TropeTwist:
Trope-based Narrative Structure Generation

Alberto Alvarez
Malmö University, Game Lab
Malmö, Sweden
alberto.alvarez@mau.se

Jose Font
Malmö University, Game Lab
Malmö, Sweden
jose.font@mau.se

ABSTRACT

Games are complex, multi-faceted systems that share common elements and underlying narratives, such as the conflict between a hero and a big bad enemy or pursuing a goal that requires overcoming challenges. However, identifying and describing these elements together is non-trivial as they might differ in certain properties and how players might encounter the narratives. Likewise, generating narratives also pose difficulties when encoding, interpreting, and evaluating them. To address this, we present TropeTwist, a trope-based system that can describe narrative structures in games in a more abstract and generic level, allowing the definition of games' narrative structures and their generation using interconnected tropes, called narrative graphs. To demonstrate the system, we represent the narrative structure of three different games. We use MAP-Elites to generate and evaluate novel quality-diverse narrative graphs encoded as graph grammars, using these three hand-made narrative structures as targets. Both hand-made and generated narrative graphs are evaluated based on their coherence and interestingness, which are improved through evolution.

KEYWORDS

Authoring Tools, Narrative Generation, Evolutionary Computation, MAP-Elites, Computer Games

1 INTRODUCTION

There exists a plethora of games[1], with diverse genres and each containing a different set of gameplay mechanics, audio, level, graphic, and narrative facets. The creation and combination of these facets make game development a hard task, commonly involving a diverse group of developers [12]. Likewise, the generation of these facets in conjunction has been categorized as one of the biggest and most challenging tasks within computational creativity [36, 37]. However, games share common elements and underlying narratives, but it is non-trivial how to identify these, how to define and analyze these games structurally, or what type of common underlying structures exist; pointed out as well by [3, 59].

Among the different facets, narrative stands out in games as it helps to create meaning, make sense of situations, and make games [stories] recognizable [2, 32, 38, 56]. Narrative structures can be used to describe how an experience or story is to be developed as argued by Barthes [11], and to create an abstract representation based on the narrative structure instead of a temporal and partially-ordered sequence of events [51]. Common narrative structures used in many domains are Aristotle's drama structure, which subdivides a story into *exposition*, *climax*, and *resolution* or Propp's analysis

---

[1]For instance, currently there are more than 68k games in steam https://store.steampowered.com/search/?category1=998.

on the morphology of russian folktale, which revealed a common structure among them, denoted as Propp's 31 "narremes" [45].

This paper presents *TropeTwist*, a preliminar system that uses Tropes [21, 54] extracted from TvTropes [26, 46] as patterns and fundamental units, which when combined can compose structures further representing other composed tropes. Common narrative structures can be identified and defined using *TropeTwist*. TropeTwist can define generic aspects of a story, leading to the identification of events, roles, and narrative elements, as well as a novel way to form narratives. As a proof-of-concept, we built, analyzed, and described structurally three game examples shown in figure 1, top row.

We propose graph grammars as indirect encoding of narrative graphs and the use of the Multi-dimensional Archive of Phenotypic Elites (MAP-Elites) [40] to generate novel variations (shown in figure 1, bottom row) using the proof-of-concept examples as roots. Simultaneously, we propose metrics to evaluate the resulting narrative graphs' coherence, cohesion, and interestingness. Our preliminary results show that we can produce more interesting structures retaining coherence based on our metrics.

2 RELATED WORK

Propp [45] analyzed Russian folktales identifying their fundamental structure in 31 steps. His work contributed to the identification of core elements, the proposal of actions and events as *functions* and narrative atoms, and roles that are recurrent within the folktales. Propp emphasized that these 31 *functions* and their arrangement were the structure and what gave meaning to the story discourse. Barthes [11] proposed three intertwined and progressively integrated levels in narrative work: *functions*, *actions*, and *narration*. His work is characterized by the proposal of fundamental narrative units in the *function* level to better assess and identify structures in a narrative. Furthermore, Baikadi and Cardona-Rivera [10] further discuss these fundamental units as *narremes* encoding narrative state and how they could be combined to narrative structures. Their work, similar to TropeTwist, proposes a graph structure of interconnected *narremes*. However, they defined narrative axes like Barthes, where each connection between *narremes* means a change along a narrative axis. In games, the narrative is usually directed by quests, which Aarseth [1] discusses as a central element in games to make sense of other elements, and which are defined by Yu et al. as a form of structure, dividing the story into achievable rewards and partially ordered set of tasks [62].

Furthermore, the generation of narratives, stories, and quests using a variety of techniques such as planning algorithms [47, 61], grammars [7, 27], or machine learning [53, 58], is a growing and important field within games research and narrative research in general [18, 22, 32, 62]. One typical approach for the generation of
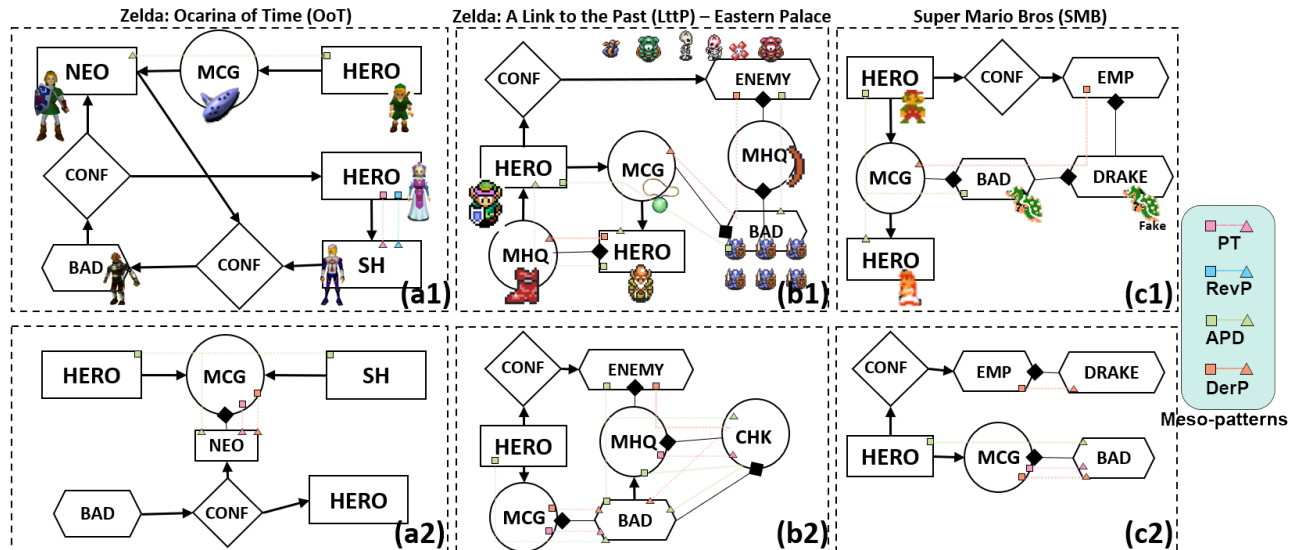
Figure 1: Proof-of-concept narrative structures of existing games (top row) created using TropeTwist with the available nodes (table 1). Bottom row shows exemplar elites generated with MAP-Elites using the respective top row narrative structure as root. Color matching squares, lines, and triangles denote different meso-patterns in the structures. Squares and triangles are the start and end of a meso-pattern, respectively.

content and stories is the use of patterns representing different elements such as level design patterns [4, 56], quest patterns and common quests in games [16, 57], or identifying fundamental units and assembling them based on various pre-conditions [20, 31]. A particular type of pattern is tropes, which are concepts that are recurrently used in transmedia storytelling [21, 46]. Horswill [28] focused on constructing an expressive language that could encode plot tropes as story fragments, composing a database of fragments combined sequentially with a planner. Similarly, Thompson et al. [54] used the idea of tropes as story bits where a system would construct valid stories from users' defined story bits with pre-and post-conditions. TropeTwist uses the idea of tropes for nodes and patterns in structures and encodes and represents these as a graph. *Scheherazade* is a system that can capture narrative structures by encoding and annotating narrative texts, which introduced the Story Intention Graph model, a formal and expressive representation of narratives [19].

Moreover, we use graph grammars and grammar recipes to generate structures. This approach is similar to how Dormans and Bakkes [17] generate missions and space using a "key and lock" structural idea. Our approach uses MAP-Elites, a quality-diversity algorithm that uses behavioral dimensions that are orthogonal to the objective function to store diverse individuals in a grid [40]. Evolutionary algorithms are a popular approach in PCG to generate diverse type of content [55], but not as much for narrative content. MAP-Elites have been used to generate content in different game facets such as levels [5, 15], mechanics [14], or enemy behavior [29].

Assessing narratives is a complex and non-trivial task. The goal is to create a narrative that is both syntactically correct (e.g., coherent and consistent) and semantically rich (e.g., novel and interesting) [13, 25, 48]. Perez y Perez and Ortiz [44] proposed a model to evaluate interestingness based on novelty and correct story recount,

Table 1: Tropes included and used in TropeTwist, extracted from [46].

Name	Symbol	Definition
Hero	HERO	A protagonist character.
Five-man band	5MA	Group composed by up-to-five archetypical characters.
The chosen one	NEO	Specific hero chosen as the one.
Superhero	SH	Specific hero with unique abilities.
Conflict	CONF	Non-specific problem to overcome between characters.
Enemy	ENEMY	A nemesis to the hero.
Empire	EMP	Collective enemy with the ambition of conquering the world.
Big bad	BAD	Specific enemy, which is the ultimate cause for all the bad.
Dragon	DRAKE	Specific enemy, which is the right hand of BAD.
Plot device	PLD	A feature or element that drives the plot forward.
Chekhov's gun	CHK	PLD relevant to the story
MacGuffin	MCG	PLD with irrelevant nature to drive the story.
May help in quest	MHQ	PLD important to resolve a conflict.

with emphasis on the story's opening, closure, and dramatic tensions. Szilas et al. [52] discuss interestingness as a paradox dramatic situation with obstacles and conflicts, albeit applicable to stories as successive events. Yet, to approach subjective measurements such as interestingness, most research turns towards having human evaluation [30, 33] or using such to form human models to be used as surrogate models [35, 50].

3 BUILDING NARRATIVE STRUCTURES WITH TROPES

In storytelling, a trope [21] is a convention or figure of speech that the storyteller assumes to be recognizable by the audience. TvTropes is an online wiki that compiles and describes several thousand tropes in many sorts of media [46]. As exemplified by [26], tropes could be interconnected in graph-like structures, called story molecules, to succinctly depict the structure behind a narrative.

3.1 TropeTwist

TropeTwist elaborates on the concept of story molecule to represent narratives using graph-like structures of interconnected tropes, called narrative graphs (NG). NGs encode narrative structures in an abstract level that show and define the game's narrative structure and certain abstract properties such as key items, roles, relations, or main events. Table 1 shows all the included tropes to be used as nodes. Nodes are depicted (fig 1) with shapes specific to their trope base type: heroes (rectangle), conflicts (diamond), enemies (hexagon), and plot devices (circle). HERO is the base pattern of 5MA, NEO, and SH. ENEMY is the base pattern of EMP, BAD, and DRAKE. PLD is the base pattern of CHK, MCG, and MHQ.

Nodes in a narrative graph are necessarily interconnected by either unidirectional or bidirectional edges (with one or both arrowheads) or by entailment edges (with a single diamond head). Given nodes A and B, A $\diamond-$ B, reads as "A entails B," whereas A $\rightarrow$ B denotes a relationship from A to B, and B $\rightarrow$ A the opposite. A $\leftrightarrow$ B denotes a reflexive relationship between A and B. As an example, HERO $\rightarrow$ CONFLICT $\rightarrow$ EMP denotes a hero who is in conflict against an empire-type enemy, whereas HERO $\leftrightarrow$ CONFLICT denotes a hero who is in conflict with themselves. EMP $\diamond-$ DRAKE $\diamond-$ NEO, denotes an empire that entails a dragon enemy that, once beaten, will lead to the appearance of a chosen one hero, creating some causal links. The system is ambiguous by design. We take advantage of the ambiguity for 1) the generation of new structures (fewer constraints), 2) removing the focus on details by designers to let them focus on the overarching picture, and 3) for other systems to define and interpret these abstract properties.

Furthermore, interconnecting tropes can give rise to other tropes and patterns, described in the following section. The nodes and their respective trope and pattern were chosen from a subset of tropes in generic categories such as heroes or plot devices. These categories were inspired and chosen based on tropes from TVTropes, the division by James Harris [26], and previous research such as Propp's morphology [45] or Greimas' actantial model [24].

3.2 Trope Patterns

Tropes and interconnected tropes (i.e., subgraphs) give rise to different types of patterns. These patterns can be **micro-patterns**, encapsulating a single trope node, **meso-patterns**, often composed by more than one micro-pattern with special meaning, and **auxiliary patterns**, denoting graph problems. We calculate the relative tropes and patterns' quality within an NG and use this to assess the general quality of the graph. These qualities are proxies for certain characteristics among the defined patterns that are used to evaluate the graphs, but they do not capture any story quality; especially, since we are only defining structures. When generating

narrative graphs from a root (explained in section 4), the quality of a narrative graph becomes relative to the root, henceforth, the "root graph" (RG). In the following descriptions, we will use **EG** referring to the "evaluated graph" we are calculating the pattern's quality (the generated individual), and **RG** to refer to the relative and root graph. When using subscript "pat," we refer to the current pattern that is evaluated.

For most patterns, we calculate three general qualities (indicated when used) that add to the quality of the pattern. $G_q(pattern)$ relates to the *Generic* quality of patterns in EG, which calculates the general occurrence of a pattern within EG compared to its occurrence in RG, calculated in eq. 1. $R_q(pattern)$ relates to the *Repetition* quality of patterns, which calculates if a trope is unique in EG ($R_q(pattern) = 1$) or its ratio among the same base pattern. Lastly, $I_q(pattern)$ relates to the *Involvement* quality of patterns in EG, which calculates the amount of associations a pattern has with **structure patterns**. Involvement means that the pattern is either *source* or *target* in a structure and is calculated as the ratio of structure pattern involvement by the structure pattern count in EG. These three metrics incentivize graphs with similar amount and type of nodes than RG, minimal repetitions, and more involvement.

$$G_q(pattern) = 1.0 - |RG_{pat} - EG_{pat}|/\max(RG_{pat}, EG_{pat}) \quad (1)$$

3.2.1 Micro-Patterns. Micro-patterns are the fundamental unit in the system, which aims at categorizing different sets of the individual patterns that are shown in table 1. Micro-patterns are single nodes and the basic building block that, when interconnected, allows the detection of meso-patterns.

Structure Pattern (SP) is any type of trope that would give some structural definition to a narrative, whether this being a conflict, specific act, or a part in a dramatic arc (e.g., climax). Currently, the only type of structure trope is the CONFLICT (CONF) trope, which represents the most basic structural interaction. The quality SP_q is calculated as the equally weighted linear combination of:

$$SP_q = G_q(SP) + I_q(SP) \quad (2)$$

Character Pattern (CP): are identified as nodes within the narrative that could be either the player, possible ally or enemy NPCs, or simple enemies. In TropeTwist, it is distinguished between heroes and villain patterns, and these are commonly used as **sources** or **targets** (or both) of other patterns, and on a few special occasions to denote a relation to another character. The quality CP_q is calculated per group (heroes and villains), and it is the equally weighted linear combination of:

$$CP_q = G_q(CP) + R_q(CP) + I_q(CP) \quad (3)$$

Plot Device Pattern (PDP) is described as the element within the narrative that moves it forward, as a goal, object, or dramatic element to be used or encountered by any of the characters. The quality PDP_q is calculated as the equally weighted linear combination of:

$$PDP_q = G_q(PDP) + R_q(PDP) \quad (4)$$

3.2.2 Meso-Patterns. Meso-patterns are the features that emerge in the narrative from dynamically combining micro-patterns and, on some occasions, these with other meso-patterns. They are always

composed of more than one pattern denoting some spatial, semantic, or usability relationship within the narrative graph. We identified a subset of Tropes (extracted from TVTropes [46]) that requires or works as the combination between more fundamental units. For instance, the *reveal pattern* relates to the "Good all along" or "evil all along."

Conflict Pattern (ConfP) is a type of structure pattern composed by a conflict node (Con), a source s node, and a target t node, which are both CPs and usually a hero and a villain or the same character as s and t. For instance, the subgraph HERO $\rightarrow$ CONFLICT $\rightarrow$ EMP, indicates that a hero CP has a conflict with an enemy CP. A conflict node can be used indefinitely to define several ConfP. A ConfP is also either EXPLICIT or IMPLICIT. EXPLICIT conflicts are explicitly encoded in the graph and directed from s to t passing through the conflict trope. On the other hand, IMPLICIT conflicts relates to the conflicts from t (or derivatives) to s (or derivatives) that are not encoded in the graph. For instance, the previous example is an EXPLICIT conflict from HERO to EMP, and at the same, the EMP has an IMPLICIT conflict with the HERO. The quality $ConfP_q$ is calculated as the equally weighted linear combination of:

$$ConfP_q = G_q(ConfP) + R_q(ConfP) \quad (5)$$

Derivative Pattern (DerP) defines a relationship between tropes connected by "entails" connections ($\diamond-$). Therefore, a DerP contains a list of patterns connected by entails, named derivatives. DerP starts from a root micro-pattern and continue until no more "entail" connections are encountered, effectively establishing a hierarchy from the root derivative to the rest. By design, the patterns within a DerP have a local and temporal order and a causal relationship. For instance, in the subgraph EMP $\diamond-$ DRAKE $\diamond-$ NEO, engaging with the *EMP*, entails both the conflict with *DRAKE* and the appearance of *NEO*. This means that only by overcoming the *DRAKE*, *NEO* will appear - as a new hero or the evolution of another. The quality $DerP_q$ is calculated (eq. 6) based on its $G_q(DerP)$, the ratio of derivatives within the DerP among the total amount of derivatives across all DerPs in EG ($ratio\theta_q$), and the derivatives' diversity.

$$DerP_q = G_q(DerP) + ratio\theta_q + \frac{\sum_{i=0}^{len(DerP_{der})} DerP_{der_i basepat}}{len(DerP_{der})} \quad (6)$$

Reveal Pattern (RevP) connects two independent CPs as one, meaning that character A was, in fact, always character B, and vice-versa. This pattern identifies confusion and surprise within an EG, as, for instance, a villain could have been, in fact, "Good All Along"[2]. In practice, a RevP is identified as a villain or hero connected with a unidirectional connection ($\rightarrow$) to another hero or villain. As a consequence, all existing conflicts between them would become *fake*. $RevP_q$ is calculated based on its $G_q(RevP)$, the number of reveals in EG in relation to characters, and the number of fake conflicts given the specific reveal.

---

[2]https://tvtropes.org/pmwiki/pmwiki.php/Main/GoodAllAlong

$$RevP_q = G_q(RevP) + \frac{len(EG_{RevP})}{len(EG_{CP})} +$$
$$\left(1.0 - \frac{\sum_{i=0}^{len(EG_{conf})} \begin{cases} 1, & \text{if } RevP \in x_i \\ 0, & \text{otherwise} \end{cases}}{len(EG_{conf})}\right) \quad (7)$$

Active Plot Device Pattern (APD) operationalize and integrate PDPs within a narrative since PDPs only describe an abstract goal or target. In practice, an *APD* is identified as PDPs that have at least one incoming connection, and optionally, one single outgoing connection. These limitations are added to limit the effect of a PDP within a narrative. APD_q is measured based on its $G_q(APD)$, and the APD's usability, calculated based on the sum of incoming and outgoing connections divided by half of the nodes in EG depicted as $bal\gamma_q$, penalizing APDs for not using all their connections.

$$APD_q = G_q(APD) + bal\gamma_q \quad (8)$$

Plot Points (PP) are key events within the EG, identified as discrete moments given some pattern. The derivatives within a *DerP*, RevP's source, and PDPs that are *APD* are considered as plot points. PP_q is measured based on the number of PPs within RG ($G_q(PP)$), and the number of PPs within EG in relation to the number nodes within it ($Balance_q(PP)$).

$$PP_q = G_q(PP) + Balance_q(PP) \quad (9)$$

Plot Twist (PT) takes advantage of plot points to identify those that could have a bigger impact on the narrative. In practice, *PTs* consider the source of *RevP*, derivatives from *DerP* that are a different micro-pattern than the root of the DerP (except PDPs), and *APDs* that are connected to other *APDs*. For instance, in the subgraph: EMP $\diamond-$ DRAKE $\diamond-$ NEO, given that NEO is a different micro-pattern than root EMP (Hero and Villain, respectively), NEO will be identified as a *Plot Twist* as it alters the "natural" order in the DerP. PT_q is based on the number of PTs within RG ($G_q(PT)$), the PT's involvement in EG, and the balance of PTs based on the PPs in EG. Involvement varies depending on the associated pattern to PT. When a PT is associated with a *RevP*, involvement is calculated as how much the structure changes based on that (i.e., how many fake conflicts are created). When it is related to *DerP*, involvement is calculated as how different the pattern is and its order within the derivatives. Finally, when it is related to *APD*, involvement is based on how usable the *APD* is within the narrative based on incoming and outgoing connections.

$$PT_q = G_q(PT) + I_q(PT_{assoc_p at}) + \frac{len(EG_{pt})}{len(EG_{pp})} \quad (10)$$

3.2.3 Auxiliary Patterns. Auxiliary patterns denote problems in the graph and sub-optimal or impractical nodes and connections within a graph. They are classified into *Nothing*, which are nodes that are not identified as part of a meso-pattern; and *Broken Link*, which are outgoing connections from a node that are not used or do not lead to any pattern.

3.3 Proof-of-Concept

TropeTwist can be used to represent different narrative structures and parts of games. To test and show TropeTwist's expressiveness, we chose to form three different narrative graphs representing different games shown in figure 1, top row: *Zelda: Ocarina of Time* (Zelda:OoT) [43], *Zelda: A Link to the Past* (Zelda:LttP) [42] - eastern palace, and *Super Mario Bros* (SMB) [41]. They represent different games from different genres (fig. 1.a and 1.b are adventure-dungeon games, and 1.c is a platformer), and represent different game's phases; in the case of fig. 1.a and 1.c, both represent the main structure of the game, while 1.b, represents a specific area and sequence of the game.

Figure 1.a represents a simplified overarching narrative structure from Zelda: OoT. The ocarina of time, given by Zelda to Link, is defined as a McGuffin (MCG) that, when collected by "young link," allows him to go forward in time to "adult link," the chosen one (NEO). This, in turn, enables explicit conflicts between hero and enemy characters, which represents the main loop of the game. The structure shows two factions, a set of heroes and the BAD.

Figure 1.b represents the structure and plot points from the eastern palace in Zelda: LttP. All palaces in *A Link to the Past* follow a very similar structure and sequence. The HERO's goal is to get the "Pendant of Courage" (MCG). However, the MCG derives from ENEMY and BAD, so the HERO must overcome them to achieve his goal. The structure shows a causal and linear narrative that could be used to identify elements that need to appear before others, similar to the work by Dormans and Bakkes [17].

Figure 1.c represents the overarching narrative structure of SMB. In SMB, the objective of Mario (HERO) is to rescue Peach (HERO) from Bowser (BAD). To do this, the player goes through a series of platform worlds that always end in a "Fake Bowser" (DRAKE). The player must continue until encountering the "Real Bowser" (BAD), which then would enable the player to get to their objective (MCG).

4 EVOLVING NARRATIVES WITH GRAPH GRAMMARS

We use the Constrained MAP-Elites [29], and adapt it to work with graph grammars, evolve production rules, and adapt the evolution towards a target similar to [5]. Constrained MAP-Elites adds feasible-infeasible two populations to each cell, effectively evolving sub-populations per cell. An individual's phenotype is a narrative graph, and its encoding genotype is the production rules of a graph grammar. A graph grammar is a context-free grammar whose productions add, remove, and modify nodes and edges of a graph. Our implementation uses the tropes listed in Table 1 as nodes, and the three available connection types as edges ($\rightarrow$, $\leftrightarrow$, $\diamond-$). Graph grammars do not apply rules sequentially; instead, every individual does a random sampling of the rules in their genotype to produce *recipes* to generate graphs. *Recipes* describe the rules' order and repetition, and their size is limited by the amount of production rules as minimum and the minimum plus five as maximum. *Recipes* do not have repetitions within them, i.e., if rule 1 is added at step 2, subsequent addition would simply add to the number of times that rule will be applied at step 2. Their size is limited by the number of production rules as minimum and up to five more samples as maximum.

Figure 2 shows a sample complete process from an individual's genotype (i.e., rules) to the phenotype (i.e., narrative graph).

Individuals move between the feasible and infeasible population depending on the feasibility constraint. NGs are deemed infeasible if the nodes are not fully connected or if there exists a conflict pattern with more than one self-conflict. Infeasible individuals are evaluated based on how close they are to be fully connected and not having any inadequate self-conflict. The fitness function assesses NGs that are deemed feasible based on their coherence (equation 12), which we use to assess how correct, coherent, and in general, syntactically correct the narrative graphs are. Coherence aims at maximizing an equally weighted sum between cohesion and consistency. Cohesion refers to the link between elements that hold together to form some group. In our implementation, it focuses on minimizing the number of auxiliary patterns by calculating the proportion of *Nothing* and *Broken Link* among all patterns in NG. A consistent NG should be regular and free of contradictions. Thus, we calculate *consistency* (eq. 11) as the collective quality of micro-patterns since they are the building blocks, and conflicts' goodness based on the number of fake conflicts. Thus, we aim at maximizing the quality of micro-patterns and minimizing contradictions created by meso-patterns.

$$f_{consistency} = \frac{\sum_{i=0}^{len(ng_{micro})} i_{qual}}{len(ng_{micropat})} - \frac{len(ng_{fakeConfP})}{len(ng_{confP})} \quad (11)$$

$$f_{coherence} = f_{consistency} + (1.0 - f_{cohesion}) \quad (12)$$

Furthermore, MAP-Elites uses behavioral dimensions in a grid shape to retain and foster diversity throughout generations. We use the following two dimensions to evaluate the diversity:

Step. Step (eq. 13) calculates the Levenshtein distance [34] between two narrative graphs, taking into consideration the number and type of nodes and connections. Step is normalized using step threshold $\theta = 11$ determined through a process of experimentation, which does not consider steps farther than θ, avoiding the generation of too dissimilar graphs.

$$D_{step} = \min(lev_{EG,RG}(|EG|, |RG|), \theta) \quad (13)$$

Interestingness (int). We aim at measuring the semantic quality of a narrative graph. A narrative graph can be syntactically correct and coherent yet lack a good semantic quality and do not evoke interest for designers or players. Therefore, we leverage **plot point**, **plot twist**, and **active plot device** patterns to measure the *interestingness* of the NGs. The nature of *interestingness* creates pressure on the fitness function since the incidence of the three meso-patterns could (if overused) "degenerate" the narrative; thus, decreasing its coherence. D_{int} is calculated as the weighted sum ($w_0 = 0.4, w_1 = 0.2, w_2 = 0.4$) of the normalized cumulative quality of **APDs**, **PPs**, and **PTs** within an NG (eq. 14).

$$D_{int} = w_0 \times \frac{APD_q}{\#APD} + w_1 \times \frac{\#PP_q}{\#PP} + w_2 \times \frac{PT_q}{\#PT} \quad (14)$$

4.1 Experiments

We conducted a series of experiments to evaluate and analyze how the system could evolve NGs into quality-diverse and valid narrative structures. We evolved the three manually constructed narrative
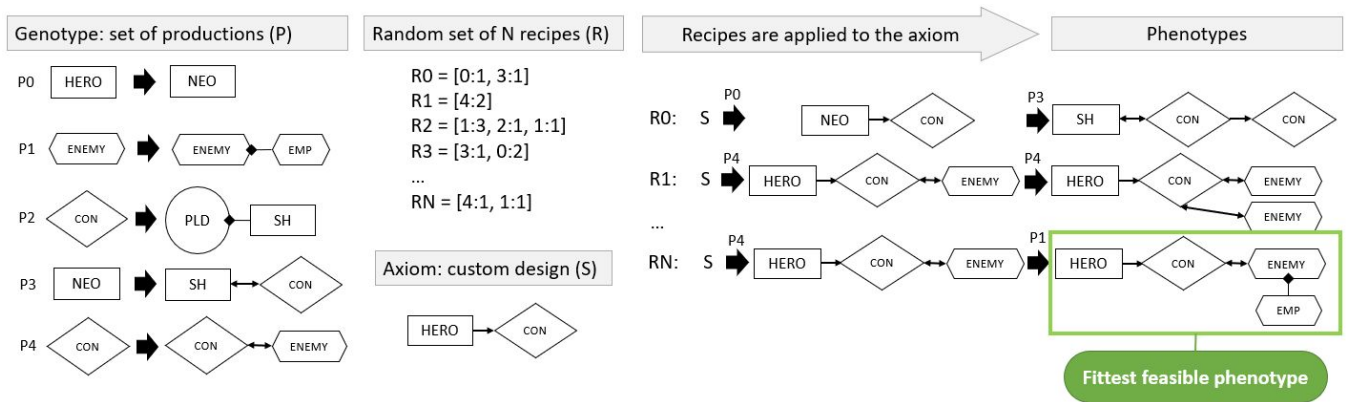
Figure 2: sample complete process from an individual's genotype to the phenotype.

Table 2: Comparative results between root graphs and generated elites (shown in fig. 1)

Graph	Cohesion	Consistency	Coherence (fitness)	Interestingness
RG (fig 1.a1)	1.0	0.66	0.825	0.61
Elite (fig 1.a2)	1.0	0.76	0.875	0.73
RG (fig 1.b1)	1.0	0.75	0.87	0.38
Elite (fig 1.b2)	1.0	0.91	0.95	0.55
RG (fig 1.c1)	1.0	0.77	0.88	0.4
Elite (fig 1.c2)	1.0	0.85	0.92	0.52

graphs shown in figure 1, top row. They were used as root graphs and axioms in the EA, and we used *interestingness* and *step* as behavioral dimensions. We did 5 MAP-Elites runs per narrative graph, ran each for 500 generations, and set the initial population to 1000 randomly created individuals. The initial population is generated by randomly creating between two and five production rules. Each feasible and infeasible population per cell has 25 individuals. Each individual is limited to test 10 recipes regardless of the chromosome size. Offspring were produced either by selecting either the left-side or right-side of a random production rule and exchanging them or with a 50% mutation chance. If an offspring was generated by mutation, there was a 10% chance to add or remove a production rule and a 90% to modify in various ways existing production rules.

We calculated the *coverage*: how much of the constrained search space is explored (i.e., constrained by the behavioral dimensions); the avg. fitness and the avg. interestingness. All experiments had little variation regarding these metrics, and got in avg. 23.5% coverage (24.9%, 21.4%, and 24.2%, respectively), 0.79 fitness (0.76, 0.8, 0.8, respectively), and 0.37 interestingness (0.39, 0.37, 0.36, respectively). These results exemplify both the hard task of generating narrative graphs and exploring the possibility space, and the seemingly competing qualities of coherence (i.e., fitness) and interestingness.

Furthermore, in figure 1, bottom row, it is shown three different example elite narrative graphs, generated from their respective root graphs on the top row and with each individual evaluation shown in table 2. The root graphs have a cohesion of 1.0 since none of them have unused nodes or connections and have similar mid-high consistency values because of using generic nodes (e.g., HERO or ENEMY), repeating them, and low involvement in structures by

characters. In the case of fig 1.a1, the **RevP** from HERO to SH creates some fake conflicts, which affect the consistency but also boost the interestingness value of the narrative graph. Both fig 1.b1 and 1.c1, are evaluated similarly with low interestingness; c1 involves a simplistic and linear structure, and b1, while in principle more complex, is also a relatively linear structure with no **PTs**.

Furthermore, all the exemplar elites have better *consistency*, *coherence*, and *interestingness* than the respective root graph. In figure 1.a2, the graph has been reduced towards a bottleneck, **RevP** (HERO → SH) is removed, and MCG is added as the objective for SH, which could point towards competition or cooperation to enable NEO. Such a change gives more *consistency* to the graph while seemingly reducing its *interestingness*, but this relation and the ◇− connection between MCG and NEO increase its *interestingness*. In figure 1.b2, the narrative has more interaction between **Plot Devices**, and the BAD has a more active role. Particularly, the fact that now HERO → MCG ◇− BAD and MHQ ◇− CHK ◇− BAD could enable and force the HERO towards two main objectives before overcoming the boss, which is reflected in the higher *Interestingness*. Finally, in figure 1.c2, the narrative did not change much (only four steps away), yet the graph is seemingly better, and the narrative could be very different. The graph has broken the loop which connected DRAKE ◇− BAD, and could could point towards a side objective. Further, the connection between BAD and MCG has been reversed; thus, the HERO does not need to face the BAD to get the MCG, rather reaching the MCG will have as a consequence the emergence of the BAD. Finally, BAD is no longer connected to EMP and DRAKE; thus, BAD could be its own enemy faction, in this case, complexifying the narrative and creating more challenge.

5 DISCUSSION AND LIMITATIONS

The trope-graph representation in TropeTwist allows for a quick definition of narrative structures. They are, by design, ambiguous, do not encode temporal information besides causal chains, and are, to some extent, generic, which makes structures relatively simple to develop but more complex to interpret. These design decisions make the system encode less rich information than others, such as Scheherazade [19], but allow the structure to be interpreted in

multiple ways. For instance, the generated graphs could equally describe different stories, and the interpretation given in this paper is just one of many. Thus, the system effectively shifts the complexity from the structure to the "interpreter." While the generated structures could already serve as inspiration for users, an interpreter could provide alternative interpretations that could be guided by or learned from users, which is part of our future work.

Furthermore, the metrics proposed and developed here were used to tune and evaluate the graph outputs without humans in the loop. However, they do not stand in or replace human judgment. The metrics are estimated heuristics mainly based on the graph functionality and relation among patterns. Most of them are related to a "root graph," which is a preliminary step for making TropeTwist interactive and have humans-in-the-loop. We aim to develop a mixed-initiative version of TropeTwist, where metrics depend on the designer's creation. This would, in turn, allow the designer to steer the MAP-Elites search, generating content adapted to them [6], and for MAP-Elites to assist designers with ideation proposing varied structures.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we have presented *TropeTwist*, a system that interconnects tropes and trope patterns to describe narrative structures. We demonstrated through three proof-of-concept structures the system's expressiveness to describe games with diverse genres and mechanics, and different game phases. Further, we illustrated how we could generate novel structures from the three proof-of-concept structures using MAP-Elites, improving them on our metrics.

Tropes could be seem as something to avoid when exploring creativity, mainly due to the possibility of showing unoriginal views by definition. However, a set of combined tropes, patterns, and structures could give rise to novel combinations that express the wanted structure. Similarly, identifying, visualizing, and defining the tropes and patterns and doing "twists" with them; thus, transforming something typical into atypical is the goal with TropeTwist.

The narrative structures show essential aspects of how the story will develop and lead, and important components such as events, conflicts, or roles. However, to further operationalize these structures, it is necessary other systems that make use of them, such as quest [7, 8] or plot [9] generators. Another interesting future work would be to explore the multi-faceted nature of games [36] and combine this type of system with generators that focus on other facets such as level design [4, 49] or game mechanics [14, 23].

Generating novel narrative structures resulted in interesting variations, but the system could not exploit all the advantages of MAP-Elites. Our results point towards difficulties exploring the space, possibly because *coherence* and *interestingness* are to some extent competing objectives. Therefore, we aim at extending TropeTwist towards a mixed-initiative co-creative system [60], and with that, evaluate with human participants. Given that our metrics are dependant on the designed graph; then, we could constantly adapt the content generation and have adaptive models, for instance, of interestingness, based on the user's creation similar to [4, 39].

REFERENCES

[1] Espen Aarseth. 2005. From hunt the wumpus to everquest: introduction to quest theory. In *International Conference on Entertainment Computing*. Springer, 496–506.

[2] Espen Aarseth. 2012. A Narrative Theory of Games. In *Proceedings of the International Conference on the Foundations of Digital Games* (Raleigh, North Carolina) *(FDG '12)*. ACM, New York, NY, USA, 129–133. https://doi.org/10.1145/2282338.2282365

[3] Espen Aarseth and Paweł Grabarczyk. 2018. An Ontological Meta-Model for Game Research. In *DiGRA &#3918 - Proceedings of the 2018 DiGRA International Conference: The Game is the Message.* DiGRA. http://www.digra.org/wp-content/uploads/digital-library/DIGRA_2018_paper_247_rev.pdf

[4] Alberto Alvarez, Steve Dahlskog, Jose Font, and Julian Togelius. 2019. Empowering Quality Diversity in Dungeon Design with Interactive Constrained MAP-Elites. In *2019 IEEE Conference on Games (CoG)*.

[5] Alberto Alvarez, Steve Dahlskog, Jose Font, and Julian Togelius. 2020. Interactive Constrained MAP-Elites: Analysis and Evaluation of the Expressiveness of the Feature Dimensions. *IEEE Transactions on Games* (2020).

[6] Alberto Alvarez, Jose Font, Steve Dahlskog, and Julian Togelius. 2021. Assessing the Effects of Interacting with MAP-Elites. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 17. AAAI, 124–131. https://ojs.aaai.org/index.php/AIIDE/article/view/18899

[7] Alberto Alvarez, Eric Grevillius, Elin Olsson, and Jose Font. 2021. Questgram [Qg]: Toward a Mixed-Initiative Quest Generation Tool. In *Proceedings of the 16th International Conference on the Foundations of Digital Games (FDG'21)*. Association for Computing Machinery, New York, NY, USA.

[8] Prithviraj Ammanabrolu, William Broniec, Alex Mueller, Jeremy Paul, and Mark Riedl. 2019. Toward Automated Quest Generation in Text-Adventure Games. In *Proceedings of the 4th Workshop on Computational Creativity in Language Generation*. Association for Computational Linguistics, Tokyo, Japan, 1–12. https://www.aclweb.org/anthology/2019.ccnlg-1.1

[9] Prithviraj Ammanabrolu, Ethan Tien, Wesley Cheung, Zhaochen Luo, William Ma, Lara J. Martin, and Mark O. Riedl. 2020. Story Realization: Expanding Plot Events into Sentences. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 7375–7382. https://aaai.org/ojs/index.php/AAAI/article/view/6232

[10] Alok Baikadi and Rogelio E. Cardona-Rivera. 2012. Towards finding the fundamental unit of narrative: A Proposal for the Narreme. In *Proceedings of the Third Workshop on Computational Models of Narrative.* European Language Resource Association, 42–44. https://doi.org/10.1.1.487.7222

[11] Roland Barthes. 1975. An Introduction to the Structural Analysis of Narrative. *New Literary History* 6, 2 (1975), 237–272. http://www.jstor.org/stable/468419 Translated by Lionel Duisit.

[12] Jonathan Blow. 2004. Game Development: Harder Than You Think. *Queue* 1 (02 2004), 28–37. https://doi.org/10.1145/971564.971590

[13] Louis Castricato, Spencer Frazier, Jonathan C. Balloch, and Mark Riedl. 2021. Fabula Entropy Indexing: Objective Measures of Story Coherence. *CoRR* abs/2104.07472 (2021). arXiv:2104.07472 https://arxiv.org/abs/2104.07472

[14] Megan Charity, Michael Cerny Green, Ahmed Khalifa, and Julian Togelius. 2020. Mech-Elites: Illuminating the Mechanic Space of GVG-AI. In *International Conference on the Foundations of Digital Games* (Bugibba, Malta) *(FDG '20)*. Association for Computing Machinery, New York, NY, USA, Article 8, 10 pages. https://doi.org/10.1145/3402942.3402954

[15] Megan Charity, Ahmed Khalifa, and Julian Togelius. 2020. Baba is Y'all: Collaborative Mixed-Initiative Level Design. arXiv:2003.14294 [cs.HC]

[16] Jonathon Doran and Ian Parberry. 2011. A Prototype Quest Generator Based on a Structural Analysis of Quests from Four MMORPGs. In *Proceedings of the 2nd International Workshop on Procedural Content Generation in Games* (Bordeaux, France) *(PCGames '11)*. ACM, New York, NY, USA, Article 1, 8 pages. https://doi.org/10.1145/2000919.2000920

[17] Joris Dormans and Sander Bakkes. 2011. Generating missions and spaces for adaptable play experiences. *IEEE Transactions on Computational Intelligence and AI in Games* 3, 3 (2011), 216–228.

[18] Mirjam P Eladhari, Philip L. Lopes, and Georgios N. Yannakakis. 2014. Interweaving Story Coherence and Player Creativity through Story-Making Games. In *Interactive Storytelling : 7th International Conference on Interactive Digital Storytelling, ICIDS 2014, Singapore, Singapore, November 3-6, 2014, Proceedings (Lecture Notes in Computer Science, 8832)*. 73–80. https://doi.org/10.1007/978-3-319-12337-0_7

[19] David Elson. 2012. DramaBank: Annotating Agency in Narrative Discourse. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. European Language Resources Association (ELRA), Istanbul, Turkey, 2813–2819. http://www.lrec-conf.org/proceedings/lrec2012/pdf/866_Paper.pdf

[20] Jacob Garbe, Max Kreminski, Ben Samuel, Noah Wardrip-Fruin, and Michael Mateas. 2019. StoryAssembler: An Engine for Generating Dynamic Choice-Driven Narratives. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (San Luis Obispo, California, USA) *(FDG '19)*. Association for Computing Machinery, New York, NY, USA, Article 24, 10 pages. https://doi.org/10.1145/3337722.3337732

[21] Pablo García-Sánchez, Antonio Velez-Estevez, Juan Julián Merelo, and Manuel Jesús Cobo. 2021. The Simpsons did it: Exploring the film trope space and its large scale structure. *PLOS ONE* 16, 3 (03 2021), 1–28.

[22] Pablo Gervas. 2009. Computational Approaches to Storytelling and Creativity. *AI Magazine* 30, 3 (Jul. 2009), 49. https://doi.org/10.1609/aimag.v30i3.2250

[23] Michael Cerny Green, Ahmed Khalifa, Philip Bontrager, Rodrigo Canaan, and Julian Togelius. 2021. Game Mechanic Alignment Theory and Discovery. arXiv:2102.10247 [cs.AI]

[24] Algirdas Julien Greimas. 1984. *Structural Semantics: An Attempt at A Method.* University of Nebraska Press.

[25] Charlie Hargood, David Millard, and Mark Weal. 2011. Measuring Narrative Cohesion: A Five Variables Approach. In *Narrative and Hypertext Workshop at Hypertext 11 (05/06/11).* https://eprints.soton.ac.uk/272275/

[26] James Harris. 2016. Periodic Table of Storytelling. https://jamesharris.design/periodic/. Accessed: 2020-03-24.

[27] Ken Hartsook, Alexander Zook, Sauvik Das, and Mark O. Riedl. 2011. Toward supporting stories with procedurally generated game worlds. In *2011 IEEE Conference on Computational Intelligence and Games, CIG 2011, Seoul, South Korea, August 31 - September 3, 2011,* Sung-Bae Cho, Simon M. Lucas, and Philip Hingston (Eds.). IEEE, 297–304. https://doi.org/10.1109/CIG.2011.6032020

[28] I. Horswill. 2016. Dear Leader's Happy Story Time: A Party Game Based on Automated Story Generation. In *Proceedings of the EXAG Workshop at AIIDE.*

[29] Ahmed Khalifa, Scott Lee, Andy Nealen, and Julian Togelius. 2018. Talakat: Bullet hell generation through constrained map-elites. In *Proceedings of The Genetic and Evolutionary Computation Conference.* ACM.

[30] Max Kreminski, Melanie Dickinson, and Noah Wardrip-Fruin. 2019. Felt: A Simple Story Sifter. In *Interactive Storytelling,* Rogelio E. Cardona-Rivera, Anne Sullivan, and R. Michael Young (Eds.). Springer International Publishing, Cham, 267–281.

[31] Max Kreminski and Noah Wardrip-Fruin. 2018. Sketching a Map of the Storylets Design Space. In *Interactive Storytelling,* Rebecca Rouse, Hartmut Koenitz, and Mads Haahr (Eds.). Springer International Publishing, Cham, 160–164.

[32] Quinn Kybartas and Rafael Bidarra. 2016. A survey on story generation techniques for authoring computational narratives. *IEEE Transactions on Computational Intelligence and AI in Games* 9, 3 (2016), 239–253.

[33] Petri Lankoski. 2013. Models for Story Consistency and Interestingness in Single-Player RPGs. In *Proceedings of International Conference on Making Sense of Converging Media* (Tampere, Finland) *(AcademicMindTrek '13).* Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/2523429.2523480

[34] V. I. Levenshtein. 1966. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady* 10 (Feb. 1966), 707.

[35] Boyang Li and Mark O. Riedl. 2010. An offline planning approach to game plotline adaptation. In *Proc. AIIDE'10.*

[36] A. Liapis, G. N. Yannakakis, M. J. Nelson, M. Preuss, and R. Bidarra. 2019. Orchestrating Game Generation. *IEEE Transactions on Games* 11, 1 (2019), 48–68.

[37] Antonios Liapis, Georgios N. Yannakakis, and Julian Togelius. 2014. Computational Game Creativity. In *Proceedings of the fifth International Conference on Computational Creativity (ICCC 2014),* Hannu Toivonen, Simon Colton, Michael Cook, and Dan Ventura (Eds.).

[38] Michael Mateas and Andrew Stern. 2003. Integrating plot, character and natural language processing in the interactive drama Façade. In *Proceedings of the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE-03),* Vol. 2.

[39] Panagiotis Migkotzidis and Antonios Liapis. 2021. SuSketch: Surrogate Models of Gameplay as a Design Assistant. *IEEE Transactions on Games* (2021), 1–1. https://doi.org/10.1109/TG.2021.3068360

[40] Jean-Baptiste Mouret and Jeff Clune. 2015. Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909* (2015).

[41] Nintendo R&D1 and Intelligent Systems. 1985. *Super Mario Bros.* Game [SNES]. Nintendo, Kyoto, Japan. Last played August 2020..

[42] Nintendo R&D4. 1991. *The Legend of Zelda: A Link to the Past.* Game [SNES]. Nintendo, Kyoto, Japan. Last played December 2011..

[43] Nintendo R&D4. 1998. *The Legend of Zelda: Ocarina of Time.* Game [N64]. Nintendo, Kyoto, Japan. Last played December 2011..

[44] R. Y. Pérez and Otoniel Ortiz. 2013. A Model for Evaluating Interestingness in a Computer-Generated Plot. In *Proceedings of the fourth International Conference on Computational Creativity (ICCC 2013).*

[45] V. Propp, L. Scott, and L.A. Wagner. 1975. *Morphology of the Folktale: Second Edition.* University of Texas Press. https://books.google.se/books?id=cyc7AQAAIAAJ

[46] Chris Richmond and Drew Schoentrup. 2004. TV Tropes. https://tvtropes.org/pmwiki/pmwiki.php/Main/Tropes. Accessed: 2020-03-24.

[47] Mark O. Riedl and R. Michael Young. 2006. Story planning as exploratory creativity: Techniques for expanding the narrative search space. *New Generation Computing* 24, 3 (01 Sep 2006), 303–323. https://doi.org/10.1007/BF03037337

[48] Jonathan Rowe, Scott McQuiggan, Jennifer Sabourin, Derrick Marcey, and James Lester. 2009. STORYEVAL: An empirical evaluation framework for narrative generation. *AAAI Spring Symposium - Technical Report,* 103–110.

[49] Anurag Sarkar and Seth Cooper. 2021. Dungeon and Platformer Level Blending and Generation using Conditional VAEs. In *Proceedings of the IEEE Conference on Games (CoG).*

[50] Manu Sharma, Santiago Ontañón, Christina R. Strong, Manish Mehta, and Ashwin Ram. 2007. Towards Player Preference Modeling for Drama Management in Interactive Stories. In *Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference, May 7-9, 2007, Key West, Florida, USA,* David Wilson and Geoff Sutcliffe (Eds.). AAAI Press, 571–576. http://www.aaai.org/Library/FLAIRS/2007/flairs07-112.php

[51] Nicolas Szilas. 2002. Structural models for Interactive Drama. In *Proceedings of the 2nd International Conference on Computational Semiotics for Games and New Media.* https://ci.nii.ac.jp/naid/10026187402/en/

[52] Nicolas Szilas, Sergio Estupiñán, and Urs Richle. 2016. Qualifying and Quantifying Interestingness in Dramatic Situations. In *Interactive Storytelling,* Frank Nack and Andrew S. Gordon (Eds.). Springer International Publishing, Cham, 336–347.

[53] Pradyumna Tambwekar, Murtaza Dhuliawala, Lara J. Martin, Animesh Mehta, Brent Harrison, and Mark O. Riedl. 2019. Controllable Neural Story Plot Generation via Reward Shaping. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019,* Sarit Kraus (Ed.). ijcai.org, 5982–5988. https://doi.org/10.24963/ijcai.2019/829

[54] Matt Thompson, Julian Padget, and Steve Battle. 2018. Governing Narrative Events with Tropes as Institutional Norms. In *Artificial Life and Intelligent Agents,* Peter R. Lewis, Christopher J. Headleand, Steve Battle, and Panagiotis D. Ritsos (Eds.). Springer International Publishing, Cham, 133–137.

[55] J. Togelius, G. N. Yannakakis, K. O. Stanley, and C. Browne. 2011. Search-Based Procedural Content Generation: A Taxonomy and Survey. *IEEE Transactions on Computational Intelligence and AI in Games* 3, 3 (Sept. 2011), 172–186. https://doi.org/10.1109/TCIAIG.2011.2148116

[56] Simon Tolinsson, Alexander Flodhag, Alberto Alvarez, and Jose Font. 2020. To Make Sense of Procedurally Generated Dungeons. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) *(CHI PLAY '20).* Association for Computing Machinery, New York, NY, USA, 384–387. https://doi.org/10.1145/3383668.3419890

[57] Marcus Trenton, Duane Szafron, Josh Friesen, and Curtis Onuczko. 2010. Quest Patterns for Story-Based Computer Games. In *Proceedings of the Sixth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Stanford, California, USA) *(AIIDE'10).* AAAI Press, 204–209.

[58] Judith van Stegeren and Jakub Myśliwiec. 2021. Fine-Tuning GPT-2 on Annotated RPG Quests for NPC Dialogue Generation. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021* (Montreal, QC, Canada) *(FDG'21).* Association for Computing Machinery, New York, NY, USA, Article 2, 8 pages. https://doi.org/10.1145/3472538.3472595

[59] Miruna Vozaru. 2022. *The Game Situation: An object-based game analysis framework.* PhD Thesis. IT University of Copenhagen. Published: PhD Dissertation, IT University of Copenhagen.

[60] Georgios N. Yannakakis, Antonios Liapis, and Constantine Alexopoulos. 2014. Mixed-Initiative Co-Creativity. In *Proceedings of the 9th Conference on the Foundations of Digital Games.*

[61] R. Michael Young, Stephen G. Ware, Bradly A. Cassell, and Justus Robertson. 2013. Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative* 37, 1-2 (2013), 41–64.

[62] Kristen Yu, Nathan R. Sturtevant, and Matthew Guzdial. 2020. What is a Quest?. In *Proceedings of the Intelligent Narrative Technolgies Workshop at AIIDE.* http://www.cs.ualberta.ca/~nathanst/papers/yu2020quest.pdf