



# Quantitative Characteristics of Human-Written Short Stories as a Metric for Automated Storytelling

Carlos León<sup>1</sup> · Pablo Gervás<sup>1</sup> · Pablo Delatorre<sup>2</sup> · Alan Tapscott<sup>1</sup>

Received: 21 March 2020 / Accepted: 29 September 2020  
© The Author(s) 2020

## Abstract

Evaluating the extent to which computer-produced stories are structured like human-invented narratives can be an important component of the quality of a story plot. In this paper, we report on an empirical experiment in which human subjects have invented short plots in a constrained scenario. The stories were annotated according to features commonly found in existing automatic story generators. The annotation was designed to measure the proportion and relations of story components that should be used in automatic computational systems for matching human behaviour. Results suggest that there are relatively common patterns that can be used as input data for identifying similarity to human-invented stories in automatic storytelling systems. The found patterns are in line with narratological models, and the results provide numerical quantification and layout of story components. The proposed method of story analysis is tested over two additional sources, the ROCStories corpus and stories generated by automated storytellers, to illustrate the valuable insights that may be derived from them.

**Keywords** Empirical quantification · Narrative features · Story components · Story metrics · Story evaluation

## Introduction

Creating story generation systems is a complex task. The number of features that can play a role in the generation or the evaluation of automatically generated stories is large, as evidenced by the heterogeneity of systems described in the literature. These

---

This work has been supported by the CANTOR project (PID2019-108927RB-I00) funded by the Spanish Ministry of Science and Innovation; by the project FEI INVITAR-IA (FEI-EU-17-23) of the University Complutense of Madrid; and by the ComunicArte project (PR2005-174/01) by BBVA Foundation Grants-Scientific Research Groups 2017.

---

✉ Carlos León  
cleon@ucm.es

Extended author information available on the last page of the article

features include aspects related with the story world like emotions, characters, locations or intentions, and structural aspects like length or narrative arc. Some of these features need explicit or implicit values for the generation, as setting the appropriate length, the number of characters, or the amount of descriptions that the story needs. Additionally, the parameters for the story features can change depending on the kind of the story, author, and context.

Choosing the optimal values for these parameters is not a trivial task, since the range of acceptable values for many of the features is large and the features are rarely independent. However, one potential source for information is stories written by humans. A quantitative analysis of the features present in stories produced by humans can provide a set of values that can be considered a characterization of high-quality stories, which can then be used to inform the construction of better stories in story generation systems. In doing so, the generative process could approximate human output, which is not necessarily optimal from the point of view of creativity. In any case, gaining more insight on how to emulate the human-writing process is a good resource.

Following this approach, the present paper addresses the task of identifying empirically a set of quantitative measures on some basic components for a story and structural relations between them as observed in human-produced samples of simple narratives. Because it is intended to inform the improvement of existing story generators, the paper starts by identifying what basic components for a story are shared by a range of existing story generation systems. For those elements identified, an experiment is proposed for constructing a sample of human-generated stories under controlled conditions. Then, a procedure for annotating these stories in terms of the desired components is designed. Statistics on the observed features are compiled and discussed, to be used in further work to inform the development of selection metrics for the improvement of story generation systems. As an example of how these measures can be applied, stories from two different sources are discussed in these terms.

The research is based on two hypotheses: (1) there is a basic set of characteristics that are generally associated with the concept of story, and (2) these characteristics tend to show up in similar proportions, given a specific scenario of simple, direct plots.

Additionally, the presented study is carried out under the assumption that a better understanding of human-like narrative structures can lead to improve computational storytelling. While this is considered plausible as a base hypothesis, it is also possible that machine-based structures or metrics produce equally good or better stories, even from a human evaluator perspective. While this is accepted, the current study is purely focused on the human-like process of story generation.

To provide insight on the validity of these hypotheses, this paper summarizes an empirical exploration in which human participants were asked to invent short plots. The participants were not required to have specific narrative skills, since the study focuses on commonly used narrative structures. This aspect contextualizes the conclusions and constrains the application of the results to stories that are not assumed to have very good literary quality, in contrast to stories written by highly qualified writers.

These plots were annotated according to a structured framework extracted from a sample study of existing computational creativity systems that produce stories. The annotation was statistically analyzed to identify the amount and characteristics of the story components that the participants use. The results, summarized in “[Results](#)” section, are intended to serve as a resource to inform future automatic story generation and understanding systems, and to kick off exploration on empirical human story production as a source of input material for computational creativity systems.

The rest of this paper is structured as follows: “[Previous work](#)” section contains previous existing work deemed relevant to this work; “[Components and relations for analysing human-written texts](#)” describes the set of features that have been deemed worthy of abstraction during the annotation process, and justifies the choice based on existing systems; the experiment and the annotation process are described in “[Experiment](#)”; and the results are presented and discussed in “[Results](#)”. Section “[Testing Resulting Insights over Additional Sources](#)” provides examples of how the proposed measures can be applied to better understand a set of stories—the ROCStories corpus—and how they can help to understand the operation of automated story generators in comparison with respect to human ability—by analyzing a small set of stories produced by such systems. The relevant conclusions are summarized in “[Conclusions and future work](#)”.

## Previous Work

Automatic story generation systems have explored distinct ways of producing narratives. Some construct a plot from a structured template, some develop narrative constituents into literary discourses relying on brute force generation, and others reproduce existing patterns found in narrative texts without concern for formalisms or structure. This procedural heterogeneity reflects the wide range of reasons the stories are built for and result in an equally rich set of creative artefacts.

Some systems attempt to reproduce a creative process [13, 33, 38, 43, 46], while others are focused on generating a human-like narrative [8, 44, 50]. Some of these systems also introduce interaction, a focus on video games or even attempt to extract and reproduce semantics [17, 26, 28].

## Relevant insights from narratology

There are significant differences in the underlying computational models of story generation systems present in the literature. Those that use a narratologist conceptual basis implement different approaches. This plurality is also extensive to narratology studies. For instance, classic narratology started as a structuralist study of narratives [3, 12]. Although, from different perspectives, structuralist narratology tries to identify the fundamental features of stories. In particular, one of the most typically agreed aspects of narratives is the difference between what is told and the particular way chosen to tell it—including the order of presentation to the reader. Different authors assign different names to these two

concepts (*story* and *discourse* [6], *histoire* and *recit* [12], or *fabula* and *story* [2]—with *text* an additional aspect representing the sequence of words that actually convey the story). Adhering to this perspective and to synthesize these concepts, we present the *fabula* as the set of all events taking place in a story, and the *discourse* as the concrete realization of these events in a linear sequence meant to be transmitted. In other words, the *fabula* is *what* happens, and the *discourse* is *how* the story is told [34].

The plot is a different concept that is often taken into consideration, understood as the structured sequential subset of story events linked by their causality. Forster argued the important difference between a *chronology*—a sequence of events described in order—and a *plot*—a sequence of events linked together by a causal thread [10]. Forster's argument suggests that volunteers asked to narrate a particular experience that they had been exposed to might have difficulty in differentiating a rendering of the experience as a story from a rendering of the experience as a straightforward report. All these aspects influence how the story is being told and read, i.e., with the discourse domain, leading to story world creation [27].

Formalizations of lower level descriptions of the constituents of narrative have been explored as a bridge between narratology, cognition, and computation. For instance, the seminal work by Schank and Abelson introduces a set of primitives for formalizing short abstract narratives, or *scripts* [45]. More specific properties of narrative like time [1, 57] or causality [16, 51, 52] have also been addressed.

Nowadays, some modern narratologists argue that narrative cannot be considered simply as a structure, and propose looking at narratives from a more cognitive perspective [5]. For instance, [19] proposes that narrative is not a story, but a specific way of reasoning about reality according to a *story logic*. More recently, some authors defend the *narrative hypothesis*, namely that narrative is a fundamental cognitive structure [47]. In this fashion, a computational model of cognitive narrative [24] and its relation with a more general cognitive architecture [25] have been explored.

This disparity of methodologies and tasks (and their corresponding combinations in the different systems) suggests that there is a lack of consensus about what constitutes a narrative in computational terms. The literature provides several definitions that inevitably span several research fields including computer science, cognitive science, narratology, computational creativity and literary studies. As such, and as another evidence of the wide range of options, each field provides specific insights and perspectives, that are some times not satisfactory or useful to other disciplines. In particular, there is a well-known gap between the views arising from cognitive and narratological perspectives and what can be actually carried out computationally. The definitions of the constituents that cognitive narratology provides, while operational and popular, can be elusive from the point of view of a systematic implementation, because these definitions are not grounded on formal elements easily usable by a computer program.

## Relevant Insights From Metrics on Narrative

The quantitative analysis of narratives has long been subject of studies, featuring a rich collection of approaches and focuses. [14] compares plot generation procedures grounding all of them on a basic reference vocabulary for the representation of narrative units using metrics distilled from the same procedure. The metrics rely on previous work based on [35] along with plot schemas mined from existing literature on plot and are mostly based on text similarity, conformance to a model, ratio of satisfied dependencies and ending validity. [48] evaluates the generated short stories (named mini narratives) using a wide range of metrics that include internal generation parameters, hits in search engines, length, amount of characters, and other varied objective measures. [20] proposes to calibrate a metric for story similarity using human judgement, which is compared to plan refinement (a formal description of the common structure shared by two formalized stories). [15] presents simple representations of a chess game in algebraic notation. A number of candidate metrics are explored in relation to their variability over a number of games and their impact on the resulting tellings of a game. [31] attempts to create story novelty metrics based on empirical aspects identified in a survey with human evaluators, mostly focused on event, character, and prop novelty. The dimensions of conflict (*balance*, *directness*, *intensity*, and *resolution*) have also been studied, quantified, and checked experimentally with success [56]. [41] presents an evaluation framework for empirically studying computational models of narrative generation that includes four complementary tools for evaluating interactive and non-interactive narrative generation. The representation and evaluation of stories is addressed in [54], proposing a linear genome representation and then using a human collective to measure stories and produce data that are then correlated with quality objective metrics. Finally, [36] introduces a set of story features that correlate with the human judgement of stories along with algorithms meant to measure such features. When it comes to our goal (comparing human and computer story invention), the main limitation of these approaches is that they rely on empirical observation of complex stories and models overlooking the story invention process. Finding essential constituent features attributed to human-invented narratives can be difficult in the context of studies framed as observations or analyses of stories that have not been conceived in a controlled manner. Overall, they represent comprehensive attempts to quantify story features. However, given the lack of insight on the conditions in which the original stories have been conceived, or the fact that they analyze formalist story models or existing computational story generators, it might not be safe to assume that the empirical factors derived from these studies inform a human-like process of story invention.

## Relevant Insights from Existing Story Generation Systems

This section analyzes a number of story generation systems that are frequently referenced in the field of computational creativity and computational narrative. A number of reported systems have not been included in this review either because they do

not include any feature not considered here or because the underlying model is not applicable to this study. The family of text-oriented machine learning systems that do not have an underlying structural representation has not been explored [37], since their processing models fall outside the application of this work.

## MINSTREL

MINSTREL [53] relies on a library of pre-existing stories to generate new ones during the execution of its generative function. The author-level planning component is responsible for connecting the solutions provided by the TRAM system, based on the stories stored in the library. The author plan is also responsible to properly order the provided solutions in a coherent way using PATs (Planning Advice Themes). Characters are introduced into the system as entities with their own goals, although secondary to the schema described by the PATs and the author plan. In order for these PATs to be used, the system constantly verifies the consistency and adds the necessary world facts/pre-condition sentences. The notions of ordering, causality, and character entities are deeply embedded in MINSTREL's design. Every story starts with a introductory scene that introduces facts into the story world through a brief description. These facts are later referenced and modified through actual action-based plot developments. The result is a story that has descriptive sentences to establish facts that are later used in the action sentences that drive the plot forward, ensuring the resulting sequence of events maintain a sense of causality and happen in an order that satisfies the author's own conception of casualty. All developments of the story are verified in terms of character plan, story world, and emotional consistency, and several constraints exist also to maintain consistency adding or modifying the story. Since the generator dynamics of MINSTREL generate actions performed by the characters, the strategy to make them more consistent relies on adding new, related descriptions and facts (e.g., the king collected trophies, Frederik wanted to impress the king) to the state of the story world that the incoherent actions (Frekerik killed the dragon) can refer to, resulting in a reinforced sense of causality.

## Universe

Universe [23] is meant to generate soap opera episode scripts. It focuses on complex data structures that represent character data, and relies on algorithms and user input to develop several overlapping storylines. Data structures for a universe of coherent and engaging characters are developed before plot units (story fragments), populating the world and establishing explicit relationships before creating the plot. The complex data for every character include traits and quantified interpersonal relationships (*Liz Chandler is basically a rich, intelligent socialite who is currently married to Tony Dimeria, but hates him*). Characters also possess goals and subgoals with pre-conditions that are meant to create conflict with other characters. The system then proceeds by picking goals with no missing pre-conditions and implements plot developments to achieve such goals, changing the state and potentially fulfilling pre-conditions for more goals. This allows developing several plots in parallel imitating

actual soap operas. Instead of adding an early description of the world and then referring to it to develop the plot, Universe uses a more indirect approach that arguably follows the same conceptual sequence. Characters are created early on and then referred to reinforce causality as the plot unfolds. The initial characters generated possess traits and relationships, but no explicit depiction of previous actions, favouring an initial descriptive representation of the world. These characters are then used to implement plot units that unfold action. These two dynamics describe a creative process of establishing a base textual description of the world (the descriptive character set) and then adding chained actions plot units that rely conceptually on the previous plot unit or the initial character set descriptions.

### **Virtual Storyteller**

The Virtual Storyteller [49] features a director agent with plot structure knowledge that has environmental, motivational, and proscriptive control of the story world. The two assumptions underlying its design are that a plot must be consistent and well constructed. To achieve these goals, the director agent manipulates the process by adding or removing elements (environmental control), changing the character goals (motivational control), or forbidding certain actions (proscriptive control). The director, however, cannot directly enforce any action. Characters have beliefs, desires, and actions. In this case, the distinction between the descriptions meant to establish facts and the chain of actions that rely on such facts is implicit in the basic creative loop. A director agent, compelled by its plot structure knowledge, uses control devices to introduce the optimal elements and motivations that trigger actions that not only represent a well-constructed story, but also are coherent inside its virtual logic. There is a clear separation between the descriptive content manipulated by the director agent that configures the story world, and the action developments conducted by the character agents. The resulting process features descriptions that are introduced before the actions. New actions keep referring to each other backwards until a new description is required to keep the story consistent with the director's structural knowledge.

### **Fabulist**

Similarly, Fabulist [40] is a multi-tiered automated story generator that relies on an intent-driven partial-order causal link (IPOCL) planning algorithm meant to produce sequences that represent causally coherent narratives and believable characters. In this case, there is an explicit formal concern for character believability and causality that results in a chain of descriptions and actions that reflect such concern. The resulting story alternates between establishing facts in a descriptive way (*King Jaffar is not married, Jasmine is very beautiful*), asserting actions that drive the story forward (*King Jaffar falls in love with Jasmine*) and stating character desires and intentions that boost the coherence and causality of the story (*Aladdin is loyal to king Jaffar*). While these loops are short and iterate very frequently, there is a clear effort to separate these sentence types and keeping them causally connected following a coherent sequence.

## MEXICA

MEXICA [32] generates short stories about the native inhabitants of Mexico in two cyclic states of engagement and reflection. The generator relies on a set of story actions to seed the story during the engagement state, and adds post-conditions necessary for the story to be coherent during the reflection state. While story actions are based on action, the pre-conditions are descriptive, establishing facts that can then be exploited in a chain of events that follows later in the resulting causal story (since the pre-conditions now allow to follow a cause-consequence logic).

## BRUTUS

BRUTUS [4] writes short stories based on pre-defined themes. Its generation process first instantiates a thematic frame, then runs a simulation-process for the characters to achieve pre-defined goals, and finally expands story grammars that the final output. The thematic frames include a sequence of events. These events are developed during the simulation process in which the characters perform character actions with a mixture of proactive and reactive behaviour. According to the author, proactive behaviour is meant to maintain the thematic theme of the story, while reactive behaviour aims at increasing the variability. While proactive behaviour relies on actions with pre-conditions, reactive behaviour relies on actions with conditions. Proactive behaviour uses pre-conditions to emphasize the necessary events or established facts that need to have happened before the candidate action. The pre-conditions found in the provided examples seem to use descriptive language (e.g., candidate is some person). Reactive behaviour relies on the consequences of previous actions to trigger character actions such as archetypical actions or dialogue. We observe again how descriptive content is established before the following actions, always ensuring that the causal implications enforced by the thematic frame are respected and the actions' own conditions are fulfilled in order.

## Author

Another view on story generation, reflected in the design of the Author system [7], is that story worlds emerge as the story is written. The Author system models the mind of the author as the story is written, representing the knowledge involved in the process (e.g., characters or memorable episodes) and how it is organized. While this does not follow the pattern of description and action loops seen in similar systems, it still holds a distinction between establishing facts (modelled after the human author's memory) and using them to have the action unfold (the story proper). The connection between a plot development and the knowledge it uses is explicit, but causal event chains are not so linear in this case. Instead of having some facts established in a description that are then developed through sequential and connected actions, this model proposes to establish a knowledge representation separated from the story. Both share an implicit causal connection, ensuring a coherent resulting story output.



## Neural Story Generation

More recently, [9] propose a story generator that relies on a large dataset of existing stories and also a complementary dataset of prompts (an online collection of inspirational story premises). Their approach seeks to improve the coherence and structure of stories using hierarchical generation from a textual premise. By fusing more traditional language-based models with statistical sequence to sequence models, it first generates a base premise which is later used to generate a larger passage of text. This hierarchical approach is used to ensure that consistency is present in the result by relying on the high-level plot (the premise or prompt) to build the actual text story. The emphasis in the inheritance of coherence, implicit in hierarchical generation, can be read as a concern for maintaining causal links between several story representations. Regarding sequence to sequence modelling, an innovative and computationally powerful application of neural networks specialized on sequence encoding and decoding, we can argue that its focus is to preserve the order of both the linguistic dimension (e.g., words and sentences) as well as the plot dimension (the order facts are introduced and ordered). Judging by the generated prompts provided as examples, it would seem they hold no regard for separating description from action, which apparently would be an argument against our hypothetical belief that this distinction is necessary. The lack of a formal narrative structure found in this approaches questions whether it is necessary to establish facts via description before articulating actions that refer to them. The authors, however, highlight the repetition as a problem arising from a short dependence distance. This could be interpreted as a symptom derived from the lack of a narrative high-level structure, a generalized design choice often found in similar modern approaches that rely on statistical techniques.

## Knowledge-Based Story Construction

[22] introduces a computational story generator in which human authors can define plot points that are then used for generation. According to the author, consistency and plot structure are achieved by incorporating the knowledge from ConceptNet, a repository for semantic relationships based on common sense. In this case, causality is fetched from an external source and incorporated into a directional graph that drives the story generation. The human author interacts with the system by defining characters and their emotions as well as a couple of initial events, which are then used to calculate the goals and their viability. If the system finds that the goals are not feasible according to its internal knowledge (based on ConceptNet), it requests modifications to the human input. If the input is acceptable, a chain of events with causal links is established through a planning strategy. While the approach is not innovative in its internal generative dynamics, it still relies on a basic world building (the initial events and characters) that are then used to develop the plot through subsequent events. Once again, we observe the distinction between descriptive world building and a sequential chain of character plot developments that establish causal links backwards up to the initial input.

## Evolutionary Story Computation

[55] presents a methodology meant for evolutionary computation that transforms a written story into an event-level and hierarchical-level grammar using a network representation. In this work, story formal representations are composed of events that depend on each other in a hierarchical configuration. The provided definition of dependence (events serving as the enabling conditions of subsequent events) suggests a causal relationship. The concern for actor characters, temporal order, and shared setting also seems to be central in determining the dependence relationships. The result of this methodology is a set of formal entities (a hierarchy of interdependent events with characters, time, space, a description a topic, and an object) that can be transformed into a chromosome (and vice versa) apt for evolutionary computation techniques that develop a story using subjective (coherence, novelty, interestingness, and quality) and objective (logical structure of events and participant arrangement) metrics to control the result. Overall, the formal representation of a story and the metrics used support our hypothetical belief that these systems rely on the notions of time and sequence to attribute causality. Regarding actions and descriptions, however, this methodology is meant to extract the story representations and then make it 'evolve' over several iterations. It makes no assumption on whether retrieved data will depict descriptions or actions, and therefore, it is conditioned by the source stories and their structure.

## Components and Relations for Analyzing Human-Written Texts

The analysis of human-invented texts requires the definition of an annotation framework that facilitates an independent, non-complex process in which the tagging of the stories is specific and reduces the inference of the annotators. Additionally, the annotation framework must be aligned with knowledge representation in existing computational narrative systems.

### Level of Granularity

It is possible to identify three levels of granularity with regard to the abstractions over a narrative: (a) the universe of all entities and events, (b) the subset of that universe that is structured by time and causality, and (c) the rendering of the narrative. In the literature, the naming convention varies between authors. To carry out the study, a specific level of granularity must be set. Since this research is focused on plot generation systems, final textual rendering will not be addressed. The analysis of the universe of all entities and events is not possible when only the rendered story is provided by the participants (which is the case).

Therefore, restricting the annotation process to a *subset of the universe that is structured in the form of events* (the top granularity-level plot), satisfies both annotation objectives. The simple annotation is made possible, because it is assumed that

the annotator will follow the sequential rendering and will align sentences and state transitions with plot events, and the alignment with existing automatic story generation systems should happen because, as summarized in “[Relevant insights from existing story generation systems](#)”, there are a large amount of computational systems that work at this level of granularity.

## Formal Components of Existing Story Generation Systems

Section “[Relevant insights from existing story generation systems](#)” has reviewed several story generation systems. The analysis was focused on the main structural components that the systems use to model and produce stories. With this information, it is possible to identify which characteristics are more frequently used, and design the annotation task, so that the tags provide coverage for the main features. The main conclusion extracted from the analysis, summarized in Table 1, is the recurring inclusion of an atomic narrative element depicting action or description. This is explained in detail in the following section..

## Story Components: Actions and Descriptions

Studying how plot elements are laid out in short plots implies the segmentation of the story into these plot elements. The process is not straightforward, because it requires setting a specific level of granularity (see 3.1). Fixing a level of abstraction constrains the generality of the results, but the annotation and the analysis require a certain abstraction level for it to be realizable.

In this research, the level of abstraction will try to match the one existing in the story generation systems reviewed in “[Relevant insights from existing story generation systems](#)” and summarized in Table 1. Each element of a story at this level of abstraction will be referred to as *story component* and the details will be explained next.

The level of granularity of story components must fulfil three requirements for them to facilitate the annotation and be applicable to computational narrative: they have to convey narrative semantics, they have to be explicit, and they have to be indivisible at that certain granularity level. From the point of view of an annotation, not having explicit information or being divisible would transform the process into a task of abstraction and understanding, which would lead to inconclusive results. Along the same lines, identifying the atomic components allows us to focus on the structural features of the story, and not the features of the story components.

All of the reviewed story generation systems consider some form of atomic element that represents either an event (a change in the state of the narrative, a plot event) or a description. The difference between actions and descriptions has important structural implications. Descriptions in a narrative do not make explicit the time span of this property. For example, an everyday story about a baker could describe him as a grumpy man but not provide a temporal reference about it, i.e., the text does not explain *when* did the baker start being grumpy or why. This can be different

**Table 1** Formal features of narrative representation and processing for the story generation systems from “Relevant insights from existing story generation systems”

System	Main characteristics
MINSTREL	Initial <i>description</i> scenes, sequence of character <i>action</i> scenes preserve character plan, story world and emotional consistency.
Universe	Complex descriptive character data structures <i>action</i> that trigger plot developments consistently with the initial characters and goals.
Virtual storyteller	Director AI exercises environmental, motivational and proscriptive control over character belief, desire and <i>action</i> model. Resulting character action plot is consistent and avoids prescriptive control.
Fabulist	Intent-driven partial-order causal link, planning algorithm produces coherent narratives and characters. Features <i>descriptive</i> fact establishing and asserting, character beliefs, desires and <i>actions</i> . <i>Descriptions</i> precede <i>actions</i> that refer to them.
MEXICA	Generator relies on two cyclic states; engagement (story- <i>actions</i> are used to seed the story) and reflection (additional post-conditions are added). Reflection stage boosts the coherence by reinforcing factual links.
BRUTUS	Uses thematic frame sequences of <i>descriptive</i> events, characters and goals. Proactive and reactive character behavior including conditions and consequences for causally chained <i>action</i> events.
Author	Develops story events ( <i>actions</i> ) from representation of the author’s mind ( <i>descriptions</i> ) keeping both data dimensions linked.
Neural generation	Relies on knowledge extracted from large datasets. Uses set of premises to hierarchically derive stories. No specific distinction between <i>descriptions</i> or <i>actions</i> .
Knowledge-based construction	Human author inputs characters and initial scenes ( <i>descriptions</i> ). Events are created transforming the story world ( <i>actions</i> ) using common sense data extracted from datasets for consistency.
Evolutionary computation	Transforms written stories into formal representation for encoding and decoding as chromosomes. Original text <i>actions/descriptions</i> are integrated later to boost coherence.

in more complex narratives, but for simple plots, the property holds and does not represent any change in state (which is the main differential feature of actions).

Causality also works differently between actions and descriptions in narrative. Descriptive components cause events in a narrative at a very global level. In a description, the grumpy baker can have a bad reaction at different points in the narrative, not necessarily after he has been described as grumpy, and no cause for the baker’s grumpiness needs to be provided. In a story action, the baker will probably complain right after some unfortunate event has taken place, thus fully linking it with a causal pre-condition. That is, actions, as defined in the level of granularity, need specific causality for them to make sense.

To provide coverage to these particularities, two tags for story components have been established: *story actions*, which refer to story components that represent a change within the story, and *story descriptions*, which represent definitions of the elements of the story (places, characters, etc.).

Finally, it is important to note that the analyzed texts do not only contain elements that fully match story descriptions and story actions: discourses in the form of text also contain stylistic figures like “once upon a time” or “they lived happily ever after”. These forms are considered communicative conventions not being really specific to the story. Therefore, the study of these figures has not been addressed in this experiment. The first step in the annotation was to identify these elements and omit them in the study.

## Relations Between Story Components

Identifying and categorizing the presence of story actions and story descriptions in the source stories (as described in “[Story components: actions and descriptions](#)”) are the first step to understand the basic features of stories, but the narrative structure of the story components, in the form of relations between them, can provide a much richer insight for quantifying the layout of plot elements.

Defining the relations between different story components is also dependent on a certain level of granularity. Since rendered text is not considered in this research, the relations must work at a purely structural and semantic level, in the same way as the segmentation in story descriptions and story actions.

The analysis provided in [Relevant insights from existing story generation systems](#)” reveals that, implicitly or explicitly, all reviewed story generation systems use a form of *causality* as an important component for connecting the plot events. Another relevant relation is *time*. Temporal ordering of events in a plot is necessary, and it is (as described in “[Story components: actions and descriptions](#)”) one of the main differentiators between actions and descriptions: actions are subject to time (occurrence instant and duration), while descriptions, as defined in this experiment, are not. Since causality and time were used to identify actions and descriptions, they have also been annotated for further analysis.

It was previously introduced that modern narratology and cognitive science claim that narratives are not only a common form of communication or artistic expression, and it is in fact a fundamental way of structuring several knowledge structures in human cognition [18, 42, 47]. It has been previously argued that this is due to a more general adaptation of human cognition to a physical environment [24, 25]. According to this idea, it would be possible to annotate the stories with several relations beyond time and causality, like agency, location, and other cognitive aspects.

Annotating agency was considered important for the analysis, since characters are fundamental in a story plot, and all computational systems address characters. In order for the annotation to be simple and generally applicable, story descriptions are tagged with the character receiving the description (“*John* was tall”), and story actions were annotated with the active and the passive characters (“*John* kissed *Mary*”). While the set of possible roles in a story component can be more elaborated than agent and object, this simple method is expected to cover a significant number of cases for a first experimental study.

Other relations are not considered in this study. Location, for instance, is many times omitted in short plots, so it is possible that a number of subjects do not include

location. Additional relation types like aggregation (actions that describe a higher-level episode), emotional impact, or others are not considered in the experiment. They have been left out, because either they might not happen in a relatively important number of the stories written by the subjects or they are not part of the common features of computational narrative systems.

In summary, *causality*, *time*, and simple *character information* have been selected as the relations linking story components in the annotation.

## Experiment

To observe the amounts, order, and relations of the fundamental formalizable components of a short plot, an experiment in which human participants wrote a short story was conducted. Our aim was to obtain simple, invented narratives from human subjects and then use annotators to observe and extract common story constituents and dependencies. The study did not address complex literary phenomena. As such, the participants were not required to have specific narrative skills. The stories were annotated according to the components and relations described in “[Components and relations for analysing human-written texts](#)”.

As described in the next sections, the objective was to create a scenario in which participants that are not necessarily experts had to invent and write original stories. The subsequent annotation was carried out to reveal the most important features of this structural representation.

## 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 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.

## Participants

The experiment was announced and those wanting to take part in it voluntarily enrolled, counting finally twenty-six students ( $N = 26$ ), 20 males (76.92%) and 6 females (23.08%), from Complutense University of Madrid (Spain), with ages ranging from 19 to 28 years ( $mean = 21.58$ ,  $stdev = 2.44$ ). All participants were native Spanish speakers. There was no compensation for participating in the experiment. They were informed about the process and the data acquisition and analysis, and they were given a written document clarifying their informed consent, which all of them accepted and signed. The experiment was conducted in three sessions.

## Experimental Guidelines

Before each experimental session, each participant was assigned a computer in the room. The participants were told not to manipulate the computers, to turn off their cell phones, and not to talk to each other during the experiment. Each participant was placed in a numbered computer with the screen switched off.

After being briefed, they were instructed to turn their corresponding screens on. Each computer was prepared, so that the only thing they would see was a Google Document with instructions about the task to be carried out. The Google Document was named after the computer number to facilitate the data collection and participant identification. The participants were told not to make any edition to the document and read the instructions carefully, which all of them did. The list of instructions that they received was the following:

- **Read the instructions carefully** before starting the task.
- **Do not start** until you are told to.
- Before starting, **ask any question you might have**. You cannot ask questions once the task has started.
- You have to **write a story** in this task.
- There is **no restriction** about the theme, characters or actions.
- Do not worry about the format, typeface or presentation: **the text is the only thing that matters**. The only important structure are sentences and paragraphs.
- You can use up to **one page** at most, starting in the next page.
- The task lasts exactly **20 min**. The remaining time will be visible in the screen.

After 3 min, all participants were permitted to ask questions about the task. After the questions were answered, the participants were told to scroll down the Google Document, where they would find a blank page where the task was to be completed. They were given 20 min. The time that remained at each point was always visible by being projected in a large projection screen by the teacher's desk.

No particular narrative skills were needed to participate in the experiment. However, 4 simple questions were added to the questionnaire to identify possible outliers (subjects with a very low self-identified skill). Therefore, after the task, the participants were told to switch to another tab in their computers that contained a Google Form with demographic data. They were asked about their narrative skills (5 levels, low-to-high), their skills when using a computer (5 levels, low-to-high), and their skills using a text processor (5 levels, low-to-high). Additionally, they were asked whether they felt they had been given full freedom to create the stories (yes or no). These questions were not intended to measure specific skills, and the corresponding analysis was just focused on the identification of non-valid data. As detailed later on, all subjects were identified as valid for the study.

## Annotation Process

Before annotating each written story, orthography and grammar were checked and corrected. The orthographic correction was carried out with the free command line tool

aspell, version 0.60.6.1 with the Spanish dictionary es. Grammar was checked and corrected with the free command line tool LanguageTool, version 4.1, with the Spanish dictionary es. Only obvious syntactic and grammar errors (mostly due to typing mistakes) were fixed.

Verbs are the most elementary, explicit information in a written discourse. They convey most of the narrative meaning of a sentence, and clearly define actions or definitions in a story. The annotation was therefore focused on verb sentences, since these maximize the conditions for story components. Each verb sentence was matched with the corresponding participant, paragraph, and sentence number.

For dialogues, the annotation assumed that each utterance is an action (*to speak*). This helped to simplify the analysis, avoiding potential confusion between the layers of narration and dialog. Therefore, every time a character participated in a dialog in direct style (e.g., “—This is a thing—, I explained”), we established to strictly annotate the character action and not the content of the communication between the participants.

The annotation process was carried out by 2 independent annotators. The annotators would go through each sentence sequentially to fill in the following fields:

- *Component type*: whether the annotated item is a *story action* or a *story description* (as described in “[Components and relations for analysing human-written texts](#)”).
- *Protagonist is the agent of the component*: a Boolean field determining whether the protagonist is the agent of the component.
- *Protagonist is the object of the component*: a Boolean field determining whether the protagonist is the object of the component.
- *Previous story action*: the identifier of the previous story action, according to the temporal ordering of the story. That is, the identifier of the story action that happens right before the current action. Only applicable to story actions.
- *Causal dependencies (multiple)*: the list of story actions and story description identifiers that cause the current item.

## Results

In total, 25 out of 26 participants (96.15%) declared to have completed the task with full freedom, and one of them declared not to have felt that freedom. All participants declared that they would not disclose any details of the experiment until all the sessions were over.

Regarding demographic data, we asked the users to provide information about their skills in a 5-point Likert scale. 53.8% of the participants declared to be highly skilled when using a computer, 38.5% declared themselves being skilled (not very highly), and 7.7% had average skills. No participant declared to have little or no knowledge about how to use a computer. Therefore, the stories from all participants were considered valid for the study.

In terms of word processor usage, all participants had previous experience with Google Document. 19.2% described their word processor expertise as professional, 50% could use advanced features, and 30.8% of them only had an average



proficiency. No participant declared to have problems with the use of the word processor.

Participants were also asked how much they knew about narrative. Half of them (50%) claimed to have average knowledge, 26.9% higher-than-average level (but not the highest level), and 15.4% lower-than-average (but not the lowest) level. One participant (3.8%) chose the highest-level option and another one chose the lowest value (3.8%). Distributed in a 1 – 5 point Likert scale, the average was 3.11 ( $sd = 0.86$ ). This suggests a normal distribution of the participants perception of their own knowledge about narrative.

It is concluded that all participants had at least the minimum required skill to use the platform. Analogously, they declared to have at least the minimum required knowledge of narrative to complete the task in our experimental conditions. It is therefore possible to rule out technical difficulties when writing the stories.

## Annotation Results

In total, 741 sentences (per participant,  $avg=28.50$ ,  $sd=13.52$ ) were annotated<sup>1</sup>. The agreement regarding the annotation of story components as *story actions* or *story descriptions* was measured by computing the number of story components that were annotated under the same category. The results were acceptable: 79.55% ( $\kappa = 0.596$ ,  $p = 0.000$ ). The agreement regarding the role of the protagonist in the story was considered acceptable: 82.69% ( $\kappa = 0.643$ ,  $p = 0.000$ ). Among all the annotation results, the causal references are some of the less agreed upon. A multidimensional vectorial analysis revealed a moderate agreement of 49% when it comes to establishing causal dependencies between sentences in the same story. While certain linguistic dependencies such as the grammatical structure are relatively agreed upon and even automatized using NLP approaches, causality dependency is far from a consensus. Additionally, dependencies between sentences in a text are also subject of study, but we still lack a standardized inter-sentence dependency parsing process. Given these circumstances, we consider a moderate agreement an acceptable result regarding the annotators' agreement on these matters.

Table 2 shows a partial brief example of the annotation process and how both annotators were only partially in agreement. It can be seen how sentence 1 is tagged both as a story action and as a story description, because it includes both. Each annotator has focused on a different part of the sentence. In sentence 4, the annotation detail differs between annotators, probably because annotator 2 assumed that the causality was not as clear. This exemplifies the two more common divergences, which are assumed to happen because of the high complexity or language and semantics and the limitations of the tagging method.

Story descriptions were slightly more frequent than story actions. 53.28% of the annotated story components were story descriptions, while 46.72% were story

<sup>1</sup> The annotated human-invented story sentences and their resulting values can be found at <https://github.com/NILGroup/annotatedShortStoryFeatures>.

actions. Story descriptions were located on average at 43.93% ( $sd = 11.46$ ) of the relative narrative sequence. Story actions tended to happen later on in the story: their positions average at 59.81% ( $sd = 13.55$ ) of the relative narrative sequence. Differences were significant ( $t = -3.897, p < 0.001$ ). This suggests that descriptions were more frequently used and happened earlier in the story. Besides, the dispersion of the relative spatial location of the story descriptions is lower. Figure 1 offers a graphical depiction of these data as two boxes. The position of the component in the story is relative to the story time (0% – 100%). The story actions are represented as circles and the story descriptions are represented as triangles. The shown points represent the average position for each story.

The number of verb sentences per story written by the participants amount ranged from 13 to 61 ( $avg = 29.64, sd = 13.52$ ). There were no significant differences between individuals declaring different narrative skills ( $Z = -0.064, p = 0.949$ ), computer skills ( $Z = 1.045, p = 0.296$ ), or text processor use skill level ( $Z = -0.015, p = 0.988$ ). Likewise, other factors such as the amount of previous story actions, causal dependencies, or the attribution of protagonist agent and/or object do not seem to be influenced by these demographic factors.

Most story components are at some point referenced by other story components: 85.42% of the story components have a causal link to other elements. No significant difference was found between story actions and story descriptions in terms of the amount or distance to references ( $\chi^2 = 0.137, p = 0.712$ ). Along the same lines, 84.82% of story components reference other story components. Again, this aspect did not yield any statistically significant difference ( $\chi^2 = 0.055, p = 0.815$ ) in terms of the kind of story component (action or description).

The layout of story descriptions is also worth analysis. The acquired data reveal that 88.04% of descriptive components link to previous descriptive components, thus producing a description in which the information is delivered sequentially. However, only 24.64% of descriptive components are referenced. This seems to suggest that a quarter of the descriptive block influences most of the subsequent descriptions (and actions). This could indicate a lesser direct impact of approximately 75% of descriptive components, although the analysis is not strong enough to conclude anything, especially given the complexity of semantics.

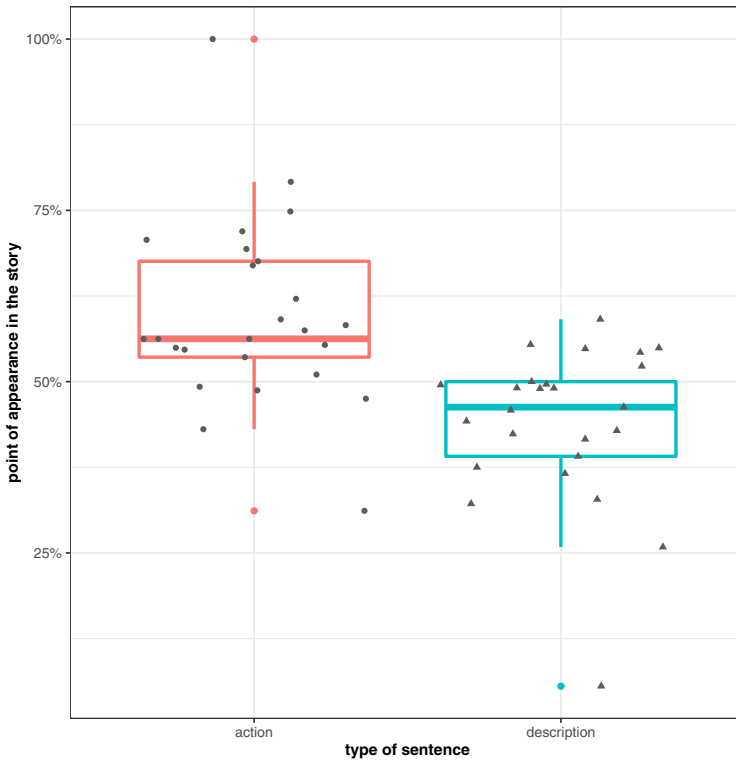
A similar but less intense pattern can be observed with story actions. 81.29% of them make causal or temporal reference to previous actions, and 43.39% of them are referenced (50.95% of story descriptions and 61.24% story actions). In the case of story actions, these amounts seem less relevant, because, semantically, referencing a previous action implies relying on the narrative state of all the actions that took place before, so the low amount of directly referenced actions can just be a rendering of a wider causal graph structuring the plot. Again, the collected data do not permit to conclude anything, but seem to point towards the need of a more qualitative study.

Most of the annotated causal references (79.23%) point to the immediate previous component, followed by components within a distance range of 2 – 6, above 1% of the times. Table 3 shows these data. A histogram of the reference frequencies is provided in Fig. 2.

The distribution of story descriptions and story actions is shown in Fig. 3. This figure shows a clear downward tendency of the frequency of story descriptions, and

**Table 2** Sample sentences and their corresponding annotation results. It can be seen that both annotations are correct, but given the complexity of linguistic aspects, the tags could differ, totally or partially, between annotators

SENTENCE	Annotator 1	Annotator 2
1. Holding the cup over the bar, I was waiting	Story description	Story action
2. The smoke loaded the ambience, it was repulsive	Story description, depends on 1	Story description, depends on 1
3. On the other side of the establishment's door, that person was about to enter	Story description, depends on 1	Story description
4. The security guard, after nodding and setting the earphone in his right ear, opened the door to that paradise, an orgy of sound and alcohol	Story action, follows 3, depends on 3	Story action
5. Walking on the red carpet towards the dance floor	Story action, depends on 4	Story action



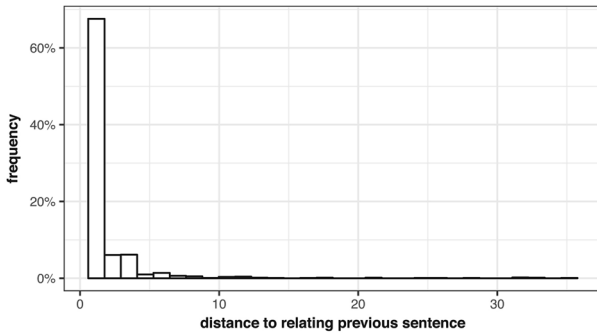
**Fig. 1** Distribution of story actions (circles) and story descriptions (triangles) along the relative story time. Story descriptions are slightly more frequent than story actions. On average, story descriptions appear before story actions

a clear upward tendency of the frequency of story actions. Additionally, it can be seen how there is a strong average increment of the proportion of story actions at around the middle of the story.

Regarding temporal dependencies, only 6.88% of the story components were annotated as having temporal dependencies with other elements, with no distinction between story action or description ( $\chi^2 = 1.247$ ,  $p = 0.264$ ). Only 3.91% of the story components have explicit temporal dependencies, with a significant predominance for story actions (65.47%,  $\chi^2 = 18.71$ ,  $p < 0.000$ ). As with the causal dependencies, most temporal dependencies link consecutive story elements (75.86%) while other temporally dependencies might link story elements separated by two (13.79%), three (5.17%), four (3.44%), or, for instance, nine (1.72%), as illustrated in Table 4.

**Table 3** Frequency of relative distance between a story component and the story component it references (both actions and descriptions). Most story components are causally related to the previous element

Causal distance	Frequency
1	79.23%
2	7.11%
3	4.33%
4	2.86%
5	1.11%
6	1.64%
7	0.73%
8	0.57%
9	0.08%
10	0.32%
[11–34]	< 0.17%
35	0.08%

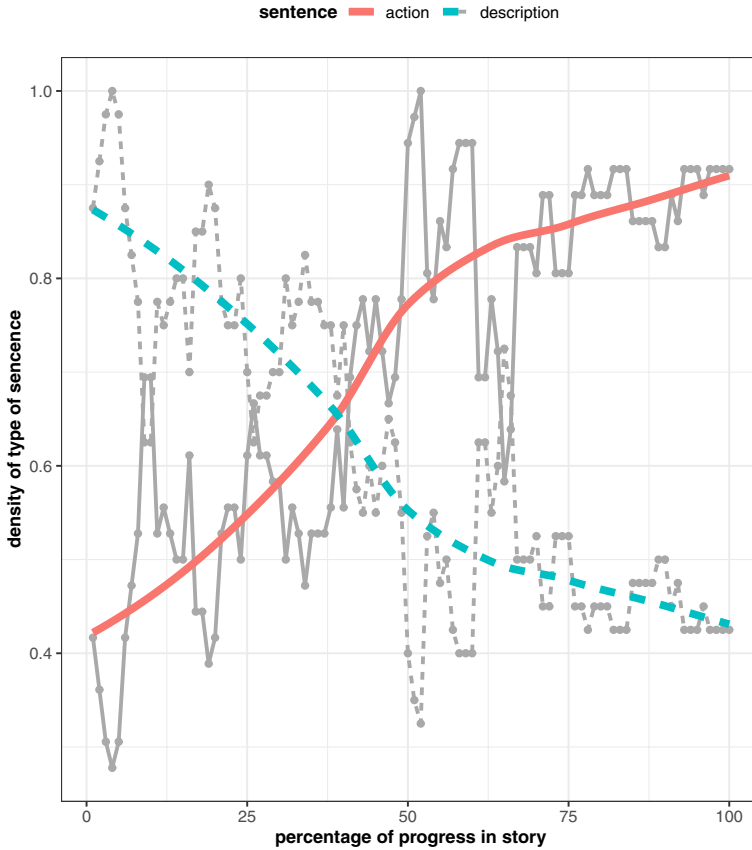


**Fig. 2** Histogram of the relative distance between a story component and the story component it references (both actions and descriptions)

### Analysis of Results

The analysis of the material generated by the annotation of the stories produced in the experiment described in “[Results](#)” has permitted to identify a number of characteristics that can be observed across the set of instances of stories produced by the participants. These characteristics are:

- Most of the descriptions are provided at the beginning, and most of the actions, at the end.
- There is a sudden increase in the proportion of actions around the middle of the story.
- A relatively low proportion of descriptions have a strong impact in terms of causality.



**Fig. 3** Average proportion of story descriptions (dotted line) vs. average proportion of story actions (straight line)

**Table 4** Story component temporal dependence distance

Temporal distance	Frequency
1	75.86%
2	13.79%
3	5.17%
4	3.44%
9	1.72%

- There is a strong relation between story components. Components in a story, except for the concluding ones, tend to be referenced by other story components.
- Both casual and temporal references to earlier components from later components tend to involve reference to the immediately preceding component in the

sequence. The distance between relations is very short, and the causal links are mostly laid out incrementally, producing a chain-like structure.

In this way, the observed results seem to align with the existing narratological descriptions of story, but the results also constraint the diversity.

The data show a very frequent pattern of an initial descriptive setting followed by a sequential chain of atomic actions. While a description followed by the set of actions was intuitively expected, it is possible to observe that short narratives describe almost all the content in the beginning and all the decisions seem to be fixed. The actions coming out from the initial setting do not modify the original decisions.

From this fixed descriptive block, the actions occur as a chain of state modifications. In this chain, actions are laid out in the chronological order of their starting time. Complex discourse features like flashbacks or flash forwards were not significantly observed, which suggests that humans tend to represent simple plots linearly.

The increase of the proportion of story actions around the middle of the story probably describes a significant point where the descriptive block at the beginning and the previous actions have provided all the required causality for a climax. This seems to be rendered as a burst of actions in which the proportion of actions is around 90%. This result could be the numerical parameter matching the climax in classic narratology models like Freytag's pyramid [11], for instance.

Causality was observed to provide a general coverage on the whole plot, and causality links between plot components happen at a relatively short distance in the discourse. This suggests a *strong correlation between causality and time* order, yielding plots in which, in most cases, the current state causes the next. In terms of agency, the protagonist clearly plays a major role, accounting for a very high percentage of the story components.

Regarding the role of causality, the results provide empirical evidence that the causal links tend to follow a relatively narrow structural pattern. This, in many cases, suggests that humans construct simple plots by representing causality as a chain of consequences. This is in consonance to findings in cognitive science about the way in which humans understand narratives [16, 52], but the literature suggests wider causal trees than those found in the annotated stories. That is, the causality for short plots tends to look like a sequence of state plus action.

The results also evidence that narratives are constructed as sequential chains of actions *along time*. This is in line with the findings about causality; that is, generally actions only cause subsequent actions. In this line, it is clear that this level of story representation, which is assumed to be fundamental in humans, follows a simple physical layout. This conclusion is well aligned with the narrative hypothesis (i.e., that narrative is a fundamental way of structuring part of our knowledge) and also is related to the physical rules of our environment.

The subjects were not expert writers. The written stories were expected to fit into the category of everyday creativity, akin to *little-c* creativity [21]. While *Big-c* creativity probably cannot be captured by these means, it is possible to argue that current story generation systems are still not able to produce top-quality narratives, so this kind of process is assumed to provide a useful contribution. In any case, the

proposed methodology would be applicable in a study in which trained story writers create the stories to potentially achieve *Big-c* creativity. This would help to gain insight on how more-developed stories are structured. In particular, combining stories produced by trained writers with evaluation from non-skilled participants would provide a perspective very useful for some story generation systems.

It is relevant to note that the experiment was designed, so that a detailed, sentence-by-sentence analysis could be applied. This perspective increases the required effort per story. From an experimental design perspective, this is a necessary trade-off that permits a detailed, qualitative output. While the process implies a level of subjectivity, it allows to provide valuable data on the story invention process from the human perspective according to the authors of the stories themselves along with the human annotators. By relying on the statistical analysis of human-invented, separated, and annotated short stories, we provide an perspective on story invention distinct to those based on formalist, cognitive, or statistical models that provides a horizontal approach across these disciplines.

## Testing Resulting Insights over Additional Sources

The observations outlined in “[Analysis of Results](#)” provide interesting insights on the corpus compiled for in the experiment. This section explores the applicability of the methods of analysis developed for the experiment to stories from sources that are radically different in many ways. The idea is to test the applicability of the methods in question not just as tools to extract meaning from the particular corpus over which they were designed, but also as tools to understand other corpora, and how they might differ from the one used in the experiment and from one another. We will also explore the insight that may be derived concerning the operation of a story generation system by applying the proposed measures to the stories it generates.

To achieve these goals, two different corpora are considered. Section “[The ROC Story Corpus](#)” applies the explained methods to the ROCStories Corpora [29] used for machine learning tasks on stories such as the Story Cloze Test [30]. Section “[Outputs of Automated Story Generators](#)” applies the same methods to a compilation of outputs from automated story generators of those reviewed in “[Relevant insights from existing story generation systems](#)”.

### The ROC Story Corpus

The ROCStories Corpora [29] are composed of stories with 5 sentences. The corpus is very structured and built with the task of evaluating story understanding. The corpus was built by asking hundreds of workers on Amazon Mechanical Turk (AMT) to write novel five-sentence stories. As reported by the authors of the corpus, volunteers were asked not to include “anything irrelevant to the story”, and the prompts used for the process were empirically refined in an iterative fashion to reduce “the number of submissions which did not have our desired level of coherency or were specifically fictional or offensive” [29].



**Table 5** Story components for the ROC corpus

Sentence number	Story actions	Story descriptions
1	15	11
2	19	7
3	20	6
4	21	5
5	18	8

As an example, 26 stories from the ROC corpus were randomly selected and annotated with the guidelines provided in “[Experimental guidelines](#)”. The results in terms of the aspects considered in this paper are summarized in Table 5.

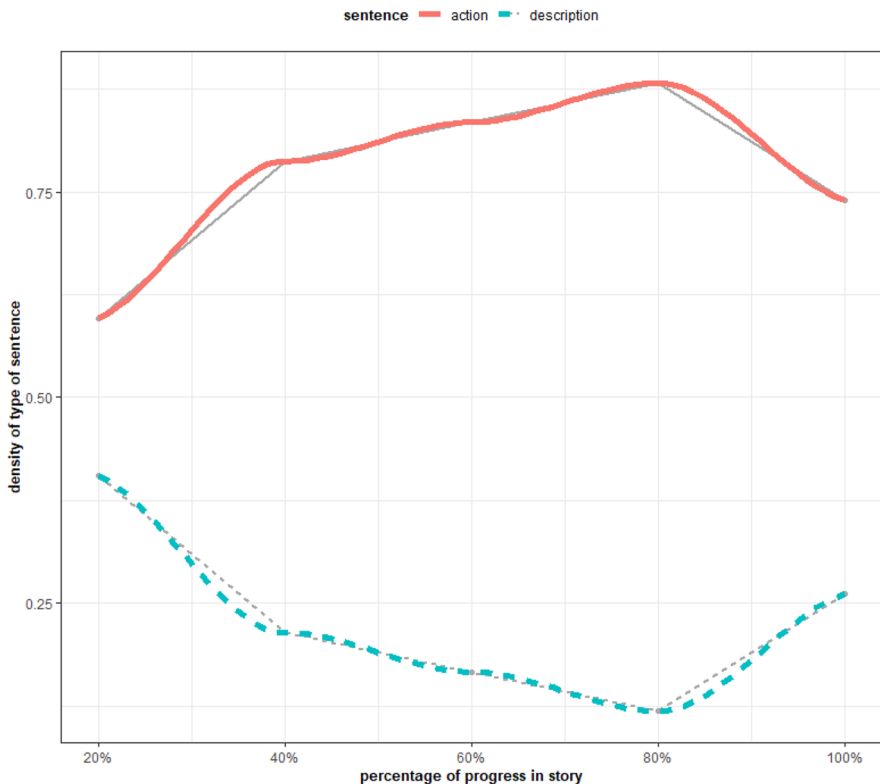
The relative average proportion of actions and descriptions for the ROC corpus is shown in Fig. 4. This graph shows the effect of the peculiarities of the ROC corpus in several ways. For a start, the fact that all the stories have the same length of 5 sentences makes for a much smoother curve in all the plotted magnitudes. The restrictions that the story have only 5 sentences and that all of them count towards the end result leads to a very high density of action (and consequently very low density of description). The overall shape of the curve is consistent with a very simple narrative, starting with a description of an initial state, a number of actions that change the state, and a brief description of the final stage.

The sample from the ROC corpus shows that 72% of the annotated sentences correspond to story actions, and 28% correspond to story descriptions, in contrast with 47% of story actions and 53% of story descriptions in the corpus which we created. These differences can be explained by the relatively strict structure that the stories in ROC present.

In terms of relative distance of causal links between elements in a story, the data obtained for the ROC corpus are shown in Fig. 5. Again, the nature of the ROC corpus significantly affects the resulting graph. In the most immediate practical terms, because the size of causal links between elements in the same story has a glass ceiling of 4, corresponding to the case when the last sentence of the story is causally related to the first one. The fact that there is a considerable percentage of causal links that span beyond the immediate sentence (around 35% of the total) and a non-trivial percentage (around 5%) that span from the last to the first sentence of a story is undoubtedly related to the particular constraints imposed on the writing process (no material unrelated to the story to be included) and the conditions for which the writing prompt was optimized (high level of coherence).

Furthermore, the analysis of causality links in the stories in the ROC corpus uncovers information that is relevant to the suitability of the corpus itself as an example of naturally occurring stories, but which may also prove valuable to the task for which the corpus is designed (evaluating machine learning approaches to story understanding).

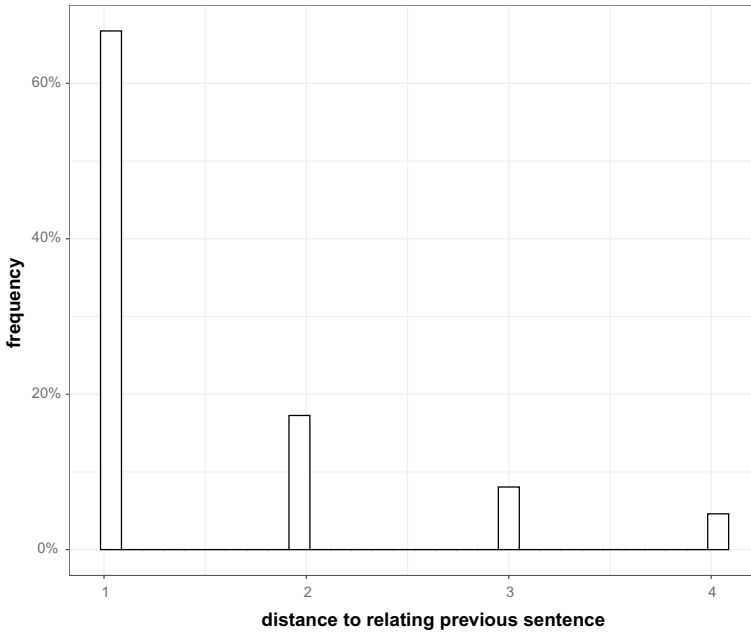
In terms of the validity of the corpus as an example of naturally occurring stories, it seems clear from this analysis that there are significant differences between



**Fig. 4** Average proportion of story descriptions (dotted line) vs. average proportion of story actions (straight line) in the analyzed stories of the ROC corpus. The shape of the curves is different from the average of the human-invented stories (shown in Fig. 3)

the stories compiled for the ROC corpus and stories compiled with less elaborate prompts and less constraints on the size and desired content, with human volunteers producing the stories in both cases. Humans with no constraints on content produce story with much higher proportions of descriptions. This has much to say about the versatility of human subjects when asked to produce stories, and how easy it is for seemingly innocent constraints to drive outcomes away from the characteristics of naturally occurring samples. This can be true of the constraints on size and content for the ROC corpus, but also on the constraints on time taken to produce the story as applied to the stories in our corpus. This point is taken up later in this section when addressing the comparison with automatically generated stories.

The machine learning approach to story understanding which is tested on the ROC corpus is the Story Cloze Test, where a system is given a four-sentence ‘context’ and two alternative endings to the story, and the system’s task is to choose the most appropriate ending. In the context of this task description, it seems clear that both the amount and the length of any causal links between the ending and the



**Fig. 5** Histogram of the relative distance between a story component and the story component it references (both actions and descriptions) in the analyzed stories from the ROC corpus. It can be compared with the human-invented output in Fig. 2

preceding sentences of the story may play a significant role in making a particular problem easy or difficult to solve.<sup>2</sup>

### Outputs of Automated Story Generators

A driving goal for the present paper is to obtain insights that might help improve automated story generators. To make this possible, it is important to consider what the analytical approach being proposed might reveal when applied to the outcomes of existing story generators. This task is faced with several obstacles arising both from the differences in nature across existing story generators and from the fashion in which past work on story generators was reported.

From the point of view of the nature of the story generators, the set of outputs that could be considered is restricted by the fact that many of the existing generators actually produce outputs that are not stories in the sense considered

<sup>2</sup> Cases where the ending is causally linked to all the preceding sentences, for instance, should be easier to solve than the case where the ending is only causally related to the immediately preceding sentence. From a slightly different point of view, the task will also vary in difficulty depending on the degree of causal relation between the alternative ending provided as distractor and the preceding sentences in the story.

in this paper. Systems such as VirtualStoryTeller [49] or author [7] actually produce interactive stories, which only make sense when an interaction by a user is meshed with system contributions. In these cases, the proposed analysis cannot be applied to their outputs. The universe system [23] produced conceptual outlines for soap opera episodes that do not correspond to stories in the sense considered here either.

From the point of view of the fashion of reporting, it is rare for reports on story generators to list large numbers of examples of system output. Where more than one example story is given, additional instances tend to consist of examples of low-quality outputs resulting from ablation of part of the system functionality (as is the case for the reports on MINSTREL [53], Fabulist [40] and, to a certain extent only, MEXICA [32]).

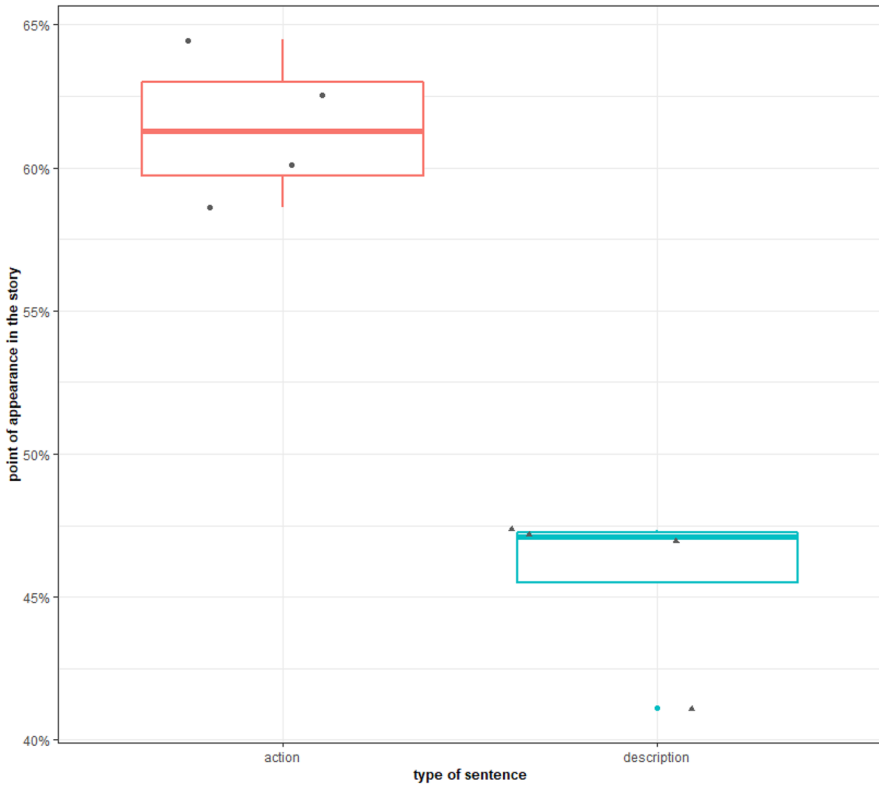
These issues make it difficult to compile a corpus of stories by automated story generators that is significant in number and balanced across different story generators. Nevertheless, the authors considered it important to include some analysis of stories of this nature. To this end, four stories generated by different automatic generation systems were analyzed and compared to the human-invented stories according to the proposed set of metrics. The automatically generated stories were selected by examining the published examples of stories produced with the story generation systems studied in “[Relevant insights from existing story generation systems](#)”. In total, only 4 systems provide full stories: MINSTREL, MEXICA, BRUTUS, and Fabulist. Among the provided examples, the longest story of each system was selected for analysis. The analysis was carried out by the same 2 annotators and according to the guidelines explained in “[Experimental guidelines](#)”.

The relative distribution of actions and descriptions over the length of the stories is shown in Fig. 6. Story actions tend to appear later in the story ( $mean = 63.22\%$ ,  $stdev = 3.15$ ), and story descriptions concentrate before the middle ( $mean = 45.65\%$ ,  $stdev = 3.04$ ). These aggregated values are slightly higher than in the human-invented stories.

Table 6 summarizes the identification of story actions and story descriptions. It is important to consider that, although the four stories under consideration have been grouped together under a single category of outputs of automated story generators, their analyses show a significant variation across them, and each one of them is represented by a single instance, so the results are provided as a source of qualitative insights rather than statistical significance. More detailed analysis of the peculiarities of each automated story generator in terms of the characteristics of its output might be fruitful, but it would require a larger sample of stories from each one, and is therefore beyond the scope of the present paper.

The graph on relative density of actions and descriptions across the length of the story for the case of the sample stories from automated story generators is shown in Fig. 7. The shape of the curves provides some interesting insights on the stories.

The first interesting feature is that the overall shape seems to match the structure already postulated for the stories of the ROC corpus: high density of descriptions at the beginning and end of the story, with actions taking over through the middle part of the story. This is consistent with a description of an initial state, narration of actions that change the state, and description of a final state. In this

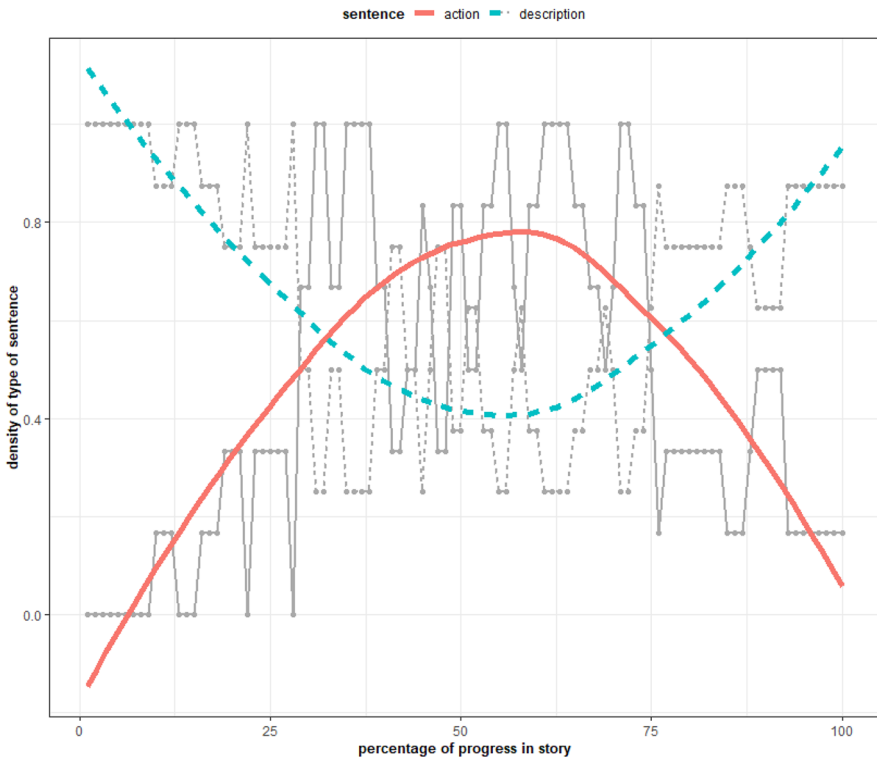


**Fig. 6** Distribution of story actions (circles) and story descriptions (triangles) along the relative story time for the analyzed story generation systems. Figure 1 shows the results for human-invented stories

**Table 6** Story components in the analyzed story generation systems

Story	Story actions	Story descriptions
Jafar and Jasmine (Fabulist)	25 (31.25%)	55 (68.75%)
Simple Betrayal (BRUTUS)	21 (47.77%)	23 (52.27%)
The Lovers (MEXICA)	24 (40.58%)	35 (59.32%)
The Vengeful Princess (MINSTREL)	15 (31.25%)	33 (68.75%)
Total	85 (36.79%)	146 (63.21%)

case, the graph shows a more balanced distribution between actions and descriptions, with the curves for one and the other actually crossing one another towards the middle of the graph. This is consistent with a set of stories not restricted by artificial constraints on size or content, which leaves freedom for the systems to inject descriptive material where the nature of the story requires it. In this



**Fig. 7** Average proportion of story descriptions (dotted line) vs. average proportion of story actions (straight line) in the analyzed stories produced by automatic story generation systems. The shape of the curves is different from the average of the human-invented stories (shown in Fig. 3)

respect, the stories by automated generators resemble more the stories produced by unconstrained human volunteers than the stories in the ROC corpus.

However, there is a second interesting feature of the graph that differs from unconstrained human-generated output. The graph for unconstrained human-generated stories showed a steady shift in density from a focus on description at the beginning to a focus on action towards the end (see Fig. 3). This contrasts heavily with the shape observed here, which shows a hump of the curve for actions around the middle of the story and a return to description towards the end. There are two possible explanations for this difference. The first one arises from consideration of the knowledge-based nature of the automated story generators in question. These systems rely on knowledge encoded in their programming to represent connections between actions and the states that result from them, which allows them to establish the state of the world after particular actions. These consequences of actions are represented explicitly, and valuable to the system in that they allow checking the consistency of the results. As such, they tend to be included in system outputs. This leads to examples of output of the type “X kills the dragon. The dragon is dead.”. This contrasts heavily with the way in which humans tell their stories, which systematically relies on the reader being able to

infer consequences of actions rather than mentioning them explicitly. This might explain the observed differences. The second possible explanation considers the observations above that the circumstances of the experiment may have affected the nature of the outcomes, for instance, by having volunteers finish off their stories with less descriptive detail than they had been including at the beginning of their story if they see that the time available to them is running out. Further elucidation of this matter would require a larger set of examples from each story generator, separate analysis of the behaviour of each story generator, and consideration of different conditions for human-produced stories.

Another interesting difference between human-invented short stories and those generated automatically is the reference distance. The graph of length of causal links between elements in the story for outputs of automated story generators is shown in Fig. 8. Because the four stories that were analyzed show marked differences in behaviour with respect to this feature, the graph shows the histogram for the average values, but superimposes on to it in colour the curves for the specific values for each of the different stories.

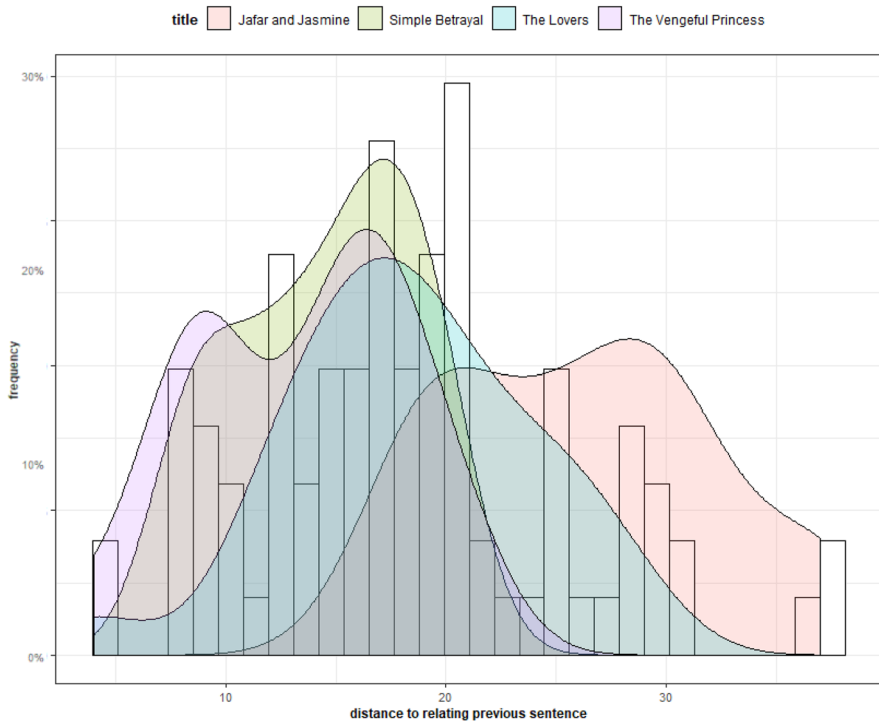
The main observed difference between the histogram for the automatically generated stories and the one for the stories in our corpus (and in the ROC corpus) is the fact that the histogram for the average values (of distance of causal links) for the automatically generated stories shows a succession of peaks spaced over the range of sizes considered. However, the coloured curves show that this pattern of peaks emerges in the process of averaging from the interaction between the distinct patterns of the different stories. In the hope of avoiding the obfuscating effect of averaging across samples that are not necessarily related, the pattern for each story is analyzed separately in search of insights.

The patterns observed for the four stories show two different trends.

Stories by MEXICA, BRUTUS, and MINSTREL show a rough bell-shape curve with a peak around a distance of 15 positions between causally related elements (around a third of the total length of the stories), and a smaller peak around a distance of 8 positions for BRUTUS and MINSTREL (around a third of the total length). In all these cases, the curve tends to flatten beyond distances of about 25 positions. This corresponds to values slightly above half the total length of the story.

The story by Fabulist differs considerably from the others in two respects. First, because it is longer (80 sentences, whereas the others stand at 44, 48, and 58). But also because the curve for it starts at a distance of 10 (almost no causal links shorter than this) and it shows two different peaks: one around a distance of 20 (a fourth of the total length) and another around a distance of 30 (a little less than a third). It also shows a substantial number of links that reach a size of almost 40 positions (which corresponds to around half of the total length of the story).

The graphs for these stories are significantly different from the one obtained for the corpus for the unconstrained human-generated stories from our experiment. The differences in each case are likely to arise from the nature of the story generators being considered in each case. Although a single story is insufficient to provide a general characterization of any story generator, the stories are analyzed in the context of how the generating system operate to obtain insights of



**Fig. 8** Histogram of the relative distance between a story component and the story component it references (both actions and descriptions) in the stories produced by the automatic story generation systems. It can be compared with the human-invented output in Fig. 2

how the proposed mechanism can extract information about story structure and system operation in each case.

The feature represented in this kind of histogram corresponds to the distance between the positions in which sentences related to one another by causal links appear in the discourse. This depends on how elements in the story world related by a causal relation end up in different relative positions in the discourse during the process of establishing a form for delivering the story. In terms of classical natural language generation [39], this corresponds to a process of *discourse planning*, by which the content for the story is organized into a linear discourse. According to Reiter and Dale, this view of the process of generating a text involves a prior stage of *content determination*, during which the choice of which content to include in the story is made. The different distributions over discourse distances between elements related causally in the story world can be interpreted in terms of the content determination and discourse planning processes applied in each story generator. Differences between the patterns observed for outputs of automated story generators and the corpus of collected human stories may arise from two different possibilities.



One is the possibility that the discourse planning procedures being applied by the story generators do not mirror closely the procedures applied by humans.

Another possibility is that the type of story that these story generators are aiming to build correspond to a type of discourse plan that differs in nature from the one being used by our human volunteers in the experiment described in the paper.

The two possibilities are considered separately.

The story generators under consideration were not built within the tradition of natural language generation, and do not in general consider a classic architecture. The MEXICA system relies on a procedure for adding a new story action to a draft of the discourse based on knowledge of whether the story action in question co-occurred in some prior story with story actions already in the draft. However, the process of knowledge extraction is agnostic as to the relative distance between the appearances of these story actions in the prior story. BRUTUS relies on a hierarchy of interrelated grammars, some of which may be considered to represent the kind of information on the discourse level that might affect this particular feature. MINSTREL uses a procedure which adapts an existing discourse (associated with a given moral used as seed) to match the desired topic. If the structure of the discourse used as seed is preserved, stories by MINSTREL should be similar to human ones in structure. Fabulist is the only generator in the set that does consider explicitly a stage of discourse planning, in which the causal graph for the plan obtained as representation of the desired story is linearised into a discourse. However, this discourse planning stage is presented as a basic solution to present the graph as text, rather than as an attempt to model human discourse planning abilities. Based on this analysis, it seems that: it is likely that MEXICA and FabulistT exhibit the final discourses for their stories that differ from human efforts in structure, but it is somewhat surprising that BRUTUS and MINSTREL would.

A further element to consider with respect to discourse planning is the already mentioned fact that some of these story generators tend to state explicitly descriptions of state arising from the actions in a way that humans would not. This is also an important difference with respect to human behaviour which is likely to introduce changes in the final shape of the discourse. The discourse of stories by MINSTREL and FABULIST is quite likely to include redundant statements that declare explicitly information that human readers can deduce easily from the context. The presence of these redundant statements would significantly alter the distances in the discourse between causally related elements.

The possibility that the story generators be following a type of discourse plan different from those applied by the human volunteers in the experiment reported in this paper has a high likelihood. The analysis in “[The ROC Story Corpus](#)” of the differences evidenced in the ROC story corpus showed that there are many possible ways of structuring a story, and that humans are masters in the art of moulding their output to the conditions imposed by constraints at different levels. As the story generators showing unexplained differences declare explicitly to be aiming for stories conforming to very specific genres (King Arthur stories with a moral for MINSTREL; stories about betrayal for BRUTUS), it is very possible that such genres employ specific discourse structures that differ from those put in practice by participants in the

experiment. Priming the participants in the experiment with indications of the type of story desired might have led to completely different discourse structures.

## Conclusions and Future Work

The paper has summarized an empirical analysis of human-invented stories that tried to gather insight on the amount and proportion of story components. The research hypothesized that simple story plots invented by humans use a relatively common set of components and relations between them, and that these components are structured similarly. The results evidence this tendency for short writing sessions, and they have permitted to provide a first approximation to the parameter values of the actual amounts used.

The results evidence that short plots invented by humans show a *clear division with descriptions at the beginning, and actions at the end, linear causality references, incremental story construction, and almost full interconnection of plot events*. Additionally, the annotation has provided actual proportions as used by humans. These proportions have a relatively common pattern and we believe they can be used to inform story generation systems in order for the programs to produce content that resembles human construction better.

Two examples of application of the proposed measures to stories from different sources have been included. The corresponding discussion shows that valuable insights can be obtained on the nature of the stories and the construction process employed to produce them. The analysis has also shown that there is no single structure for a story, but rather a number of configurations of the basic parameters that may be used to characterize types of stories. Further work is required to identify and describe these configurations. In this context, the proposed measures have shown their value as mechanisms to make explicit the features that define these different configurations.

The proposed measures have also shown their value as auxiliary tools for the design and validation of automated storytelling systems. By constraining the story to representations based on observations like the ones presented in this paper, the systems' output can approximate human-like behaviour. As such, the outcome of this research is intended to contribute as one of the first steps towards a set of commonly accepted metrics of storytelling when creating formal creativity-based programs like story generation systems. The long-term objective is to be able to develop a set of evidence-based models, agreed by the community as a common framework to compare and share resources.

The current results do not provide coverage for the vast amount of phenomena that takes place in full, rich story creation as carried out by humans. Such phenomena are currently beyond the scope of the representational capabilities of computational systems. Further work should extend the approach to explore experiment-based models for other aspects that might be relevant, such as: use of characters beyond the protagonist, emotional relations between characters, scene separation, and more.

As introduced in "[Introduction](#)", the study has addressed the particular perspective of human-like narrative structures. The experimental design and the results do not provide coverage for other types of narrative; for instance, narrative construction based on computational metrics that do not resemble human processes. Future work

in Computational Creativity must address this, both as a study of the non-human metrics and a comparative analysis of the differences between human and machine-based metrics for narrative structures.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References


1. Allen, J.F.: Time and time again: the many ways to represent time. *Int. J. Intell. Syst.* **6**, 341–355 (1991). [citeseer.ist.psu.edu/allen91time.html](http://citeseer.ist.psu.edu/allen91time.html)
2. Bal, M.: *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, Toronto (2009)
3. Barthes, R., Duisit, L.: An Introduction to the Structural Analysis of Narrative. *New Literary History* **6**(2), 237–272 (1975). <https://doi.org/10.2307/468419>
4. Bringsjord, S., Ferrucci, D.: *Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, a StoryTelling Machine*. Lawrence Erlbaum Associates, Hillsdale, NJ (1999)
5. Bruner, J.: The narrative construction of reality. *Crit. Inquiry* **18**(1), 1 (1991). <https://doi.org/10.1086/448619>
6. Chatman, S.: *Story and Discourse: Narrative Structure in Fiction and Film*, vol. Second. Edi. Cornell University Press, USA (1986)
7. Dehn, N.: Story generation after tale-spin. In: *Proceedings of the international joint conference on artificial intelligence*, pp. 16–18 (1981)
8. Delatorre, P., Arfè, B., Gervás, P., Palomo-Duarte, M.: A component-based architecture for suspense modelling. In: *Proceedings of AISB 2016's 3rd international symposium on computational creativity (CC2016)*, pp. 32–39. Sheffield (UK) (2016). <http://hdl.handle.net/10498/18328>
9. Fan, A., Lewis, M., Dauphin, Y.: Hierarchical neural story generation. [arXiv:1805.04833](https://arxiv.org/abs/1805.04833) (2018)
10. Forster, E.M.: *Aspects of the Novel*. Harcourt, New York (1927)
11. Freytag, G.: *Technique of the Drama*. S. Hirzel (1863). <http://books.google.es/books?id=CDYLA AAAIAAJ>
12. Genette, G.: *Narrative Discourse: An Essay in Method*. Cornell University Press, New York (1979)
13. Gervás, P.: Computational approaches to storytelling and creativity. *AI Magn.* **30**(3), 49–63 (2009)
14. Gervás, P.: Comparative evaluation of elementary plot generation procedures. In: *Proceedings of the 6th International Workshop on Computational Creativity, Concept Invention, and General Intelligence* (2017)
15. Gervás, P., León, C.: Reading and writing as a creative cycle: the need for a computational model. In: *5th International Conference on Computational Creativity, ICCO 2014*. Ljubljana, Slovenia (2014)
16. Graesser, A.C., Singer, M., Trabasso, T.: Constructing inferences during narrative text comprehension. *Psychol. Rev.* **101**, 371–395 (1994)
17. Gustafsson, A., Bichard, J., Brunnberg, L., Juhlin, O., Combetto, M.: Believable environments: generating interactive storytelling in vast location-based pervasive games. In: *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, pp. 24–es (2006)
18. Herman, D.: *Narratology as a Cognitive Science*. Image and Narrative (2000)
19. Herman, D.: *Story Logic: Problems and Possibilities of Narrative* (2002)
20. Hervás, R., Sánchez-Ruiz, A., Gervás, P., León, C.: Calibrating a metric for similarity of stories against human judgment. In: *CEUR Workshop Proceedings*, vol. 1520 (2015)

21. Kaufman, J.C., Beghetto, R.A.: Beyond big and little: the four c model of creativity. *Rev. Gen. Psychol.* **13**(1), 1–12 (2009). <https://doi.org/10.1037/a0013688>
22. Khalpada, P., Garg, S.: Balancing consistency and plot structure in computational storytelling. In: 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEM-CON), pp. 0550–0555. IEEE (2019)
23. Lebowitz, M.: Creating Characters in a Story-Telling Universe. *Poetics* **13**, 171–194 (1984)
24. León, C.: An architecture of narrative memory. *Biol. Inspired Cognit. Architect.* **16**, 19–33 (2016). <https://doi.org/10.1016/j.bica.2016.04.002>. <http://www.sciencedirect.com/science/article/pii/S2212683X16300184>
25. León, C.: Narrative as a fundamental information unit in the mind. In: AAAI fall symposium—technical report, vol. FS-17-01 (2017)
26. Lim, C.U., Harrell, D.F.: An approach to general videogame evaluation and automatic generation using a description language. In: 2014 IEEE Conference on Computational Intelligence and Games, pp. 1–8. IEEE (2014)
27. Livytska, I.A.: The art of narration and artificial narrative intelligence: implications for interdisciplinary research. *J. Narrat. Lang. Stud.* **7**(13), 309–318 (2019)
28. Mateas, M., Stern, A.: Procedural authorship: A case-study of the interactive drama *Façade*. *Digital Arts and Culture (DAC)* **61**, (2005)
29. Mostafazadeh, N., Chambers, N., He, X., Parikh, D., Batra, D., Vanderwende, L., Kohli, P., Allen, J.: A corpus and evaluation framework for deeper understanding of commonsense stories. *arXiv:1604.01696 [cs]* (2016)
30. Mostafazadeh, N., Roth, M., Louis, A., Chambers, N., Allen, J.: LSDSem 2017 shared task: The story cloze test. In: Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics, pp. 46–51. Association for Computational Linguistics, Valencia, Spain (2017). <https://doi.org/10.18653/v1/W17-0906>
31. Peinado, F., Francisco, V., Hervás, R., Gervás, P.: Assessing the novelty of computer-generated narratives using empirical metrics. *Mind. Mach.* **20**(4), 565–588 (2010)
32. Pérez y Pérez, R., Sharples, M.: MEXICA: a computer model of a cognitive account of creative writing. *J. Exp. Theor. Artif. Intell.* **13**(2), 119–139 (2001)
33. Pérez y Pérez, R., Sharples, M.: Three computer-based models of storytelling: BRUTUS, MINSTREL and MEXICA. *Knowl. Based Syst.* **17**(1), 15–29 (2004)
34. Porter Abbott, H.: *The Cambridge introduction to narrative*, second edition (2014). <https://doi.org/10.1017/CBO9780511816932>
35. Propp, V.: *Morphology of the Folk Tale*. Akademija, Leningrad (1928)
36. Purdy, C., Wang, X., He, L., Riedl, M.: Predicting generated story quality with quantitative measures. In: 14th Artificial Intelligence and Interactive Digital Entertainment Conference (2018)
37. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: *Language models are unsupervised multitask learners* (2019)
38. Rank, S., Hoffmann, S., Struck, H.G., Spierling, U., Petta, P.: Creativity in configuring affective agents for interactive storytelling. In: M.L.M.K.H.A.P.R.P. y Pérez; Dan Ventura; Geraint Wiggins (ed.) *International conference on computational creativity*, pp. 165–169. ICCS (2012)
39. Reiter, E., Dale, R.: *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge (2000)
40. Riedl, M.: *Narrative Planning: Balancing Plot and Character*. Ph.D. thesis, Department of Computer Science, North Carolina State University (2004)
41. Rowe, J.P., McQuiggan, S.W., Robison, J.L., Marcey, D.R., Lester, J.C.: Storyeval: An empirical evaluation framework for narrative generation. *AAAI Spring Symp. Intell. Narrat. Technol.* **11**, 103–110 (2009)
42. Ryan, M.L.: Narratology and cognitive science: a problematic relation. *Style* **44**(4), 469–496 (2010). <https://doi.org/10.1007/s13398-014-0173-7.2>
43. Ryokai, K., Cassell, J.: Computer support for children’s collaborative fantasy play and storytelling. In: Proceedings of the 1999 Conference on Computer Support for Collaborative Learning, CSCL ’99, p. 63–es. International Society of the Learning Sciences (1999)
44. Sakamoto, M., Nakajima, T.: Incorporating fictionality into the real world with transmedia storytelling. In: Marcus, A. (ed.) *Design, User Experience, and Usability: Design Discourse*, pp. 654–665. Springer, Cham (2015)
45. Schank, R., Abelson, R.: *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. L. Erlbaum, Hillsdale, NJ (1977)

46. Schneider, O.: Storyworld creation: Authoring for interactive storytelling. In: The 10-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2002, WSCG 2002, University of West Bohemia, Campus Bory, Plzen-Bory, Czech Republic, February 4–8, 2002, pp. 405–412. DBLP (2002). [http://wscg.zcu.cz/wscg2002/Papers\\_2002/D67.pdf](http://wscg.zcu.cz/wscg2002/Papers_2002/D67.pdf)
47. Szilas, N.: Towards Narrative-Based Knowledge Representation in Cognitive Systems. In: Proceedings of the 6th Workshop on Computational Models of Narrative (CMN'15), pp. 133–141 (2015)
48. Tapscott, A., Javier, G., León, C., Smailovic, J., Znidarsic, M., Gervas, P.: Empirical evidence of the limits of automatic assessment of fictional ideation. In: T. Besold, O. Kutz, C. León (eds.) Proceedings of the 5th International Workshop on Computational Creativity, Concept Invention, and General Intelligence, pp. 58–71 (2016)
49. Theune, M., Faas, E., Nijholt, A., Heylen, D.: The Virtual Storyteller: Story creation by intelligent agents. In: Proceedings of the Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) Conference, pp. 204–215 (2003)
50. Theune, M., Meijs, K., Heylen, D., Ordelman, R.: Generating expressive speech for storytelling applications. *IEEE Trans. Audio Speech Lang. Process.* **14**(4), 1137–1144 (2006)
51. Trabasso, T., van den Broek, P., Suh, S.Y.: Logical necessity and transitivity of causal relations in stories. *Discourse Processes* **12**, 1–25 (1989)
52. Trabasso, T., Sperry, L.: Causal relatedness and importance of story events. *J. Mem. Lang.* **24**, 595–611 (1985)
53. Turner, S.R.: MINSTREL: a computer model of creativity and storytelling. Ph.D. thesis, University of California at Los Angeles, Los Angeles, CA (1992)
54. Wang, K., Bui, V., Petraki, E., Abbass, H.A.: From subjective to objective metrics for evolutionary story narration using event permutations. In: Asia-Pacific Conference on Simulated Evolution and Learning, pp. 400–409. Springer (2012)
55. Wang, K., Bui, V., Petraki, E., Abbass, H.A.: Human-guided evolutionary story narration. *IEEE Access* **6**, 13783–13802 (2018)
56. Ware, S.G., Young, R.M., Harrison, B., Roberts, D.L.: Four quantitative metrics describing narrative conflict. In: International Conference on Interactive Digital Storytelling, pp. 18–29. Springer (2012)
57. Zarrì, G.P.: NKRL, a knowledge representation tool for encoding the 'meaning' of complex narrative texts. *Nat. Lang. Eng.* **3**(2), 231–253 (1997)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Carlos León<sup>1</sup>  · Pablo Gervás<sup>1</sup> · Pablo Delatorre<sup>2</sup> · Alan Tapscott<sup>1</sup>

Pablo Gervás  
pgervas@ucm.es

Pablo Delatorre  
pablo.delatorre@uca.es

Alan Tapscott  
alan.tapscott@ucm.es

<sup>1</sup> Department of Software Engineering and Artificial Intelligence and Institute for Knowledge Technologies, Complutense University of Madrid, Madrid 28040, Spain

<sup>2</sup> Department of Computer Science, University of Cadiz, Cadiz 11001, Spain