narrative composition. This is reflected even in the case of computational systems that actually try to follow a cognitive explanation (like MEXICA, for instance), in which the implementations have to make strong assumptions in order not fully backed by the cognitive counterparts. Following the example with MEXICA, implementing the full engagement and reflection model would require a very elaborated computational definition of concentration and memory which the actual system does not provide.

This motivates the search for a more focused model of narrative composition. This model needs to address processes that are close to what is already possible to formalize and do not depend that much on high level assumptions of the functional capabilities of the human mind.

However, the literature does not provide such a model. The focus of the existing studies is put on the high level description of the psychological or cognitive processes and not so much on the engineering details required for artificial intelligence.

A low level model of narrative composition requires filling the gap between the narrative construction processes described in the literature and the current computational representation possibilities. However, given what is currently demonstrated about the mental processes involved in creativity, this is quite a challenging task. Our approach to this problem is to iteratively refine what we know about the process by focusing on what the cognitive tasks do instead of what they are. For instance, we focus on the effect on narrative planning rather than on the biological mechanisms, and particularly on how the narrative draft is actually evolved after the planning has been applied to it. This has the relative benefit of making the observation possible (the draft, albeit through indirectly, can be observed).

Additionally, it permits a comparable way of developing computational systems that mimics human behavior. It might seem that looking at the effect of the cognitive tasks does not pursue human-level computational models, but the long-term approach of iteratively refining the observation, consistently identifying the constituent subparts of each effect, is intended to narrow down the problem at each step and get near the cognitive phenomenon while advancing on its computational modelling.

The rest of the paper focuses on following this approach on narrative creation. Based on the available cognitive accounts for writing narrative, we experimentally examine what modifications human participants apply on an evolving draft. Not only the high level cognitive tasks are observed, but also what elements are added or removed from the draft. By doing this, we expect to provide additional insight to computational cognitive models through the identification of the objects to be included. This is assumed to be closer to formalization of the process.

A. IDENTIFYING THE REQUIRED COMPONENTS FOR NARRATIVE COMPOSITION

Narrative composition has commonly been divided between plot and discourse [27]. The plot is the whole universe taking place in the story: all the characters and actions that happen and form the sequence of events. The discourse is the manifestation of that sequence of events. In the discourse, not only the right set of words is important, but also the order, the references, the perspective, etc.

In this research, we focus on the plot and those elements that take part in it. This is so because most computational narrative systems focus on the narrative structure, as it can be seen in Section II-B. The experiments that have been carried out also involve discourse generation and writing, but the intention is to analyse how this affects the plot elements.

In terms of plot generation, existing computational approaches to narrative composition perform by making specific assumptions about what has to be represented (see Section II-C). While there is a diverse family of approaches, most systems build their knowledge based on symbols for representing basic cognitive concepts for characters, actions, places and moments. There are systems that address story generation by other means like machine learning [28], [29] but they do not focus on any cognitive aspect and are out of scope for this experiment.

While strong commonalities can be identified in the design of many of these computational approaches to narrative creation, their knowledge schemas do not follow a standard model. In order to provide a framework for consensus, a cognitive framework for this kind of representation [6] proposes as an architecture of how narrative is encoded as a high level set of structures. Besides, the implications of this architecture as a main component of human cognition have been discussed in [30]. In order to find a common ground between the cognitive task of narrative composition and the computational implementations this study is meant to inform, this architecture for narrative memory has been used as the reference set of main components.

Since this narrative architecture focuses on the components of narrative from a cognitive perspective, it provides a well-defined language to propose a model of the outcome of the cognitive tasks, which is exactly the desired approach for this computational cognitive model.

B. BASIC CONCEPTS RELEVANT TO THE TAGGING EFFORT

The experiments described in this paper were recorded on video and later analysed in order to identify the evolution of the draft along the creative process of composing a narrative (this is detailed in Section IV). For the analysis, the participants' performance was annotated using specialized software for tagging multimodal dialogues. While the tagging process was designed to be simple and reproducible, finding the right tag set for the videos was not straightforward because there is no available standard for annotating narrative creation.

In order to overcome this limitation, and without the intention of providing a standard for annotating narrative creation, we designed a tag set. The design objectives were 1) to enable a straightforward, easily observable annotation allowing trivial reproducibility and 2) provide coverage for both the kind of actions that are usually applied and 2) the content that is being included or modified in the text. For instance, ''invent'' would be the kind of the modification and ''a character named John'' would be the content.

The kind of modification has been previously studied in the literature both by [7] and [21]. In both cases, there is a clear identification of a creative process subject to little restriction, and a more refined step in which the current draft is revised and fixed according to an evolving set of constraints. According to Flower and Hayes, the process involves planning, translating and reviewing. Sharples applies an analysis in two levels that can be mapped into much the same elements. At the top-level, Sharples considers the process as the alternation stages of engagement and reflection, with the engagement stage involving the production of new material and the reflection stage focused on reviewing the material so far and reacting to it. At a second level of analysis, the stage of reflection is described further by Sharples as a combination of steps of reviewing, contemplating and planning. So both Flower and Hayes and Sharples consider stages of planning and reviewing, and each address the operation of putting the plans into practice in slightly different ways. Flower and Hayes describe it more broadly in terms of translating the ideas into text, Sharples considers an additional step of contemplating – which involves operating on the results of the interpretation built during reviewing, and describes the plans produced during the reflection stage as being acted upon during the engagement stage.

These cognitive processes, while plausible, are complex and tagging videos with them would not fulfill the first objective of the tag set, namely the need for a straightforward annotation. In order to overcome this limitation, we simplified the operation suggested by the literature into their observable counterparts. Finally, we defined creation, fixation and review:

- **Creation**: Any activity in which the participant comes up with new content for the story is a creation. Having evidence that the user is thinking of something new is enough, it is not important whether she actually adds it to the story.
- **Fixation**: When some original content makes it to the story in an explicit way, the operation is called a fixation.
- **Review**: Any modification in which previously existing content is modified. There must be clear evidence of the previous existence of some information and a clear intention to modify it.

These tags are not as detailed as the ones proposed by [7], but the annotations do not require much more than a trivial interpretation of the process. As it will be explained, the participants were asked to reflect on their process out loud, which made the annotation of these tags simple.

Regarding the content of the modification, we chose to design our tag set based on the narrative architecture referred in Section III-A. This was done so because this architecture proposes a categorization of narrative elements that is both intended to have a strong cognitive background and a computational orientation. As such, we consider it a good candidate for a tag-based transcription of the experiment, given that the overall objective of the study is to provide insights to fill the gap between computational and high-level cognitive models of narrative composition.

According to the aforementioned narrative architecture by [6], these are the basic knowledge constituents of narrative content:

- **Kernels and satellites**: The narrative-cognitive definition of kernels and satellites in the narrative architecture are based on the work by [31]. Kernels are the main part of a narrative object, and satellites are narrative constituents that are connected to the kernel and complete it by adding causation, contextualization or explanation. In general, these are assumed to be the narrative's basic actions.
- **Time**: Time, absolute or relative, represents the narrative abstraction of the physical time. According to the narrative hypothesis, narrative is a fundamental way to represent knowledge, therefore requiring a representation of basic physical magnitudes.
- **Location**: Location defines the physical context in which the action happens, and sets a background in which the perception and interaction ranges of the characters are well bounded, hence its importance.
- **Causality**: Causality provides coherence to a narrative and links the facts taking place in it. Causality in narrative serves two roles: provides a justification of each occurring event and creates a network in which all included events lead to a final outcome, certainly setting the limits of a narrative [10].
- **Agency and characters**: Characters, as the agents of the occurring actions, are a fundamental narrative object.
- **Composition**: The kernels in a story are connected together in a variety of ways. By letting kernels be satellites of other kernels, it is possible to create more complex narrative structures.
- **Abstraction**: Narratives can be managed with different degrees of abstraction. The narrative architecture defines abstractions as new levels or narrative (more or less abstracted), and considers abstractions to be narratives on their own.

Since this study only focuses on the modifications applied to the draft, the structures focusing on the overall structure of the narrative are not of interest because they are not concrete modifications but higher level structures: causality, composition and abstraction. While we acknowledge their fundamental role in narrative, the analysis of these features are beyond the objective of this study. Therefore, the list of modification tags for the annotation process are set to those making reference to time, location, agency and characters, and kernels as the main unit constituting the narrative.

The final set of tags is listed next. These tags will be used in the video annotations to make reference to draft modifications in which the most important aspect is a character, action, place, moment or other:

- **Characters**: Modifications tagged as character are those operations in which the most important aspect is a definition or an evolution of a character, its agency, or its physical or mental state.
- **Actions**: Modifications tagged as action are those in which the action (and therefore its consequences) are the relevant part of the operation.
- **Places**: When the modification refers to a physical location, relative or absolute, at any level.
- **Moments**: Modifications referring time. Again, this can be relative or absolute.
- **Other**: Any other modification that does not fall into any of the previous categories. This tag was used to annotate specific modifications that, while clear, were not focused on any of the previous categories. For instance, references to the environment or the weather.

IV. EXPERIMENT

As advanced in Section I, the experiment was designed as an exploratory analysis. Given the current problem to be studied, the closest we can get to a hypothesis is to base the experimental design on the previous cognitive models. The experimental task is then approached as a first step towards the identification of main modifications that writers apply to a draft.

A. ETHICS STATEMENT

The present study was carried out in accordance with the recommendations of national and international ethics guidelines, Código Deontológico del Psicólogo and American

FIGURE 2. A photograph of the setup for the experimentation. The experimenter is located by the window and the participant sits in from of him.

Psychological Association. The study does not present any invasive procedure, and it does not carry any risk to the participants' mental or physical health, thus not requiring ethics approval according to the Spanish law BOE 14/2007. All subjects participated voluntarily and gave written informed consent in accordance with the Declaration of Helsinki. They were free to leave the experiment at any time.

B. EXPERIMENT SET UP

The experiment was design as a four-stage version of the analysis that Flower and Hayes carried out [7]. In their analysis, they assign a writing task to the participant, who writes a narrative in one hour. The main characteristic of the experiment is that the participant is asked to verbalize everything that goes through her mind.

The experiment took place from 12th of February 2018 to the 28th of February 2018, in the Computer Science Faculty of Universidad Complutense de Madrid. A dedicated office was prepared. In this office, two desktop tables were put together side by side, one for the experimenter and the other for the participant. A single desktop PC (HP ProDesk 490 G3 MT Business PC) was connected to two sets of screen, keyboard and mouse. One set was used by the controller of the experiment and the other one was available to the participant. The participant's screen was always off except when she was required to use the computer.

Additionally, the participant's desk was provided with paper (a pile with A4 paper sheets), ball pens (green, red, blue and black), color post-it notes, small post-it page annotators, a pencil, a pencil sharpener, paper clips and glue. All this material was made available for the participant to have the freedom to annotate her thoughts with all flexibility. The use of the material the participants made use of is summarized in Section V. Figure 2 shows a picture of the set up. The experiments were run independently, participant by participant. The sessions were schedule to take one hour.

C. PARTICIPANTS

The experiment was publicly announced in the Computer Science Faculty of Universidad Complutense de Madrid with a public link to a Google Form querying for their name an email. The link was also disseminated in the student associations, which made it possible to reach non-students too since most of them also include alumni and people not enrolled in the university.

It is important to note that the objective of the experiment was to obtain information about how simple plots are constructed by subjects that are not professional writers. That is, the focus was put on the underlying cognitive aspects of basic narratives, and not on the literary or aesthetic aspects of writing. Because of this, no specific background on writing was asked from the participants.

All interested participants filling the form automatically received an automatic link to a Google Calendar with Appointment Slots in which participants could automatically reserve a one-hour slot among the available ones.

After the experiment, a questionnaire gathering both the demographics and some experiment-related information was given to the participants to fill in. In total, 14 participants took part in the experiment, 10 males (71.42%) and 4 females (28.57%). There was no remuneration for participating in the experiment. They all were informed about the process and the data acquisition and analysis. All participants accepted to be recorded and all the conditions. Their age ranged from 19 to 41 ($avg = 23.37$, $sd = 6.29$). They were asked to write their current activity. 11 of them (78.57%) were undergrad students, 2 (14.28%) were researchers (not related to this research) and 1 (7.14%) was a programmer.

Some the participants could potentially know each other and that the experiment required them to create an original story from scratch. Therefore, in the questionnaire there were also asked if they were willing not to disclose any information on the experiment until it had finished, and all of them agreed.

D. EXPERIMENT SCRIPT

For each session, the participant was asked to enter the room, close the door, leave all her things in a table away from the desk, and turn her cell phone off. She was welcomed and thanked, and she was informed that the cameras were turned off. After that, she was informed and asked for consent on the recordings and the data to be stored analysed after the session.

The participant was told that the objective of the experiment was to observe the quality of the stories created by humans and compare them among the participants according to their profile. This was done in order to have the participants focus on creating a potentially valuable story and to promote creative behaviour. Besides, that would keep their focus off the process itself, which was the aspect that was going to be observed. After this short explanation, the participant was asked to report all her thoughts out loud. A short video of the experimenter creating a story and verbalizing the process was shown to them as an example. She was given 3 minutes to practise and compose a short story with one character. The participants had the stopwatch always visible so that they could adapt the plot size and complexity to the available time. Then, the actual experiment, divided in 4 stages, followed the next steps:

1) **Stage 1** (G_1): The actual experiment started. First, the participant was asked to compose a short story in 10 minutes (again, the stopwatch was visible). The story was not meant to be written, only composed. Before the time was over, the participant had to tell the story verbally. There were no restrictions beyond the number of characters. The participants were asked to create a story with exactly 3 characters, and no extra restrictions were given to them.
2) **Stage 2** (D_1): Once the time was over, the participant's screen was turn on. The experimenter had prepared her desktop so that a blank document in Microsoft Word 2016 was the only running application. The participant was then given 10 extra minutes to type the story in the text processor. The participant was allowed to apply any modification she would deem necessary to the story, but the overall story had to be the same. She was informed that the format and the final text layout was not relevant for the experiment, so the participant only had to focus on the words, sentences and paragraphs. Again, the participant could see the remaining time.
3) **Stage 3** (G_2): The participant's screen was turned off and the keyboard and mouse were moved away to make more free space in the desk. Then, she was asked to make a modification of the original plot by adding one extra character to the story (a total of 4 characters). This character had to have "some important role in the story", but apart from this restriction (meant for the participant not to add one extra irrelevant character) the participant was told she could be freely creative. Another slot of 10 minutes was assigned for this. The participant did not have to use the computer, just come up with the new story.
4) **Stage 4** (D_2): The screen was turned on, a new blank document in Microsoft Word 2016 was presented to her, and she was again given 10 minutes to type the story composed in stage 3.

After stage 4 was over, the participant was informed that the tasks were over. The experimenter took control of the experiment computer and presented a Google Form questionnaire with demographic, consent and experiment-related questions. Then the recording was stopped, the participant was thanked again, and the session was finished.

The script was followed in all experiments with no major issue. All participants agreed to be recorded in the session.

E. FINAL QUESTIONNAIRE

As introduced in Section IV-D, the last part of the experiment consisted of a questionnaire given to the participants. This questionnaire was divided in a short description of the questionnaire itself plus 7 sections. Given the simplicity of the format of the questions and that the users were already using a computer, Google Forms was used as the platform.

The first section of the questionnaire was focused on the participant's own perception of the creative process. The participants were asked:

1) Whether she liked her story with 3 characters (4 point Likert scale).
2) The amount of perceived modification when writing the first text (4 point Likert scale).
3) Whether she liked her story with 4 characters (4 point Likert scale).
4) The amount of perceived modification when writing the second text (4 point Likert scale).
5) Which story required more effort.
6) Which one, according to her perception, she had modified more.
7) How did reporting all mental processes out loud had affected the process.

The second section was meant to gather information about the knowledge of narrative with questions about:

1) The number of short narratives read per year.
2) The number of long narratives read per year.
3) Their knowledge of narrative in general (5 point Likert scale).
4) Whether the participant used to write narrative.
5) Their skills for writing narrative (5 point Likert scale).
6) Whether the participant considered herself a creative person.

The third section was devoted to get information about the participant's skills regarding the used tools:

1) Use of personal computers (5 point Likert scale).
2) Typing (4 point Likert scale).
3) Text processors (5 point Likert scale).

The fourth section gathered free comments and these two questions:

1) Whether the participant felt she had freedom when creating the story.
2) Whether the participant felt comfortable during the experiment.

The fifth section asked about the permission for using the recorded material, the sixth about demographic information (age, gender, occupation) and the seven asked the user not to disclose any detail about the experiment.

F. ANALYSIS OF THE RECORDINGS

The creative part of the experiment consisted of 4 stages in which the participants created an original story by alternating conception and writing. The videos were annotated according to the tags described in Section III-A. Two tracks were used for the tagging: one for the kind of modification in which the current activity could be classified as creation, fixation or revision and another one for marking the content that was the object of the modification: character, action, time, place or other.

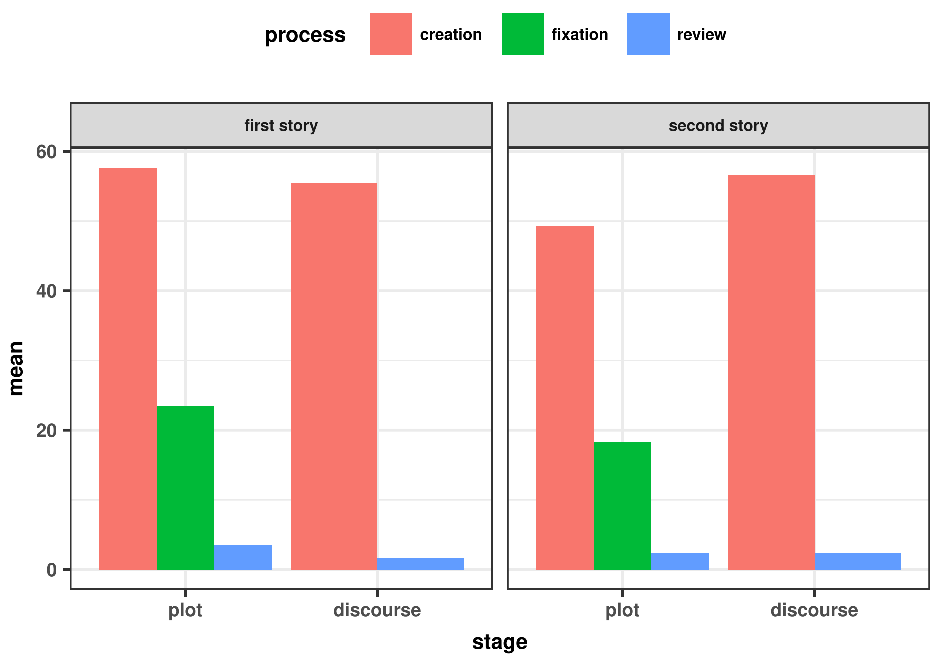

FIGURE 3. Bar chart representing the average modifications per stage (average values and standard deviations).

The videos were annotated with ANVIL 6.0 [32]. As expected, annotating the videos with this tag set was straightforward, since the kind of modifications and the content that was modified were obvious tasks. However, we discovered one limitation in the tag set.

When adding the tags for the revisions during plot invention (stages G_1 and G_2), it was not always obvious when the participants were applying modifications to the story, since the fact that they omit content when speaking out loud did not mean that the content was actually taken out from the story. Therefore, the deletions, as part of the modifications, were only extracted from the written discourse (stages D_1 and D_2). This means that the gathered data for the revisions is partially incomplete in G_1 and G_2.

Likewise, in stages in which the text was written (D_1 and D_2), fixations would constantly occur. Since transcribing existing text is actually a fixation, we decide not to include this in the analysis to avoid redundancy.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In total, 3, 704 tags were created during the analysis of the videos. This amount corresponds to 264.57 ($sd = 56.08$) tags per participant. Additionally, specific information was annotated for each tag, for a total of 39, 129 data.

The average values and standard deviations for the annotations for the kind of modification (as defined in Section III-B) are shown in Table 1. In this table, creation means all modifications that alter the draft by adding new content, fixation refers to the process by which a conception becomes set and makes it to the plot and revision makes reference to any modification to the draft that changes previously fixed content.

TABLE 1. Modification process per stage (average values and standard deviations).

process	G_1	D_1	G_2	D_2
creation	57.64 (11.96)	55.43 (13.31)	49.28 (18.54)	56.64 (9.67)
fixation	23.50 (9.02)		18.36 (7.89)	
review	3.50 (1.36)	1.67 (1.15)	2.33 (1.94)	2.33 (1.15)

Figure 3 contains a bar chart representation of the data in Table 1. The figure and table show a strong prevalence of creations over fixations and revisions. In the plot creation stages, the number of creations is slightly less than twice as much as the number of fixations.

There is a relatively low number of revisions. We believe that this is due to the constraints given to the participants. Since they were committed to create a story in intervals of 10 minutes, they seemed to opt for a straightforward generation strategy in order to have a finished story on time. This is discussed in Section VI.

This assumption is backed up by the qualitative analysis of the videos. During this process, we discovered that participants overestimated the time in the first exercises (include the test exercise in which they were given three minutes). As such, they iteratively refined the process in order to make the generation fit the available time. This aspect has not been included in the statistics because there is no clear way to tag this behavior. However, all involved researchers agree, according to the observations, that the participants focused on finishing a full story instead on having an incomplete, more polished one.

Creations happen both in the creation of the plot and the discourse. It is important to note at this point that the creations in the discourse are creations related to the plot, not the linguistic aspects.

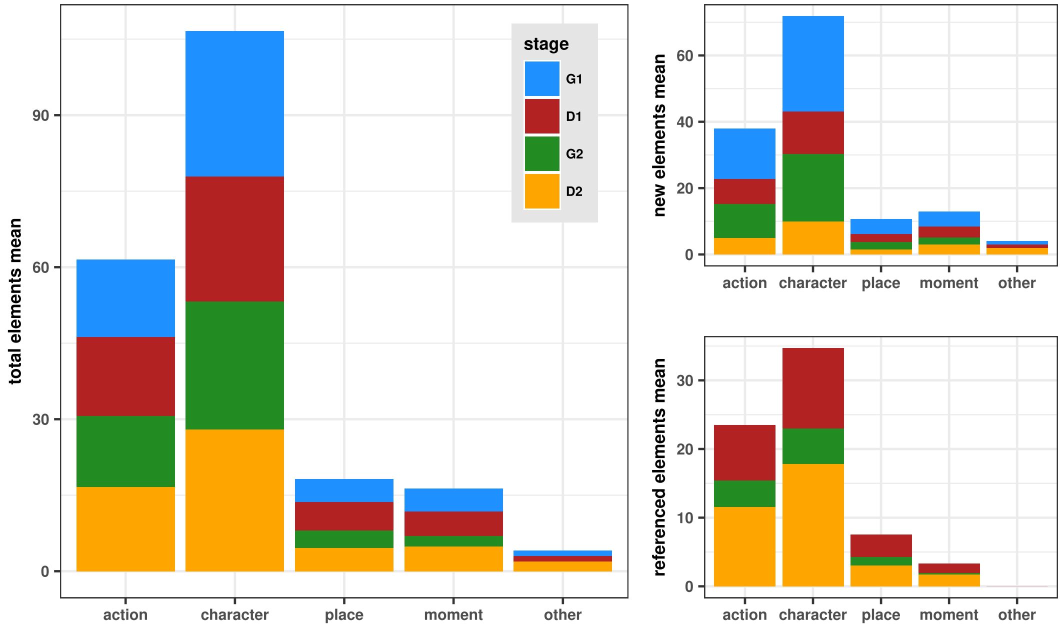

FIGURE 4. **Elements per stage. Left = used elements, top-right = average new elements per stage, bottom-right = average reused elements per stage.**

Once the kind of modifications have been analysed, we can do the same for the content. Table 2 shows the mean values and standard deviations of the elements per modification object. An element is the object of a tagged: the character, action, place or time the modification refers to. Figure 4 corresponds to the graphical representation of these data. In the figure, the leftmost chart represent all the used elements and the top-right chart represents the average new elements per stage, and the bottom-right chart represents the average reused elements per stage.

The table displays three different aggregations: the total elements, the new elements (those that appear for the first time in the corresponding stage) and the elements that are mentioned in the corresponding stage but have been already mentioned in a previous stage.

It can be seen that the standard deviations are relatively high. It will be seen that this is a common pattern in all the gathered data, and it is assumed to be inherent to the kind of experiment that has been conducted. It is discussed in detail in Section VI.

Table 2 summarizes the number of elements per stages. Most of the modifications applied to the draft are focused on the characters (53.22%). Actions are the second most frequent modifications (28.26%), followed by place (8.36%) and moment (8.28%). A marginal set of modification, tagged in the video analysis as other, represent 1.18% of the modifications.

These results strongly support that narratives are basically constructed as relations between characters, actions, times and places. The existence of reminder marginal tags also reveals that, while there are other aspects that take part in story creation, their statistical impact in terms of the plot is very low.

Table 2 shows the total amount of new elements created on each stage and the corresponding means and standard deviations, respectively. It is important to note that the references to actions, characters, moments or places do not make reference to the creation of these elements, but to the creation of parts of the story in which these elements play a main role. These tables focus on the new tags versus the old ones and therefore those modifications tagged as other do not play a role here and have been omitted in the table.

Both Table 2 and Figure 4 show that the number of creations decreases as the experiments advance, which can be seen as the process of a converging draft. That is, the successive refinements applied to the draft represent fewer changes for the four relevant aspects of a narrative. Besides, among these four relevant aspects, characters and actions are what take the most creations, and the subsequent changes are more noticeably reduced than for moments and places. We believe this corresponds to the participants focusing on the most relevant aspects of the plot first (characters and actions, according to the results in Table 2) and polishing off the less relevant parts (which would correspond to time and place) during the draft reviews.

Table 3 displays the corresponding means and standard deviations for the elements per stage. It can be seen how the stages in which the participants had to write a discourse (D_1 and D_2) present a high reuse, and most of the new tags appear on the first invention stage (G_1). Again, this supports the idea that participants built the narrative by conceiving concrete setting and actions, and then refined over iteratively.

TABLE 2. Elements per stage (mean/standard deviation). The table displays the total elements, those elements that are newly added, and elements mentioned in a previous stage.

	element	G_1	D_1	G_2	D_2
total	action	15.21 (7.42)	15.64 (6.19)	14.00 (8.13)	16.64 (5.14)
	character	28.64 (10.05)	24.64 (9.76)	25.36 (10.49)	27.93 (9.66)
	place	4.50 (4.01)	5.57 (2.79)	3.54 (2.79)	4.57 (2.95)
	moment	4.46 (2.33)	4.78 (2.72)	2.16 (1.40)	4.85 (2.82)
	other	1.00 (-)	1.00 (0.00)	0.00 (-)	2.00 (-)
new	action	15.21 (7.42)	7.57 (4.39)	10.14 (5.57)	5.07 (3.06)
	character	28.64 (10.05)	12.93 (6.92)	20.21 (8.42)	10.07 (6.65)
	place	4.50 (4.01)	2.33 (1.44)	2.30 (1.25)	1.50 (0.83)
	moment	4.46 (2.33)	3.42 (2.02)	2.00 (0.81)	3.07 (2.59)
	other	1.00 (-)	1.00 (0.00)	0.00 (-)	2.00 (-)
previous stage	action		8.07 (4.97)	3.86 (2.83)	11.57 (4.82)
	character		11.71 (4.45)	5.15 (3.84)	17.87 (6.53)
	place		3.25 (2.34)	1.24 (0.79)	3.08 (1.97)
	moment		1.36 (0.84)	0.16 (0.05)	1.78 (1.29)
	other		0.00 (-)	0.00 (-)	0.00 (-)

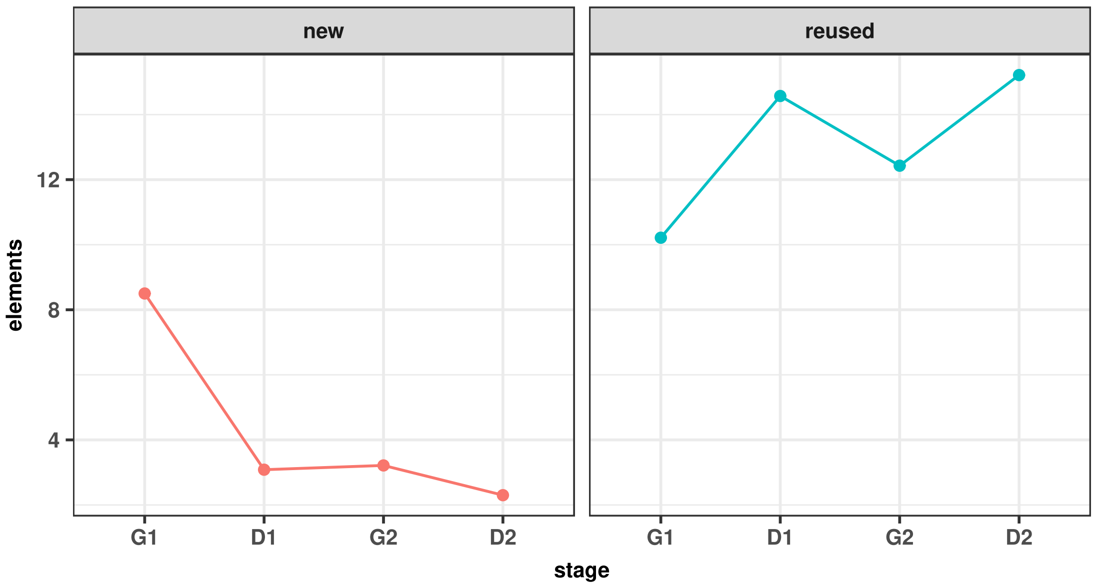

FIGURE 5. Graphical depiction of the mean new and reused elements along the four stages of the draft creation experiment.

TABLE 3. Means and standard deviations of new and reused elements per stage.

stage	new elements	reused elements
G_1	8.36 (4.87)	9.86 (5.07)
D_1	3.08 (2.53)	14.50 (4.45)
G_2	3.00 (1.52)	11.86 (5.67)
D_2	2.30 (1.70)	15.14 (3.66)

It is also relevant to note participants created new content in the stages in which they were supposed to transcribe the story (D_1 and D_2). There were explicitly allowed to create new content if they deemed it necessary, but they were not encouraged to do so: the task was to put down the draft in words, and they did not need to create an elaborate, lengthy discourse. However, all participants added novel content not purely linguistic or literary, but related to characters, actions, moments and time.

Figure 5 compares *new* and *reused* tags graphically. It depicts that G_1 is the really creative stage and the rest of the process refines the initial ideas, but these are not strongly modified.

Figure 6 shows one pie chart per stage. Each pie chart is divide in proportional sections that represent the objects that are modified on each stage. The diameter of each pie chart is proportional to the absolute number of created elements (the Y axis in Figure 5). Each labeled arrow between the diagrams represent the dependencies in terms of the reused tags: for instance, the arrow from G_2 to D_1 means that, on average, there has been 1.43 additions in D_1 that have been reused in G_2.

Figure 6 shows pie charts with the relative uses of tags. The charts also displays the average reuses as arrows between charts. The source of the arrow is the stage in which the element is referenced, and the destination is the stage on which the element was created. This figure summarizes two important findings: 1) most of the modifications are related to actions, character, places and time (which means that, at this level of draft creation, other aspects are marginal) and 2) there is a strong convergence in draft creation for the kind of setting that we used for the experiment: most new elements of an evolving story are created at the beginning, and the rest

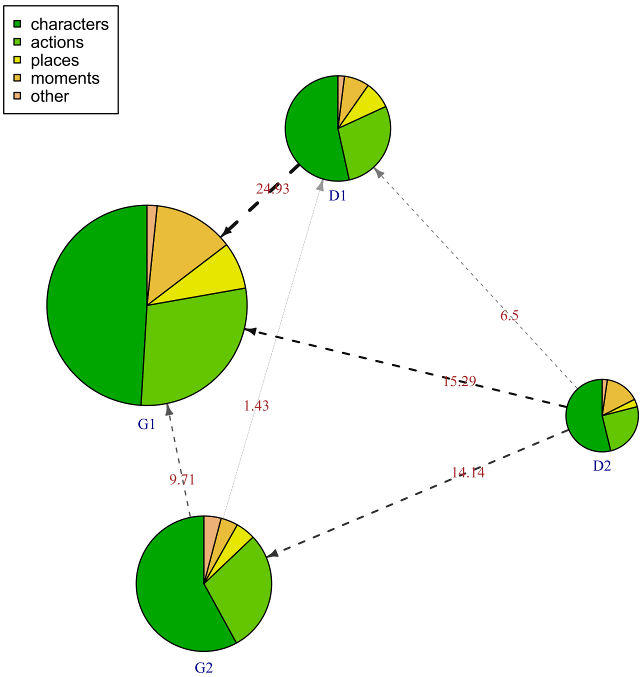

FIGURE 6. Pie charts with the relative uses of tags. The chart also displays the average reuses as arrows between charts.

TABLE 4. Average character mentions by story.

story	mentions
first	62.74 (21.50)
second	47.34 (18.12)

TABLE 5. Number of mentions per character.

	mentions	
character	first story	second story
1	70.92 (25.52)	56.43 (20.56)
2	62.64 (26.35)	46.57 (18.12)
3	54.64 (23.77)	41.50 (15.88)
4		44.86 (10.52)

of the process refines these objects and adds less relevant details. The setting itself, however, has an influence om the participants' performance. This is discussed in Section VI.

The previous results led us to analyse the use of character tags in more detail. Since characters seem to be the most relevant element in draft creation, an extra analysis can provide insight on the way in which they are used.

By taking a closer look at the amount of references, we found out that the references to characters are different between stories. Table 4 shows the average mentions per story. Characters are more frequently mentioned on the first story than in the second one ($\chi^2 = 11.438, p < 0.000$).

Moreover, characters created earlier are more frequently mentioned. The first mentioned character corresponds to the protagonist and is the most mentioned one ($\chi^2 = 9.643$, $p < 0.03$). Table 5 summarizes the mentions by stages. The information is depicted in Figure 7.

Only color ball pens (green, red, blue and black), white paper sheets and color post-it notes were used by the participants. On average, participants used 2.28 paper sheets ($max = 3, min = 1, sdev = 0.72$) and 2.42 colors on average ($max = 4, min = 1, sdev = 1.34$). Three (3) participants used post-it notes. Most notes taken were names, actions, lists and simple diagrams or arrows between the characters. The annotations were mostly one to three words long, but some participants did write short sentences. Most of the annotations can be interpreted as textual, although some participants added schematic drawings to their notes. A qualitative examination of the videos suggests that the participants were taking notes more as a canvas to lay down their ideas during the creation of the story than actual notes for future reference. During stages D_1, G_2 and D_2 they did refer to their notes, but it seems that they did so to look for some details, like names or places. There is not any note set from which the story can be reconstructed just by reading the annotations.

In terms of the demographics, we found a difference between the number of characters created by female and male participants (Table 6). Female participants seem to refer to characters more often than male participants ($Z = 1.69$, $p < 0.05$) and make references to previous episodes in the story less often ($Z = 1.77, p < 0.05$). However, the number of female participants was relatively low and the number of data is not sufficient to draw any strong conclusions. Again, the standard deviations were relatively high, which is addressed in Section VI.

TABLE 6. Average mentions to characters (character tags in the annotations) for female and male participants (standard deviations between parentheses).

gender	characters
female	141.33 (36.11)
male	102.10 (22.68)

We found that there is a strong correlation ($r = 0.836, p < 0.000$) between being satisfied with the first story (the one conceived in G_1 and D_1) and being satisfied with the second one (G_2 and D_2).

Participants liking their first story also liked their second one ($r = 0.7, p < 0.000$). Besides, those liking their first or second story used to read narrative ($r = 0.5, p < 0.000$) and consider themselves creative people ($r = 0.55, p < 0.000$). We also found a correlation between liking the first story and the number of actions in that story ($r = 0.68, p < 0.03$). A similar correlation could be found in the case of liking the second story and adding actions to it ($r = 0.64, p < 0.05$). This could indicate that making more modifications to a story means taking more creative decisions and the participants like it more, but these correlations still cannot provide conclusive causality on this effect.

There was a medium correlation, although not statistically significant ($r = 0.47, p = 0.17$) between liking the experiment and the use of secondary characters—characters which are assumed not to be the protagonist because they are not the first created ones nor the most used ones. In any case, this seems to be just a trend which the results cannot confirm yet. A strong correlation was found between those declaring

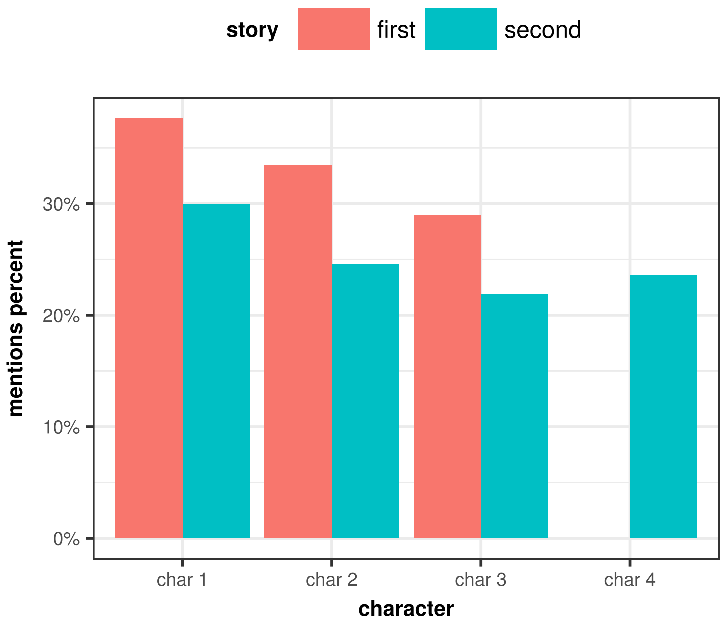

FIGURE 7. Percent of mentions per character, according to the character tags.

that reflection on the process out loud was actually helpful for composing the story and the number of creations ($r = 0.83$, $p < 0.03$) and uses ($r = 0.62$, $p < 0.05$) of tags related to secondary characters.

We asked the user for their skills with the tools involved in the experiment. All participants considered that they had a high (8 participants, 42.85%) or medium-high (6 participants, 57.14%) level of typing. All of them knew how to write simple text in a word processors: 1 (7.14%) declared medium level 8 (42.85%) declared high level and 5 (35.71%) declared very high level. Regarding the general use of a computer, 3 (23.97%) declared to have high level skills for using a computer, and 11 (78.56%) declared to be highly skilled. These results indicate that using a computer and a text processor to input the stories was not a problem for the participants.

All participants (100%) considered that they created the story with full freedom. The experiment was pleasant for 13 (92.85%) and not particularly pleasant, but not unpleasant by 1 (7.15%). This indicates that the process, while controlled, let the participants be creative without problems. We hypothesize that this high user satisfaction was partly because the subjects were not professional writers, and they were not extremely concerned with the quality of the produced material.

Other correlations were tested (for instance, between narrative skills and number of modifications) but no conclusive results could be found.

VI. DISCUSSION

The experimental results previously summarized in Section V support the main theories on which this work has been founded. The observed behaviour indicates that participants do carry out operations (namely creation) that correlate well with the generation of material, contemplated as translating ideas into text by Flower and Hayes and as engagement by Sharples. But they also carry out a number of operations (fixations and revisions) that correlate well with activities of revision (Flower and Hayes and Sharples) and contemplation (Sharples). The prevalence of creations over fixations and revisions seems to indicate that, in terms of Sharples' analysis, the participants in the experiments gave priority to engagement over reflection. This is consistent with the fact that they were operating under a tight time constraint. This point is elaborated further below.

The focus of this research has been put on lower level details of narrative draft construction, and as such it potentially provides useful insight on new aspects of the process. While the provided results seem stable, there are a number of open issues that are worth discussion.

The statistical analysis yielded relatively high standard deviations for most of the gathered values. While, in general, this might indicate a low level of commonality between participants, we believe these levels are acceptable because of the very nature of the experiment.

When measuring any form of creative performance, differences are to be expected. A very low standard deviation would have meant that the experimental process was too restrictive in terms of the freedom that the participants would have needed, and this would have produced invalid results. In particular, narrative generation is a very complex cognitive process involving many high level cognitive features: memory, emotional state, narrative skills, and many others. The fact the task set to the participants involved relatively unconstrained generation of stories allows for a wide range of fictional situations. Each participant was free to choose characters, events, and settings with no interference from the experimenters. Despite the broad range of options that this allowed, we have found a set of common patterns for story invention shared between all the participants.

The experiments consisted of four stages of 10 minutes each. Setting this hard time constraint made the participants focus on shaping the story so that the generation was doable within that time period. Participants were constantly aware of the time, and all of them focused on finishing the story at the expense of adding more details or polishing off the final result. We believe that the resulting stories would have been different if no time constraints were given. Additionally, more revisions could have taken place in a setting in which the participants had more time. However, the quality of the stories was not the objective of the study, and we argue that having the participants focus on the story and not on the process made it possible to actually observe what they do in order to produce a story.

Nevertheless, we cannot ignore that the process of producing long, elaborate narratives is more complex than we were able to observe in the experiments. In principle, this might seem to restrict the application of the results to the creation of short narratives in a short time. While we have not provided any empirical evidence supporting the applicability of the results to other forms of narrative construction, we hypothesize that the fundamental aspects that have been studied do not exclusively happen in the context of small narratives. We make this hypothesis because all our findings are in line with the reviewed cognitive models of narrative creation. While this might have been driven by the fact that the experimental setup is based on the theoretical premises they propose, we believe our statistical results are robust enough to evidence their validity. This compatibility between the large scale model and the low level evidence makes it plausible to believe that the conclusions could be extended to larger draft creation.

One important aspect to discuss is the relevance of the current results in terms of implementability. We believe that the current results can provide helpful insight about two aspects of draft construction: what are the main components that humans use to build plots, and the relative importance they play in the story. The results show a relatively stable set of procedures and components, and it is both clear and in line with previous work, that a few components are the most relevant. This information can inform computational models by establishing parameter ranges, for instance, or by helping modelling knowledge bases meant to work with narrative systems. In any case, further research is needed in order to provide a full computational model of story draft construction.

VII. CONCLUSION

The creative process of narrative text composition has been studied thoroughly, but the literature only provides high level accounts for it. In order to provide a lower level model that is still backed by empirical evidence of the actual cognitive processes, we have run and analyzed a number of experiments to study how human subjects evolve and refine a narrative. In these experiments, participants evolved a narrative in four stages with varying constraints. The results were analysed under cognitive and computational models of writing.

According to the results, and in line with the conclusions of previously described cognitive models of writing, there is a very strong similarity between participants in terms of the process they follow. Moreover, additional conclusions can be drawn when looking at the draft as the creative object and its successive refinements as applied by the writer. These are enumerated below.

Participants created the draft by focusing on the most relevant parts at the beginning, and they came up with a complete story in which most of the details were included in subsequent stages. This means that **most of the main ideas are conceived at the beginning**, and the subsequent modifications refine, fix and complete them. This contrasts with most computational approaches of story generation in which the production happens in a more greedy way by chaining events until a valid plot is found.

Most of the operations happen at the level of **characters** and **actions** (accounting for roughly 70% of the objects) and **places** and **moments** in a lesser, but still important amount (around 20%). This is certainly relevant in terms of what objects must be the constituent ones in computational implementations of narrative composition, but it is also informative in terms of a further study on what are the important elements of stories for humans. Additionally, this suggests that these four tags are a valid model for future annotation tasks, given that all other modifications accounted for little more than 1.15% of the operations. With regard to computational systems, this result suggests that focusing on the semantics and relations of these four aspects could be enough for representing short stories.

The characters have different relevance, and **the protagonist is created first**. This suggests that the narrative draft is mostly created as the story of one character in which other characters interact. It is quite likely that longer narratives need more protagonists and that really complex plots can even change the protagonist, but the results show that focusing on one single protagonist and centering the action around her is a common strategy. This can also inform computational approaches, since it is clear that the protagonist is not only semantically important, but is also mentioned more and is the agent and subject of more actions in the story.

Plot elements are created during story conception and during discourse creation. That is, discourse generation also affects the plot, and new elements are usually added to it during the realization of the story as text. However, the kind of elements that are created is different. While writing the discourse, authors tend to focus on details not taking part in the main plot, but decorating it and providing more context and appropriate environment. This seems to be less important during raw plot composition. This provides evidence to theoretical models proposing that narrative creation is far from a linear task, and writers swap from generation to transcription constantly.

It is quite likely that the reason why plot content is also created during discourse is because the task of producing a story merges linguist phenomena with memory retrieval, revision and other tasks without specific order. Given that computational systems do not necessarily have this constraint, full plots could be created without the need of alternating plot and discourse generation. However, our findings suggest that chaining plot generation and discourse generation does not properly model how humans write.

Further analysis must be carried out in order to identify more observable modification of narrative drafts by humans. We believe this is a potentially useful path for discovering specific actions, and not only high level descriptions. This is expected to provide material that is more easily translatable to computational models of narrative.

ACKNOWLDEGMENT

The authors would like to thank the voluntary contribution of all the participants. We would also like to thank L. Zardaín and J. Muñoz for their help with the experiments.

REFERENCES

[1] J. Bruner, "The narrative construction of reality," *Critical Inquiry*, vol. 18, no. 1, p. 1, 1991.
[2] D. Herman, *Narrative Theory and the Cognitive Sciences*. Stanford, CA, USA: CSLI Publications, 2003. [Online]. Available: https://books.google.com.au/books?id=6bsbAQAAIAAJ
[3] D. Herman, *Storytelling and the Sciences of Mind*. Cambridge, MA, USA: MIT Press, 2013. [Online]. Available: https://books.google.es/books?id=DSQbAAAAQBAJ
[4] T. S. Anderson, "From episodic memory to narrative in a cognitive architecture," in *Proc. 6th Workshop Comput. Models Narrative*, 2015, pp. 2–11.
[5] N. Szilas, "Towards narrative-based knowledge representation in cognitive systems," in *Proc. 6th Workshop Comput. Models Narrative*, 2015, pp. 133–141.
[6] C. León, "An architecture of narrative memory," *Biologically Inspired Cognit. Archit.*, vol. 16, pp. 19–33, Apr. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212683X16300184
[7] L. Flower and J. R. Hayes, "A cognitive process theory of writing," *College Composition Commun.*, vol. 32, no. 4, pp. 365–387, Jan. 1981.
[8] M. Sharples, "An account of writing as creative design," in *The Science of Writing: Theories, Methods, Individual Differences and Applications*, C. M. Levy and S. Ransdell, Eds. New York, NY, USA: Lawrence Erlbaum Associates, 1996.
[9] T. A. van Dijk and W. Kintsch, *Strategies of Discourse Comprehension*. New York, NY, USA: Academic Press, 1983.
[10] T. Trabasso, P. vand den Broek, and S. Suh, "Logical necessity and transitivity of causal relations in stories," *Discourse Processes*, vol. 12, pp. 1–25, Jan. 1989.
[11] P. Gervás and C. León, "Reading and writing as a creative cycle: The need for a computational model," in *Proc. 5th Int. Conf. Comput. Creativity*, Jun. 2014, pp. 365–387.
[12] P. Gervás and C. León, *Corregir e Innovar Corrigiendo: Modelos De Creatividad Literaria Como Ciclo De Reflexión-revisión*. Mexico, NM, USA: UAM-Cuajimalpa-Patria, 2015.
[13] P. Gervás and C. León, *Integrating Purpose Revision Into a Computer Model Literary Generation*. New York, NY, USA: Springer, 2016. [Online]. Available: http://www.springer.com/us/book/9783319244013
[14] N. J. Dehn, "Computer story-writing: The role of reconstructive and dynamic memory," Ph.D. dissertation, Dept. Comput. Sci., Univ. Yale, New Haven, CT, USA, 1989.
[15] M. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *J. Artif. Intell. Res.*, vol. 39, pp. 217–268, Apr. 2010.
[16] Y.-G. Cheong, R. M. Young, and R. M. Young, "Suspenser: A story generation system for suspense," *Trans. Comput. Intell. Artif. Intell. Games*, vol. 7, no. 1, pp. 39–52, Mar. 2014.
[17] B. O'Neill, "A computational model of suspense for the augmentation of intelligent story generation," Ph.D. dissertation, School Interact. Comput., Georgia Inst. Technol., Atlanta, Georgia, 2013.
[18] B.-C. Bae and R. M. Young, "A computational model of narrative generation for surprise arousal," *Trans. Comput. Intell. Artif. Intell. Games*, vol. 6, no. 2, pp. 131–143, 2014.
[19] J. Niehaus and R. M. Young, "Cognitive models of discourse comprehension for narrative generation," *Literary Linguistic Comput.*, vol. 29, no. 4, pp. 561–582, 2014. doi: 10.1093/llc/fqu056.
[20] R. P. Pèrez "Mexica: A computer model of creativity in writing," Ph.D. dissertation, Univ. Sussex, Brighton, U.K., 1999.
[21] M. Sharples, *How We Write*. Evanston, IL, USA: Routledge, 1999.
[22] P. Gervás, "Evolutionary elaboration of daily news as a poetic stanza," in *IX Congreso Español Metaheurísticas, Algoritmos Evolutivos Bioinspirados*, Sep. 2013, pp. 229–2238.
[23] P. Gervás, "Computational modelling of poetry generation," in *Proc. Artif. Intell. Poetry Symp.*, 2013, pp. 11–16.
[24] C. León and P. Gervás, "A top-down design methodology based on causality and chronology for developing assisted story generation systems," in *Proc. 8th ACM Conf. Creativity Cognition*, Jan. 2011, pp. 363–364.
[25] C. León and P. Gervás, "Prototyping the use of plot curves to guide story generation," in *Proc. 3rd Workshop Comput. Models Narrative*, May 2012, pp. 1–12.
[26] P. Gervás, "Composing narrative discourse for stories of many characters: A case study over a chess game," *Literary Linguistic Comput.*, vol. 29, no. 4, pp. 511–531, Aug. 2014.
[27] S. Chatman, *Story Discourse: Narrative Structure Fiction Film*. Ithaca, NY, USA: Cornell Univ. Press, 1986. [Online]. Available: https://books.google.es/books?id=ZE42XwAACAAJ
[28] B. Li, S. Lee-Urban, G. Johnston, and M. O. Riedl, "Story generation with crowdsourced plot graphs," in *Proc. AAAI*, 2013, pp. 598–604. [Online]. Available: http://dl.acm.org/citation.cfm?id=2891460.2891543
[29] M. Roemmele, "Writing stories with help from recurrent neural networks," in *Proc. AAAI*, Feb. 2016, pp. 4311–4312. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11966
[30] C. León, "Narrative as a fundamental information unit in the mind," in *Proc. AAAI*, Oct. 2017, pp. 408–413.
[31] S. Chatman, *Story Discourse*. Ithaca, NY, USA: Cornell Univ. Press, 1978.
[32] M. Kipp, "ANVIL: The video annotation research tool," in *Handbook Corpus Phonology*. Oxford, U.K.: Oxford Univ. Press, 2014.

CARLOS LEÓN is currently an Assistant Professor with the Complutense University of Madrid. His research interests include the use of artificial intelligence, computational creativity techniques for leveraging interactive media and narrative, cognitive aspects involved in interactive communication to the engineering of systems that communicate humans and computers. He is a member of the Steering Committee of the Association for Computational Creativity.

PABLO GERVÁS is currently an Associate Professor with the Department of Software Engineering and Artificial Intelligence, School of Informatics, Complutense University of Madrid. He is also the Director of the Natural Interaction based on Language (NIL) Research Group. Over the years, his research interests have shifted toward studying the role of narrative in human communication, with a view to applying it in human–computer interaction. Besides his generic interest in understanding and modeling language, he tries to tackle computational models of human creativity. His research interests include natural language generation, natural language analysis, NLP for accessibility, and NLP and literary artifacts.

PABLO DELATORRE received the B.Sc. and M.Sc. degrees in computer engineering from the Universidad de Sevilla, and the Ph.D. degree from the Complutense University of Madrid (UCM). He combines this research activity with his responsibilities as a Teacher with the University of Cadiz. He worked ten years in the ICT sector, managing national and international corporate projects. His research interests include narrative and interactive suspense applied to automatic story generation.