Mixed-initiative story co-creation with TaleMaker

Mijael R. Bueno Perez
M.R.BuenoPerez@tudelft.nl
Delft University of Technology
The Netherlands

Rafael Bidarra
R.Bidarra@tudelft.nl
Delft University of Technology
The Netherlands

ABSTRACT

Most storytelling games bring people together to co-create stories. However, they often require considerable creative effort and skills from all players, possibly discouraging less resourceful participants and impairing stories' quality. Moreover, most stories created within these games are usually only kept in players' minds rather than on storage, despite being a valuable and original asset, with a large potential for the narrative research community. We address these challenges with a novel mixed-initiative approach aimed at supporting a group of players to incrementally co-create a story, one sentence at a time. Our method features a *hand generator* that offers a unique set of tokens (words) to each player in each turn. This generator carefully combines tokens relevant to the *ongoing story*, to each *individual player*, to the *group as a whole* and *random*. We implemented this method in *TaleMaker*, a multiplayer online game that stimulates playful co-creation of a story. TaleMaker gives players considerable creative freedom to compose their sentences, combining a gentle structural steering with the wisdom of the group to determine the best direction for the story. The collected output of TaleMaker consists of annotated stories, with slots (e.g. action, character, location) filled with words associated with a WordNet synset. From a preliminary evaluation, players reported that TaleMaker effectively stimulated story authoring, and perceived TaleMaker-created stories of considerable quality. In addition, a first analysis of the collected tokens confirms that players mostly collected story-related tokens, rather than those randomly offered.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning algorithms**; • **Human-centered computing** → **Collaborative and social computing theory, concepts and paradigms**.

KEYWORDS

story co-creation, mixed-initiative, collaborative storytelling, storytelling game, hand generation, synset vector space

ACM Reference Format:
Mijael R. Bueno Perez and Rafael Bidarra. 2022. Mixed-initiative story co-creation with TaleMaker. In *FDG '22: Proceedings of the 17th International Conference on the Foundations of Digital Games (FDG '22), September 5–8, 2022, Athens, Greece.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3555858.3555876

FDG '22, September 5–8, 2022, Athens, Greece

ACM ISBN 978-1-4503-9795-7/22/09.
https://doi.org/10.1145/3555858.3555876

1 INTRODUCTION

Storytelling games are a fun activity that allows people to co-create a story together. They come in different forms like role-playing games, board games, improvisation games, digital games, and many others. Usually, after introducing a few characters, locations, situations, and/or objects, players iteratively form creative associations among them, to unfold the story. However, this poses non-trivial creative challenges which, if not properly supported, may hinder and discourage less skilled players, as well as impair the quality of the story. In addition, most of these stories are, at most, kept in each player's mind, not on storage, and over time they simply get lost. Yet, such stories are a valuable product of human creativity, and we argue that keeping them is important from both a cultural and a research perspective.

To address these challenges, we designed a novel mixed-initiative approach aimed at supporting a group of players to incrementally co-create a story, one sentence at a time. We implemented this approach in *TaleMaker*, a collaborative multiplayer online game that stimulates playful co-creation of open-ended stories [4]. TaleMaker features two essential mechanics: a *hand generator*, providing players with a unique set of tokens (words) to compose their sentence with; and a *submission and voting system*, that promotes collective participation and story steering. This mixed-initiative approach gives players creative control over the story while providing them with subtle guidance. Tokens are offered by a *hand generation method*, which aims at maintaining a balance between assisting players' creativity and ensuring story quality. Each hand generated combines tokens that previously appeared in the story, with tokens related to the story, relevant to the player and to the group, as well as random. We derive related and relevant tokens based on novel and common contextual associations among characters, actions, locations, and objects, as encoded in an embedding of synset vectors [5]. Moreover, rather than controlling the story structure, TaleMaker allows the group to decide on the next direction for the story.

Fortunately, in TaleMaker stories are not forgotten, and every story is fully annotated and incorporated into a database of stories [3]. In this paper, we dive deeper into the design choices and the technical aspects of our approach demonstrated with TaleMaker. In addition, we report the preliminary results of an experiment in which we examined player choices of tokens, evaluated a few aspects of the game, and assessed the quality of stories.

2 RELATED WORK

In this section, we focus on examining several storytelling games, including some that incorporate mixed-initiative aspects in their design. A broader survey on story generation techniques can be found elsewhere [15].

2.1 Storytelling Games

In analog storytelling games, the use of cards or tokens engraved with words or phrases seems to be a recurring theme that inspires people to tell stories. For example, *Once upon a time* is a collaborative and competitive card game in which players are given a hand of cards and tell a story based on them [9]. These cards include words that appear frequently in fairy tales. In addition, players receive an ending card. To win the game, players are required to use the majority of their cards and drive the story to the conclusion of their own ending card. Players accomplish this by interfering and completing each other's contributions to the story.

Other analog storytelling games employ voting rounds as a common pattern. *Dixit* is a storytelling game that uses an illustrated deck of cards [22]. In each turn, a player (the teller) is selected to tell a sentence inspired by a card in their hand. Then, other players give their most similar card to the sentence of the teller. The teller shuffles the cards and everybody has to vote for the original card that inspired the teller. *Tall Tales* is a kickstarter project for a competitive story-writing game [19]. Players must write a short scene for the tale that the group is making together. They can use cards for creative inspiration and guidance. Afterwards, players vote for a scene to be included in the tale, and they are then rewarded according to the results.

Cards Against Humanity is less of a storytelling game and more of a party game for adults [10]. Nevertheless, since it has simple rules, it could be turned into a storytelling game, if the phrases could be extended between turns. In this game, players take turns filling in the blanks of pre-made phrases (black cards) with other words or phrases (white cards). The player who comes up with the funniest or most despicable phrase wins the turn. The game continues as long as players desire.

Several digital storytelling games draw upon the elements of analog games. As an example, *The Newborn World* is a co-located, competitive and collaborative game that follows analog game patterns to encourage players to tell stories [16]. The game is played on a single mobile device shared by all players. The goal of the game is to create a complete story about the First City. This can be accomplished by filling out event cards containing phrases that have two gaps. Each player must complete the event by selecting words from a common pool of words. Then, players vote on other events, and the event with the most votes becomes part of the story.

Similarly, *4Scribes* is also a storytelling game which employs a single device among few players [7]. In this game, players take turns to add to a story, with the objective of steering the story towards their own secret ending. Players are given creative direction through *Creative Elements* cards, which serve as seeds to stimulate their creativity. In each turn, players select one of these cards, and when adding to a story, they must take the card's phrase or word into consideration. In the end, the player with the ending that makes more sense for the story is chosen as the winner.

2.2 Mixed-Initiative Co-Creators of Stories

Say Anything was designed as a mixed-initiative system, in which a user and the computer alternate adding sentences to a story [24]. As a user writes a sentence, the system searches an extensive corpus of stories for a similar sentence and returns the next sentence based on that. This principle is known as case-based reasoning, which draws inferences about a new instance by comparing it to one previously observed [25].

Creative Help works on the same principle as *Say Anything*, however, users have the option of modifying or removing suggestions within the story as they see fit [21]. These tools inspired *Writing Buddy*, a playful, mixed-initiative tool that integrates design principles of casual creators [6], and uses a social simulation engine called *Ensemble* to help authors create story beats and actions for their stories [23].

TaleBox is a game prototype to help players collaboratively create stories [18]. In each turn, a hand generation system provides players with a collection of words that players can use to form their story. TaleBox relies on *GluNet*, a knowledge base that integrates a variety of lexical databases to derive semantic relations for the hand generator [14].

Why Are We Like This? is a digital storytelling game that uses AI as a mixed-initiative component [12]. In this game, two or more players co-create a story. It is the goal of the AI to support players' storytelling practices by providing intelligent plot direction suggestions derived from a social simulation engine called *Felt* [13].

TaleForge [5] is a mixed-initiative system that given a story enables guided exploration of entities for a narrative world, defined as an environment which supports enacting a given story [1]. The system can be used as inspiration for stories or to plan narrative worlds for games or films. Each entity in TaleForge can be selected in order to obtain other associated entities, which in turn can also be selected to obtain further related entities. These entities are derived by using an embedding of synset vectors that encode common and novel associations among characters, locations, objects, and actions within a large dataset of photo captions. TaleMaker uses the same embedding of synset vectors to compute tokens relevant for players and for the ongoing story.

3 TALEMAKER

The TaleMaker game demonstrates how our mixed-initiative approach allows players to be creative in authoring sentences for a story while receiving subtle creative guidance. Instead of strictly controlling the story structure, TaleMaker relies on the creativity of players and the wisdom of the group to determine the best direction for the story.

3.1 Mixed-Initiative Approach

The mixed-initiative character of our approach lies in providing a selection of tokens (words) to be used by the players. As they create their sentences using these tokens, we identify additional tokens that could be beneficial to them in composing subsequent sentences for the story. Therefore, rather than offering only random tokens, we propose a hand generation method that uses the current story and relevancy information gathered from past turns, in order to offer tokens that might be of interest for the player and serve the story. This method seeks to strike a balance between creative assistance, story quality and the competitive nature of the game. Hence, players are always encouraged to make a creative effort when planning ahead their sentences and composing each sentence based on the tokens they have collected.

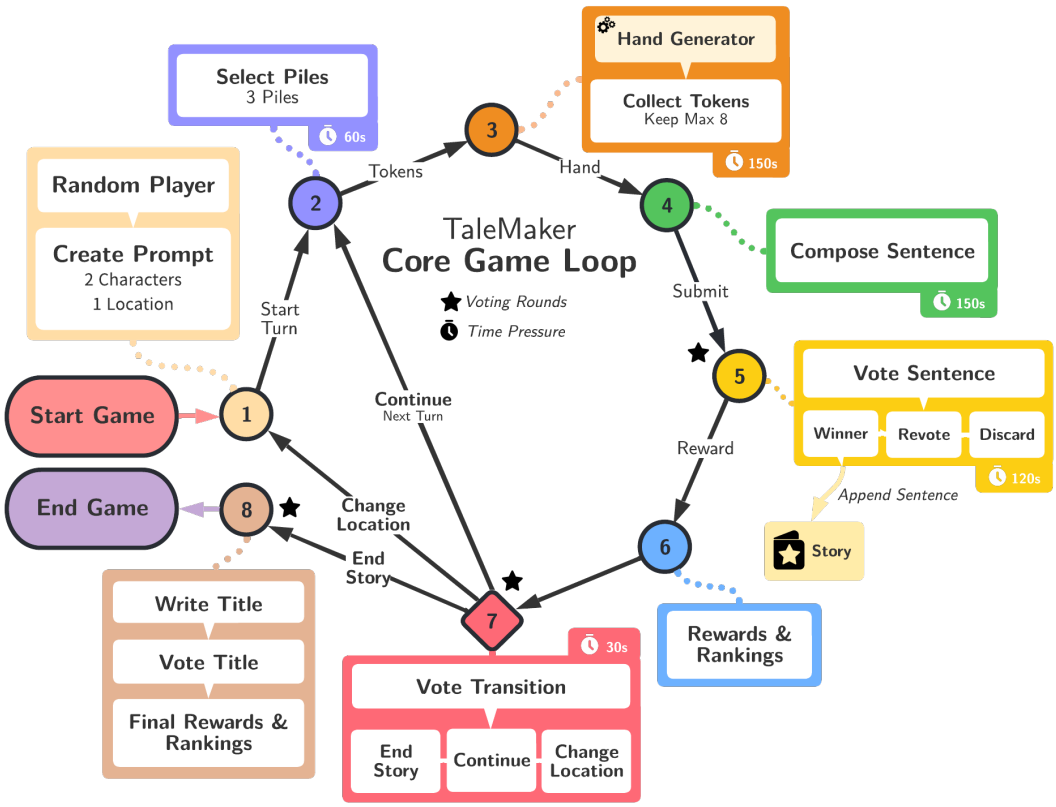

Figure 1: TaleMaker's Core Game Loop.

3.2 Game Design

TaleMaker features both competitive and collaborative elements, designed to provide a playful experience while co-creating stories. In each turn, each player collects tokens and uses them to compose a sentence that is compelling enough to convince other players that it belongs in the story. Then, all sentences are submitted for a voting round, where the most popular sentence is appended to the story and its author is rewarded. In most steps, players are under time pressure, to keep the tempo, add to the competitive gameplay, and avoid long waiting times. The game concludes when the majority of players decide to end the story, and its title is then chosen.

TaleMaker is played online on a computer or a mobile device. In order to play, players need to create a game room which can accommodate two to eight players. No limitations exist as to whether people choose to play co-located or remotely. Moreover, players will always be able to join and leave the room at any time, although the room remains open until the story is completed or after a prudent period of idle time has passed. Any player may join a public game room during an ongoing story. This facilitates the flow of new ideas into the story. In addition, one can also create a private game room, accessible only with a secret code.

3.3 Core Game Loop

The core game loop of TaleMaker consists of 8 basic phases, as shown in Figure 1. These phases involve (1) creating a story prompt, (2) selecting piles (categories) of tokens (words), (3) collecting and/or discarding tokens, (4) composing a sentence, (5) submitting and voting for the winner sentence, (6) receiving rewards and rankings, (7) voting whether the story should end, change location or continue, and finally (8) writing and voting for a title. In this subsection, we explain in detail each of these phases.

3.3.1 Story Prompt Phase. At the beginning of the game, a player is randomly selected to create two characters and a location. Since there is no story at this point, this player has greater creative freedom and responsibility, as after this, players have to develop the story centered on these characters and location. Once the players have added a certain number of sentences at one particular location, they can decide whether to change location, by which another player is selected to create a new prompt, thus continuing the story in a different location, with two more characters.

As illustrated in Figure 2a, a player creates characters by first searching for a particular type of character, for example, a *pirate*. In response to this search query, the system retrieves various tokens that can be inserted into the character slots. The player is then asked to chose and name their characters, and a similar process is repeated for a location. After having completed the prompt, all

(a) phase 1 (b) phase 2 (c) phase 3

(d) phase 4 (e) phase 5 (f) phase 6

Figure 2: The various screens of TaleMaker in which the player (a) creates a story prompt (characters and location), (b) selects piles of tokens, (c) collects tokens offered by the hand generation method, (d) composes a sentence, (e) votes for a sentence, and (f) reads the ongoing story after being rewarded.

players are informed of these two new characters and their location, and move on to the next phase.

3.3.2 Pile Selection Phase. This phase aims at allowing players to choose the type of tokens they wish to receive, a choice that might be influenced by their view on the current story. For this, tokens are classified into well-categorized groups called *piles*. Players begin their turn by selecting three piles (categories) of tokens (see Figure 2b). The pile of verbs is selected by default, in order to ensure that players do not run out of new actions for their sentences. Thus, players receive tokens from a total of four piles.

3.3.3 Token Collection Phase. In this phase, players receive a collection of tokens drawn from their chosen piles and the pile of verbs (see Figure 2c). Here the creative guidance comes into play, as the hand generator determines the tokens that will be *served* to each player, based on the piles they selected. Yet, it is the responsibility of each player to decide which served tokens to keep, and which previously kept tokens to discard, based on their own authoring plans. Players are given a limited amount of token slots for their *collected hand*, from which they draw tokens to plan and create their sentences. Currently, TaleMaker's collected hand allows a maximum of eight tokens so, at some point, players must decide which tokens to discard to make room for new ones.

From the perspective of a player, tokens are simply words. However, internally tokens are associated with a synset in WordNet [17], which allows us to access semantic knowledge from this lexical database. This semantic knowledge helped us to equip a token with a corresponding pile (category or part of speech) indication, description and an illustration (see Figure 3). Players can use this information to ensure that the meaning of a token's word corresponds with what they are trying to convey in their sentences, as well as to better understand the sentences of their peers.

3.3.4 Sentence Composition Phase. In this phase, each player composes a sentence, as their creative contribution to advance the story (see Figure 2d). A sentence consists of a variable number of slots that can be filled with tokens from the collected hand. A slot is a placeholder for a token, and it holds information regarding the type of token it must hold, such as a main *actor*, *character*, *action*,

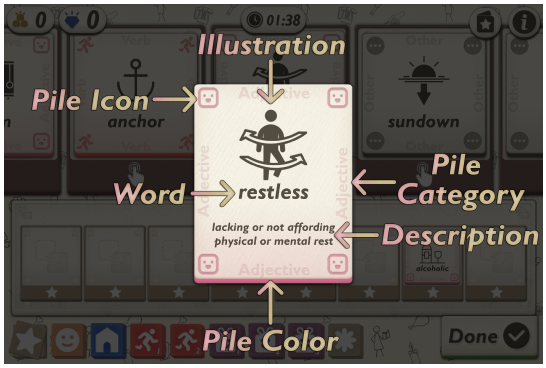

Figure 3: A token contains additional information that can be utilized by players in order to better understand its meaning.

location, and *free*. Each particular slot can only hold certain tokens, for example, an action slot can only hold a verb token.

Initially, players receive a basic sentence template with an actor slot, an action slot, a free slot filled with a preposition 'at' token and a location slot filled with the current location of the story, as shown in Figure 2d. Typically, players place the tokens from their collected hand into the sentence slots. In addition, they may also use tokens from the pile of characters, the location and a few other tokens that are included by default, such as a few basic verbs, prepositions, and others. Whenever desired, slots can be added, rearranged or removed (except the actor and action of the sentence).

To avoid creating a lengthy sentence, we limited the number of slots in a sentence to a maximum of twelve. Moreover, players are limited to one action per sentence, so as to not complicate their sentences or create 'compound sentences', which could likely give them an advantage. In addition, players have to use at least one character. In this way, we encourage players to form a simple sentence in active rather than in passive voice, where the subject performs the action stated by the verb (e.g. Bob eats an apple). Furthermore, sentences are not strictly required to be grammatically correct, but should convey the essence of an action in the story. Finally, once the players have completed their sentences and there are no slots left empty, they can submit their sentences for a voting round.

3.3.5 Sentence Voting Phase. In this phase, players anonymously vote for other submitted sentences; the sentence with most votes is appended to the story and its author is rewarded. As shown in Figure 2e, when more than two sentences are submitted, players cannot vote for their own sentence, only for other sentences. However, to avoid impasses, when there are only two sentences, the authors can vote for their own sentence, after considering whether they deserve their vote or whether the other player has a better sentence for the story. In any case, the total votes of the group will determine the sentence that makes it into the story. In case of tie, one can repeat the voting or, if it persists, possibly discard the whole turn and start a new one.

3.3.6 Rewards and Ranking Phase. The reward system aims at encouraging competition between players as well as motivating them to advance the story, contributing quality sentences as much as possible. In this phase, players receive two types of rewards: gold and diamonds. Players who submit a sentence on time receive one unit of gold as a reward. In addition, the number of votes they received for their sentence is added as one gold per vote. A high number of gold indicates that a player is actively contributing to the story with sentences that are among the most preferred ones. Diamonds are only awarded to players whose sentences are selected by the voting round and appended to the story. A player with a high number of diamonds will, therefore, have a greater sense of ownership over the story.

3.3.7 Transition Voting Phase. After a certain number of sentences within a given location, players are asked whether they wish to change location or end the story. When a majority of players decide to change location, the game will ask another player at random, from among those who have not yet been selected, to create two new characters (added to the pile of characters), as well as the new location to continue the story. This process is repeated a few times throughout the game, until a majority of the players decide to end the story, at which point they are required to choose a title for the story.

3.3.8 Story Title Phase. In this phase, each player reads the story and proposes a title for it. As such, there are no restrictions based on tokens, and players write text directly. Once players have their titles ready, they submit them to a last voting round, where they can vote for someone else's title. The title with the highest number of votes will be displayed above the story, and its author will be rewarded. This is the last opportunity for players to improve their score and obtain a higher ranking. Consequently, the game ends here, and players are free to read the completed story (see Figure 2f).

3.4 Implementation

TaleMaker's client was designed in English and developed with GameMaker Studio 2. The game server was developed in Python from the ground up using the *socket* module as the network interface; the stories and gameplay information are stored in a PostgreSQL database. In addition, we obtained illustrations of tokens from TheNounProject website [20].

4 HAND GENERATION

The hand generation method is at the core of our mixed-initiative approach. In this section, we present a general overview of the method and examine the technical aspects of its implementation.

4.1 Method Overview

The main goal of the hand generation method is to strike a balance between supporting creativity among players and maintaining the competitive nature of the game. For this, in each turn, each player receives a collection of tokens (*the served hand*) from which they will retain a subset (*the collected hand*), to be used when composing a sentence. Tokens in the served hand are sampled from five distinct *bags of ranked tokens*, each of which is designed to promote players' creativity and story quality in different ways. These bags hold an ordered list of tokens that are ranked based on relevance or relatedness to different aspects of the story and the game. The purpose of each bag is summarized as follows:

- *Related Bag*: consists of tokens, ranked by the degree to which they are related to other tokens in the story. This bag allows players to keep the context of the story consistent. For example, an *island* is typically associated to tokens such as *coconut*, *swim*, *sand*, *shipwreck*, etc.
- *History Bag*: consists of only the tokens previously found in the story. This bag enables players to stay on topic. The story might contain objects of narrative importance, for example, a *treasure map*; therefore, it is convenient to offer the token again, so that players can further develop the story using it (e.g. what happens to the treasure map?).
- *Player Bag*: includes tokens ranked according to their predicted relevance to each individual player. This bag is intended to support players' creativity and planning. For instance, a player planning sentences in the context of a love

story, might collect tokens such as *love* and *kiss*; that player might then wish to receive other relevant tokens, such as *hug*, *gift*, *date*, etc.

- *Group Bag*: includes tokens ranked according to their predicted relevance to the whole group. With this bag, players are encouraged to consider the previous contributions of others. For example, former sentences that did not get approved, might still contain tokens that are relevant to the present part of the story.
- *Random Bag*: includes all possible tokens, sampled in a random manner. This bag aims to stimulate players to think creatively about seemingly unrelated tokens, possibly coming up with new associations or story directions.

4.2 Bags of Ranked Tokens

In the sequel, we will use the following notation: T_r denotes the ranked tokens of the Related Bag; T_h, those of the History Bag, T_p, those of the Player Bag, T_g, those of the Group Bag, and T_n, those of the Random Bag. Rankings within each bag are updated in each turn to ensure that we capture the latest relevant information for the current turn. Altogether, these five bags form the set B_T:

$$B_T = \{T_r, T_h, T_p, T_g, T_n\} \tag{1}$$

This setup provides controllable means for sampling tokens from a bag and for randomly selecting which bag to sample a token from. In addition, we denote T as the set of all available tokens in the game. A story is composed of the sequence of sentences S_τ appended after each turn. Each sentence S_τ contains several tokens $S_\tau = \{s_1, s_2, ..., s_n\}$ that were added by a player. The subscript τ in S_τ indicates the turn of the game in which the sentence was chosen to be part of the story. For example, S_0 corresponds to the first sentence, and S_1 to a sentence appended to it. In addition, we treat a story prompt S^{p}_τ as a sentence that only includes the tokens of two characters and the location. However, a story prompt shares the turn with a sentence, for example, the first prompt S^{p}_0 shares the turn with the first sentence S_0 created afterwards. Prompts are mostly used for computing the ranks of the Related Bag following the start of the game or a location change. This addresses the cold start problem of our hand generator, in which we do not have prior information to serve tokens from other bags.

We keep only the tokens of S_τ which are associated to a synset vector $\vec{s}$. For this, we use an embedding of synset vectors obtained from prior work [5]. This embedding encodes common and novel associations between locations, actions, and entities learned from co-occurrence information within a large dataset of photo captions. We use the *cosine relatedness score* to measure the degree of similarity or relatedness between two synset vectors included in the embedding. Thus, the cosine score θ between the synset vectors $\vec{s_i}$ and $\vec{s_j}$ of two tokens is calculated as follows:

$$\theta = \cos(\vec{s_i}, \vec{s_j}) = \frac{\vec{s_i} \cdot \vec{s_j}}{\|\vec{s_i}\|\|\vec{s_j}\|} \tag{2}$$

In the next subsections, we describe the characteristics and technical details involved in computing each of the five ranked bags.

4.2.1 Related Bag. First, we compute the overall cosine scores measuring the relatedness of the most recent sentence S_τ or prompt S^{p}_τ to each synset vector within the entire set of available tokens $\vec{s}_j \in T$:

$$\Theta'_\tau = \{\theta'_j \in \mathbb{R} : \theta'_j = \sum_{\vec{s_i} \in S_\tau | S^{\mathrm{p}}_\tau} \cos(\vec{s_i}, \vec{s_j}),\ \forall \vec{s_j} \in T\} \tag{3}$$

In order to ensure consistency between turns, we min-max normalize the scores Θ'_τ to values within the range from 0 to 1. Following that, we obtain the current scores Θ_τ by considering prior relatedness information. Therefore, we combine the new relatedness scores Θ'_τ by aggregating them with the previous normalized scores $\Theta_{\tau-1}$ as follows:

$$\Theta_\tau = (1 - \lambda)\Theta_{\tau-1} + \Theta'_\tau \tag{4}$$

Once again, we min-max normalize Θ_τ so as to maintain consistency among turns. The *time decaying factor* λ controls the rate at which relatedness information from prior prompts or sentences is forgotten. In our implementation, we empirically chose a factor of $\lambda = 0.2$. This reinforces the importance associated to newly added tokens in the story while still retaining some information from recent turns. Furthermore, for the sentence that follows a location change, we treat $\Theta_{\tau-1}$ as Θ_τ, since the story prompt and the sentence share the same turn. In addition, at the beginning of the game $\Theta_{\tau-1}$ does not exist, thus, we consider $\Theta_0 = \Theta'_0$, which corresponds to the cosine scores of the initial story prompt.

Finally, the Related Bag T_r is obtained by considering only the top k tokens with the highest cosine relatedness score $\theta_j \in \Theta_\tau$. We chose a $k = 100$ in order to allow a considerable variety of related tokens to be sampled and offered to players.

4.2.2 History Bag. This bag contains tokens that were used so far in the story, restricted to a controllable window size of n most recent sentences. After that, we rank tokens based on their respective relatedness scores $\theta_j \in \Theta_\tau$ calculated from Equation 4. Thus, the tokens are still sampled based on their relatedness to the ongoing story. In our experiments, we used a window size n of 10 sentences. However, this parameter value can also be raised, possibly to consider all tokens in the story.

4.2.3 Player Bag. In each turn, players decide which tokens to keep, discard, or use for a sentence. We use this information to iteratively construct the Player Bag T_p, with tokens ranked according to their predicted relevance to the player. For this, we first use a relevance feedback technique to generate new training data each turn. Each token s_i is assigned a positive or negative relevance score r_i, depending on the player's choices: a negative value of -1 if the token was discarded, a positive value of 1 if it was kept in hand, or a value of 2 if it was used in a sentence. Tokens s_i and their relevance r_i are stored in a set along with their corresponding turns τ_i:

$$R = \{(s_i, r_i, \tau_i) : i \in \{1, ..., n\}\} \tag{5}$$

Since the direction of the story may change from turn to turn, players are led to reconsider which tokens to keep or discard. In consequence, tokens that have been discarded may become more relevant, or tokens that have been kept may become less relevant. Hence, we use a time decaying mechanism for updating the relevance of each token. The relevance scores R are updated based on the current turn number τ and the time decaying factor λ to decay

negative and positive relevance values towards zero. Time-decayed relevance scores R' are calculated as follows:

$$R' = \{(s_i, r'_i, \tau_i) : i \in \{1, ..., n\}, r'_i = (1-\lambda)^{\tau-\tau_i} r_i\} \quad (6)$$

We then aggregate relevance scores based on unique tokens. This implies that tokens kept in hand gain relevance after each turn. We help players to use them by offering tokens that may fit together, unless they choose to discard them at some point. Furthermore, we filter out tokens which absolute relevance scores fall below a certain threshold γ (currently $\gamma = 0.5$). Since, at this point, we are uncertain whether the token is still relevant or not, we rely on features of the most relevant tokens to make this judgement.

After updating the relevance scores R' of tokens, we fit a *gradient boosting regressor* (GBR) model, a supervised learning algorithm that builds a prediction model with an ensemble of weak learners such as decision trees [8]. We used the *LightGBM* implementation of this model, since it delivers more accurate predictions, requires less memory and is computationally faster than other implementations [11]. In each turn, we train this model with the features of synset vectors $\vec{s}_i$ of tokens as predictors of their corresponding relevance score r'_i. Then, we use the model to predict the relevance scores of synset vectors from the entire set of possible tokens T. As a final step, the Player Bag T_p is obtained by keeping only the top k tokens with the highest predicted relevance scores.

4.2.4 Group Bag. In the Group Bag T_g, the same approach is used as for the Player Bag. However, instead of collecting relevance information from a player's hand of tokens, we use information from all submitted sentences and their received votes.

The tokens s_i are extracted from submitted sentences and their relevance score r_i is determined by the number of votes received by the sentence. Furthermore, a *winner factor* w_f greater than one (e.g. $w_f = 1.2$) is used in order to increase the relevance score associated with the winner sentence of the turn. Then, we apply the same procedure to fit a gradient boosting model using synset vectors $\vec{s}_i$ of tokens as predictor features of the time-decayed relevance scores R'. Similar to the Related Bag and the Player Bag, we only keep the rank of the top k tokens with the highest predicted relevance scores.

4.2.5 Random Bag. As the name suggests, the Random Bag does not rank tokens, rather it uses the entire set of tokens T for the purpose of sampling tokens randomly.

4.3 Tokens Sampling

Many analog and digital games use randomness as a means to guarantee variety and challenge. We use randomness as a way to encourage players to come up with original sentences that are unlike those of their peers. Since the Related, History, Group, and Random Bags are shared among all players, if tokens were drawn solely based on ranking order, players would receive basically the same served hand and, as a result, they might compose very similar sentences. For this reason, we use a stochastic token sampling mechanism, which ensures that each player receives a different served hand, while still maintaining relatedness and relevance information.

Specifically, we use the *probability density function* (PDF) of a *beta distribution* to sample tokens based on their ranking order within each bag, with the top ranked tokens most likely to be sampled. We chose this distribution due to its versatility and finite interval of continuous values bounded between 0 and 1. Beta distributions are characterized by two parameters, α and β, as follows:

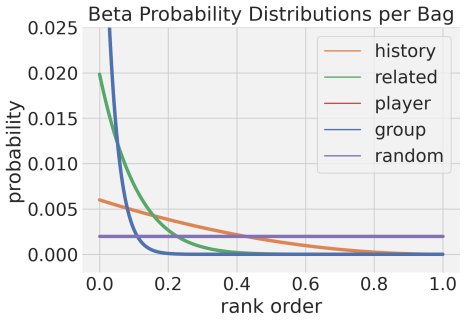

Figure 4: Beta probability distributions for sampling tokens from each ranked bag. Note that the Group and Player Bags have the same Beta value (30), and hence follow the same distribution (blue curve).

$$f(x, \alpha, \beta) = \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad (7)$$

The beta function $\mathrm{B}(\alpha, \beta)$ is the normalization of a constant determined by the gamma function $\Gamma(n)$:

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \Gamma(n) = (n-1)! \quad (8)$$

The parameters α and β can be adjusted to represent different probability distributions. For example, in our implementation, we fixed $\alpha = 1$ and empirically chose different values of β in order to obtain different probability distributions for each bag. The value of β for the Related Bag was set to $\beta = 10$ to have a higher chance of sampling tokens that are the most related to the story, while still allowing lower ranked tokens to be sampled. The value β for the Player and Group Bags was set to $\beta = 30$ to emphasize sampling primarily from the top ranked tokens. For the History Bag, the value of β was set to $\beta = 3$ to allow lower ranked tokens to be sampled in order to consider as many tokens as possible that appeared in the story. And, the value of β for the Random Bag was set to $\beta = 1$, which simply corresponds to a uniform random distribution. In Figure 4, we illustrate the probability distributions for each bag.

In addition to sampling tokens from each bag individually, the bag where a token is drawn from is also randomly selected (see Figure 5). This is repeated in order to gather a number of n_s tokens per pile. Although we currently select a bag at random for each token, this could be adjusted in order to favor certain bags more than others (e.g. sampling more frequently from the Player Bag).

As part of the game, players are asked to choose piles (categories) of tokens to draw from. When sampling for a specific pile c, we filter the bags in B_T, drawing only tokens for that category. Finally, as depicted in Figure 5, all sampled tokens are shuffled at random before being served to the player. In the current implementation, players are required to select three piles, and they will receive a sample of $n_s = 5$ tokens from each pile. In addition, the Verb pile is always included, resulting in a served hand consisting of 20

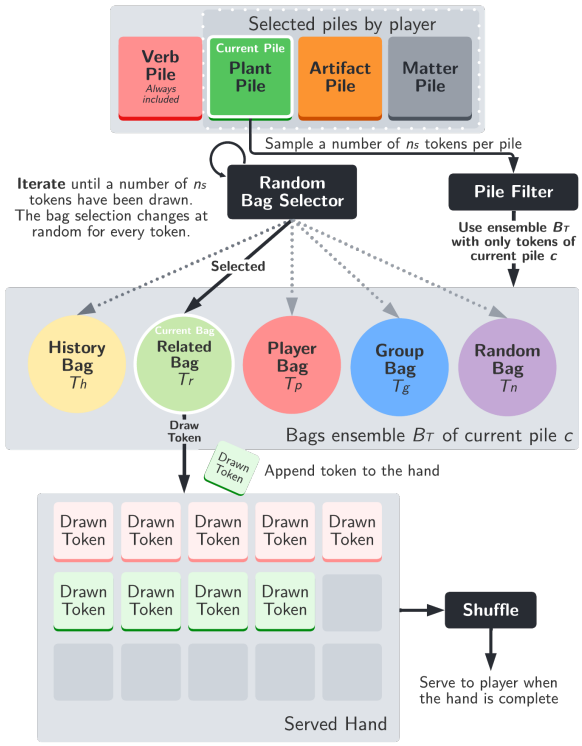

Figure 5: Token sampling procedure, from each pile and from different bags, to build a served hand.

tokens per player per turn. The served hand size was empirically determined to ensure that players feel neither overwhelmed with too many tokens nor deprived of options.

5 EXPERIMENTS

In our experiments, we examined the influence of our mixed-initiative method on players' choices. Specifically, we performed an analysis of the probability of players undertaking 3 different actions on tokens from the various bags: *collecting* a token from a specific bag when it is served, *using* it in a sentence or *discarding* it once collected in their hand. We hypothesized that players would collect more tokens from the specialized bags (History, Related, Player, and Group bag) than those randomly selected from the Random Bag. We argue that validating this hypothesis can confirm that our method provides players with tokens perceived to be more valuable for the story than those obtained by random sampling. Note that during the game, players are unaware of the origin bag of each token. In addition, we assessed the perceived quality of TaleMaker's stories as well as a few game aspects.

5.1 Procedure

For this experiment, we conducted several game sessions in an online and casual environment. At the start of the game session, players were asked to read the in-game instructions. These instructions provide a short introduction to the game. Then, players played a single turn to familiarize themselves with the game mechanics. As soon as they understood the game, we moved on to the actual game run. During the session, we collected data of the players' choices throughout the game, which was recorded in the server's database.

Upon completion of the game, players were asked to complete an in-game questionnaire concerning certain aspects of the game (see Table 1). Players evaluated the difficulty of creating a sentence using the served words, their enjoyment of creating a story with TaleMaker, the fairness of the game, and its reward system.

Once completed the first questionnaire, participants were given a second questionnaire to evaluate the perceived quality of their own story and, optionally, of other stories from other game sessions (see Table 2). The involved questions were adapted from a prior study with the intent of quantifying perceived story quality through assessing different aspects of the story [2]. In addition, we incorporated an additional question to assess whether a story is on topic with the title [26].

Table 1: Subjective Game Aspects Evaluation

Questionnaire used to evaluate aspects of the game
1. How do you rate the *difficulty* to make a sentence with the provided words? 1 (*very difficult*) to 5 (*very easy*)
2. How *enjoyable* was the creation of a story with this game?
3. How *fair* was the selection of sentences?
4. How useful was the *reward system* (gold and diamonds) to encourage you to create better sentences for the story?
Note. All questions are answered with a 5-point Likert scale. Answers for Q_2, Q_3, Q_4 are labeled with 1 (*not at all*), 2 (*slightly*), 3 (*somewhat*), 4 (*moderately*) and 5 (*extremely*).

Table 2: Perceived Story Quality Assessment

Questionnaire used to assess the quality of stories
1. To what extent is this story *entertaining*?
2. To what extent is this story *coherent*?
3. To what extent is this story *"true to life"*?
4. To what extent is this story *emotional*?
5. To what extent is this story *memorable*?
6. To what extent is this story *original*?
7. To what extent is this story *rich in imagery*?
8. To what extent is this story *engaging*?
9. (*Fidelity*) To what extent is this story on topic with the title?*
Note. All questions are positively worded and answers to the questions were made on a 5-point Likert scale. Answers are labeled with 1 (*not at all*), 2 (*slightly*), 3 (*somewhat*), 4 (*moderately*) and 5 (*extremely*).
* Additional question not included in the original questionnaire [2]

5.2 Participants

In this experiment, a total of 22 individuals volunteered to play the game. These individuals came from diverse backgrounds and were

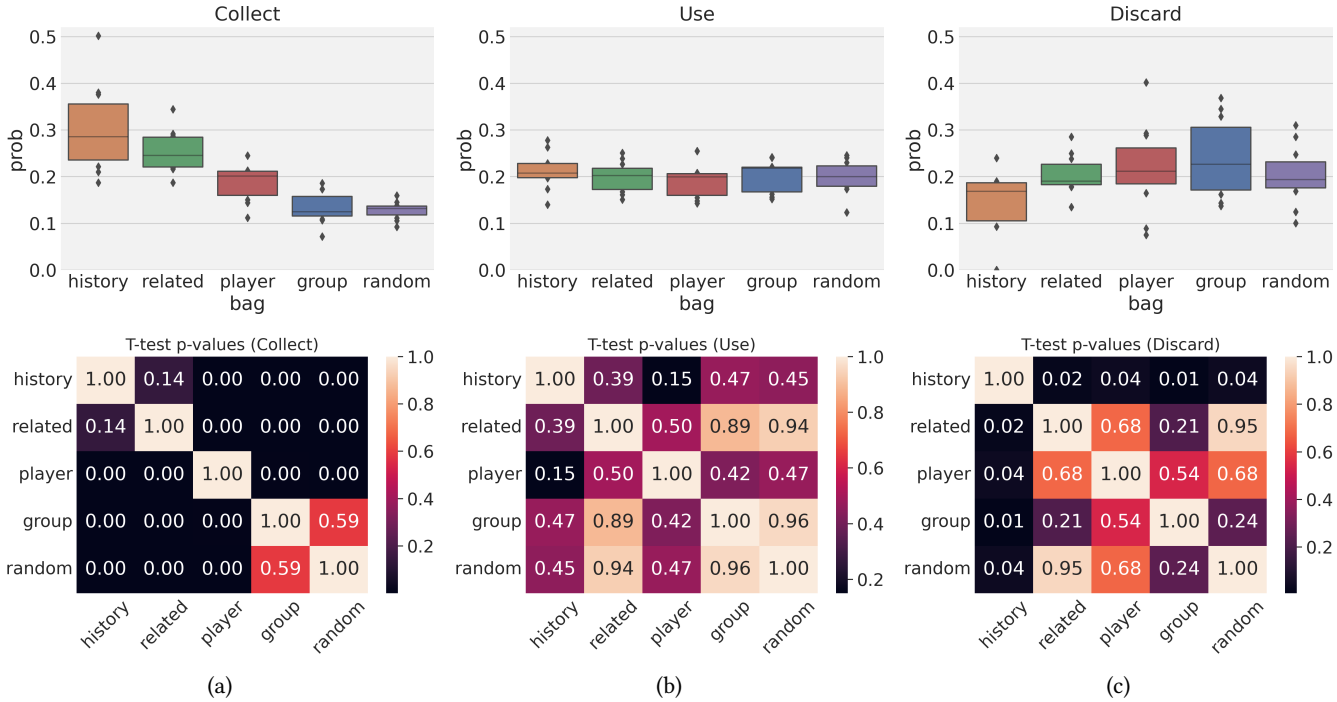

Figure 6: Distributions of the probabilities for collecting a token from each bag, and whether a token will be used or discarded once it has been collected. The p-values for a pairwise T-test are plotted in the form of a heatmap. The probabilities were sampled from each game (a total of 11 games). The black dots represent the probabilities of each game session.

gathered from multiple channels, including friends, colleagues, and anyone who wanted to participate. A total of 14 game sessions were conducted, resulting in a total of 14 stories. However, only 11 stories were considered in our analysis, as three of them were too short compared to the rest. The game was played on different days, with groups varying from two to six players. Furthermore, since this preliminary experiment had a particular focus, we allowed players to participate more than once.

5.3 Results

5.3.1 Bags Analysis. In our analysis, we looked at the data obtained from the players' choices regarding the tokens they received every turn. Table 3 shows the probability of *collecting*, *using* or *discarding* tokens from distinct bags. The probability that the token will be used or discarded is only considered once players have the token in their hand. In addition, we present in Figure 6, a graphical representation of the distributions, and, since normality assumptions were met, we include a pairwise T-test (as a heatmap) that demonstrates whether there were statistically significant differences among the distributions of probabilities between bags for each action.

Interestingly, Figure 6a shows that players are more likely to collect tokens that have previously appeared in the story (History Bag) and that are related to the story (Related Bag), as opposed to random tokens from the Random Bag ($p < 0.05$). This emphasizes the importance of these two bags for helping players to keep the story on topic and consistent with the context. Furthermore, tokens from the Player Bag appeared to be more likely to be collected than random tokens from the Random Bag ($p < 0.05$). These findings partially support our hypothesis that players prefer tokens from our specialized bags. Still, players collected tokens from the Group Bag about as seldom as from the Random Bag ($p > 0.05$). This might indicate a preference for their own envisioned story direction, rather than for considering other players' attempted directions.

Table 3: Mean probabilities of players' action per bag (collect, use and discard) and two tests of normality (data was collected from the first turn onwards)

					Tests of Normality			
					Kolmogorov-Smirnov		Shapiro-Wilk	
Action	Bag	N	Mean	Std.	Stat.	Sig.	Stat.	Sig.
collect	history	11	0.301	0.093	0.574	0.001	0.934	0.454
	related	11	0.252	0.045	0.574	0.001	0.944	0.568
	player	11	0.185	0.039	0.544	0.001	0.938	0.494
	group	11	0.134	0.034	0.529	0.002	0.955	0.703
	random	11	0.128	0.019	0.537	0.002	0.973	0.911
use	history	11	0.212	0.038	0.556	0.001	0.969	0.873
	related	11	0.199	0.032	0.560	0.001	0.966	0.847
	player	11	0.189	0.033	0.557	0.001	0.917	0.292
	group	11	0.201	0.034	0.560	0.001	0.871	0.080
	random	11	0.200	0.036	0.549	0.001	0.943	0.556
discard	history	11	0.135	0.079	0.500	0.005	0.879	0.101
	related	11	0.204	0.041	0.553	0.001	0.953	0.678
	player	11	0.217	0.092	0.530	0.002	0.949	0.632
	group	11	0.240	0.082	0.554	0.001	0.933	0.437
	random	11	0.203	0.062	0.540	0.002	0.973	0.914

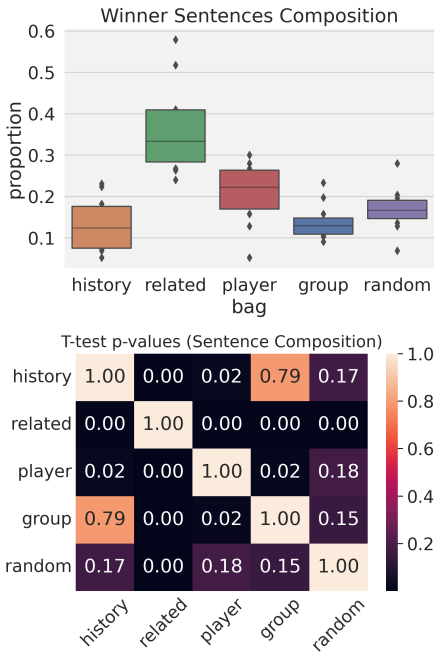

Figure 7: Composition of winner sentences. The average proportions of tokens from each bag, along with a p-values heatmap of T-tests comparing the means between distributions.

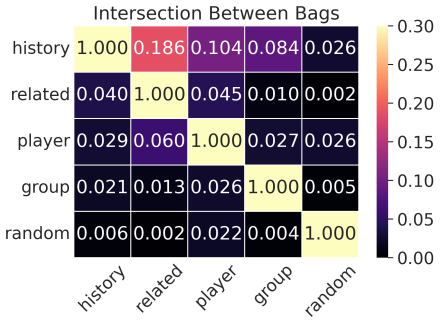

Figure 8: The proportion of intersection of matching tokens between bags.

Figure 6b shows that once tokens are collected in a player's hand, tokens from each bag are equally likely of being used ($p > 0.05$). This may be explained by the fact that players are maintaining their own collection of tokens, so most tokens that they keep are those that they are actually planning to use. In the discard action shown in Figure 6c, it appears that tokens from the History Bag are less likely to be discarded once they have been collected. In addition, tokens from other bags are almost equally likely to be discarded from players' hands ($p > 0.05$).

We examined the composition of the sentences selected to be part of the story. For this, we omitted from our analysis the default tokens included in the game. In Figure 7, we show the average

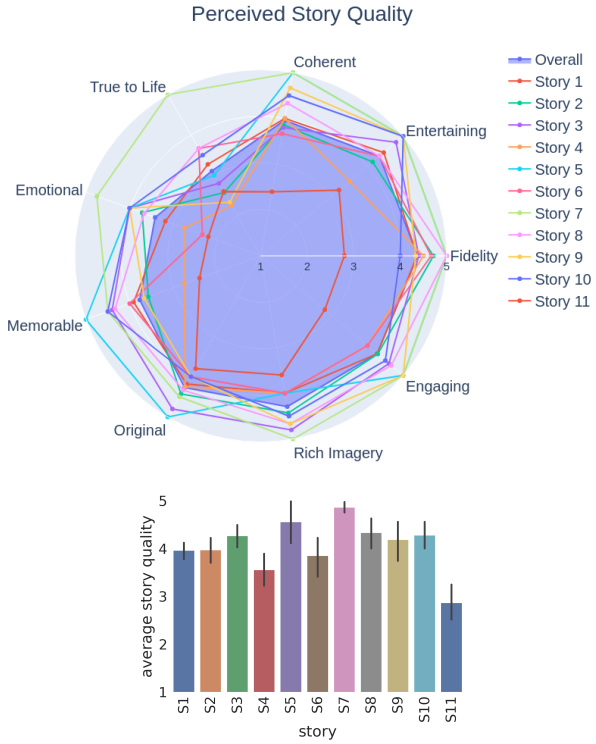

Figure 9: Perceived quality of TaleMaker stories

proportions of tokens from each bag for the winner sentences. A significant portion of tokens offered from the Related Bag dominated their composition, most likely because players were more interested in creating sentences that kept the context of the story consistent. Tokens from the History Bag dominated the least, since they were served less often, especially at the beginning, because the story simply takes time to grow (no more than one sentence per turn).

We also examined the differences in bag composition. For this purpose, we calculated the proportion of shared tokens among bags to determine whether the tokens in each bag were unique. As shown in Figure 8, the composition of the bags differs in most cases, except for the History Bag, which contains tokens collected from other bags and incorporated into the story. Note that the figure is not symmetrical since the intersection is divided by the bag size. For example, the History vs. Related Bag intersection is divided by the size of the History Bag, whereas Related vs. History Bag is divided by the size of the Related Bag. Thus, in this way we computed the proportion of how much one bag is contained within another.

5.3.2 Perceived Story Quality Assessment. On average, each story received four assessments. In Figure 9, we show a radar chart that summarizes the perceived story quality average ratings for each story, along with the overall ratings. Also included are the average story quality for each story. Stories were generally considered to be of good quality ($\overline{Q} \approx 4$).

5.3.3 Game Aspects. As illustrated in the radar chart of Figure 10, in general, players enjoyed creating a story with TaleMaker. A

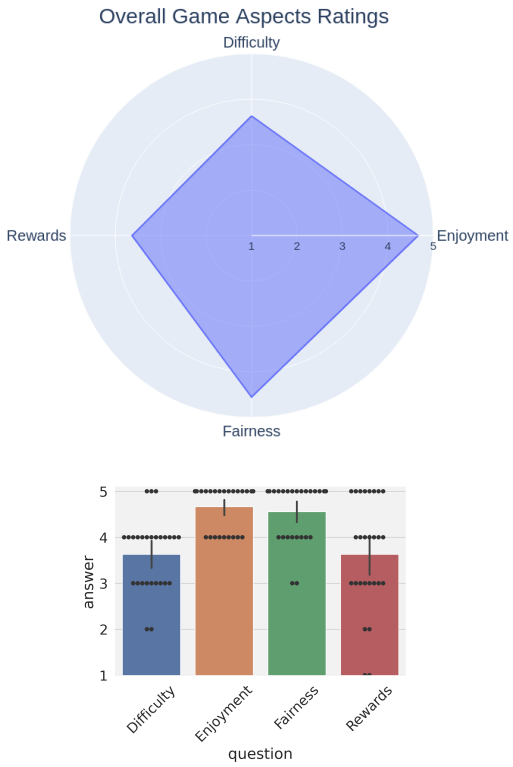

Figure 10: Subjective Game Aspects Evaluation.

further finding was that the game was deemed fair, highlighting the strength of a voting system that enables players to create a better story. Additionally, given the competitive nature of the game and its ability to support collaborative play, rewards were helpful for some players, but not for others, resulting in varying ratings for question 4 in Table 1. In average, players found it easy to construct sentences with the served words, confirming the effectiveness of our mixed-initiative method for supporting players' creativity.

The game sessions lasted from one to three hours. In addition, the average time, from selecting piles to submitting a sentence, was around four minutes. As shown in Figure 11, participants took longer to complete the collection phase. We can attribute this to the creative significance of this phase, in which players plan their sentences, not only for the current turn, but also for the turns ahead. In addition, since players accumulate tokens they are interested in, it can take time to decide which tokens to trade in exchange for new ones.

Moreover, as shown in Figure 12, we found that the standard piles of *adjective*, *activity*, and *other* were the most frequently chosen piles, and their tokens the most frequently used for a sentence. The *other* pile contains tokens of nouns that were not classified into specific piles (e.g. feelings, natural processes, time, etc.), but which many players found useful.

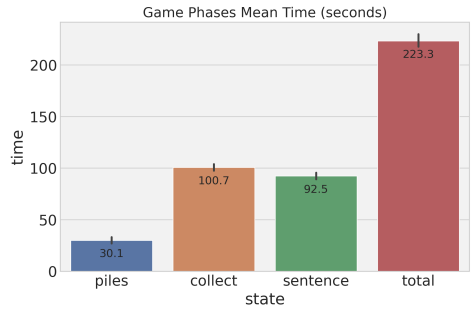

Figure 11: Average duration (in seconds) for selecting piles, collecting tokens and composing a sentence.

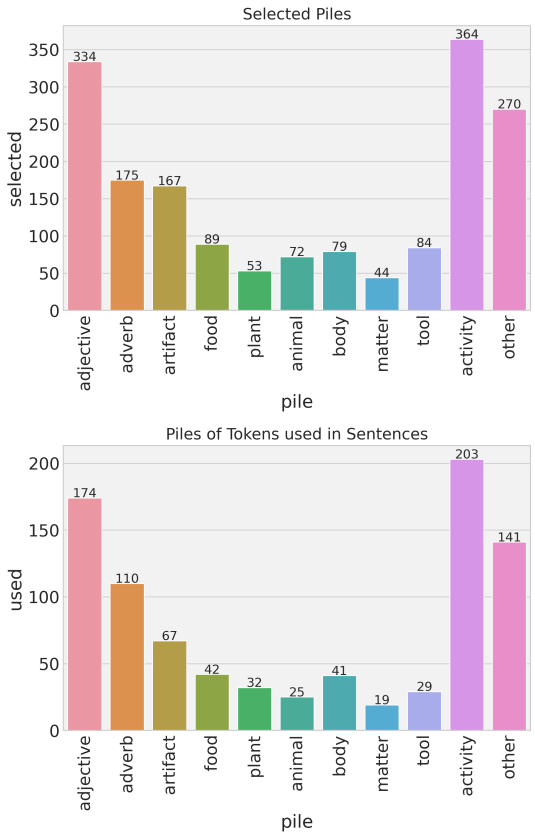

Figure 12: Piles selected throughout the game and the counts of tokens' piles used for sentences

6 DISCUSSION

In general, we observed that players were positively engaged when co-creating a story with TaleMaker. Several players reported that playing in a large group was enjoyable, since many original, often humorous, sentences were authored in each turn. Interestingly, TaleMaker could also be played between two players, which was more engaging when played in a co-located setting than in an online setting. By doing so, both players could discuss which of their

sentences contributes most to the story and brainstorm on which direction the story should take. Furthermore, in many players' comments, TaleMaker was described as a creative puzzle game in which they had some ideas of where they wanted their story to go, but were constrained by the tokens they had at their disposal. Nevertheless, according to them, this dynamic enhanced the experience of playing the game.

Regarding future improvements to TaleMaker, several issues need to be addressed and investigated. Firstly, we used a vocabulary of approximately 50k words, although many of these words would normally not appear in a casually crafted story. Probably, a smaller vocabulary would enhance the experience of creating a story with the game. In addition, the synset vectors used by our hand generation method were learned on a specific corpus of photo captions [5], which may have its own shortcomings and biases that could potentially be transferred to players' stories. It is, therefore, worthwhile to explore other synset vectors learned on other corpora and evaluate their potential for improving the story creation process with TaleMaker. In any case, one of the major advantages of our method is that it is (i) modular, i.e. vocabulary and synset vectors can be replaced, and (ii) extensible, i.e. additional bags of ranked tokens can be computed for different purposes. It would be interesting to perform experiments with bags that are computed using other methods and corpora, to examine their effects on players' creativity and on the output stories created.

Additional feedback from players revealed other challenges posed by our current implementation. Several individuals stated that they were frustrated about not being able to compose grammatically correct sentences and that they missed proper word conjugations (e.g. plurals) and some basic words (e.g. a, an, the...). Yet, due to the co-creation focus of our work, we chose to favor the core elements of our mixed-initiative approach, rather than those natural language aspects. In contrast, some players considered this to be a positive limitation because, rather than attempting to be a good writer, they were primarily concerned with the story and how it could be developed. Several players reported that the incidental ambiguity of a sentence encouraged discussion among co-located players, as well as allowed them to use their imagination to interpret the events of the story. We found particularly amusing to hear how different people were telling the story once the story was completed, and sometimes even justifying some of the actions taken by the characters of the story.

Some players felt restricted by the limitation of a single verb per sentence, as they wanted to express something that required two verbs (e.g., Bob wants to go shopping). In the future, this restriction might be softened to requiring all sentences to contain at least one verb, while maintaining a limit on the maximum number of tokens per sentence, to prevent overly lengthy sentences. Also, some players felt that the maximum amount of 8 tokens per hand was sometimes too restrictive, and that they frequently had to decide about which words to discard from their hand. In future iterations, we will explore increasing the maximum hand size, to check whether it improves players' experience with the game, as well as whether it benefits the story.

As a result of our experiments, we think piles may need to be refined, in order to provide players with a greater variety of token types. For example, the current implementation features the *other* pile, which could be divided into several other piles. However, this refinement must be done sensibly, as granularity in the pile categories could complicate the game, or make it too easy for players to construct a sentence.

Another challenge with the current TaleMaker implementation is the amount of time required to co-create a 'reasonable story'. Increasing the game's pace and allowing for more efficient flow through its phases, would also enable conducting more and quicker playtest experiments, to address research questions that require more elaborate experimental designs.

Last, but not least, TaleMaker's open-ended stories present the important challenge of putting in place appropriate moderation, so that its stories do not feature offensive or improper content. Although this is partly prevented by not allowing players to write text directly, a sentence might still become offensive as a result of the arrangement of tokens. As of now, at the end of the game TaleMaker asks players if their story can be considered offensive or obscene. However, in the future, a more rigorous, player-centered process could be more effective. For example, during the voting round, one could also report offensive sentences from other players, and even flag their author, possibly issuing their exclusion from the group.

7 CONCLUSION

In this paper, we introduced a novel mixed-initiative approach that addresses challenges commonly encountered in traditional and digital storytelling games. In order to demonstrate this approach, we developed TaleMaker, an online multiplayer game that allows a group of players to playfully co-create an open-ended story. Our approach is centered on a novel hand generation method, which selects tokens in such a way as to strike a balance between supporting players' creativity, maintaining story quality, and ensuring the competitive nature of the game. Moreover, TaleMaker does not impose the structure of the story but, rather, allows the group to decide where the story should go, and when to end it. Players deemed it fair to add sentences to the story by using the democratic process of voting. In a preliminary evaluation, we found that TaleMaker enables the enjoyable creation of high-quality stories. Moreover, the digital and online nature of TaleMaker can easily lead to a steady flow of new stories being authored and shared on a daily basis. The TaleMaker's database of stories is freely available to the research community, as described elsewhere [3].

REFERENCES

[1] J. Timothy Balint and Rafael Bidarra. 2022. Procedural generation of narrative worlds. submitted for publication.
[2] Jacqueline M Baron and Susan Bluck. 2011. That was a good story! Preliminary construction of the perceived story quality index. *Discourse Processes* 48, 2 (2011), 93–118.
[3] Mijael R. Bueno Pérez and Rafael Bidarra. 2022. The TaleMaker database of mixed-initiative co-created stories. In *17th International Conference on the Foundations of Digital Games*. ACM, Athens, Greece, 4 pages.
[4] Mijael R. Bueno Pérez and Rafael Bidarra. 2022. TaleMaker Game and Database. https://mijaelb.github.io/talemaker/.
[5] Mijael R. Bueno Pérez, Elmar Eisemann, and Rafael Bidarra. 2021. A synset-based recommender method for mixed-initiative narrative world creation. In *Interactive Storytelling - 14th International Conference on Interactive Digital Storytelling (Lecture Notes in Computer Science, Vol. 13138)*, Alex Mitchell and Mirjam Vosmeer (Eds.). Springer, Tallinn, Estonia, 13–28. https://doi.org/10.1007/978-3-030-92300-6_2

[6] Kate Compton and Michael Mateas. 2015. Casual Creators. In *Sixth International Conference on Computational Creativity*, Hannu Toivonen, Simon Colton, Michael Cook, and Dan Ventura (Eds.). Brigham Young University, Park City, UT, USA, 228–235. http://computationalcreativity.net/iccc2015/proceedings/10_2Compton.pdf

[7] 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 (Lecture Notes in Computer Science, Vol. 8832)*, Alex Mitchell, Clara Fernández-Vara, and David Thue (Eds.). Springer, Singapore, 73–80. https://doi.org/10.1007/978-3-319-12337-0_7

[8] Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational statistics & data analysis* 38, 4 (2002), 367–378.

[9] Atlas Games. 1993. Once Upon a Time - Storytelling Card Game. https://www.atlas-games.com/onceuponatime.

[10] Cards Against Humanity. 2010. Cards Against Humanity. https://www.cardsagainsthumanity.com/about.

[11] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). Curran Associates, Inc., Long Beach, CA, USA, 3146–3154. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

[12] Max Kreminski, Melanie Dickinson, Michael Mateas, and Noah Wardrip-Fruin. 2020. Why Are We Like This?: The AI architecture of a co-creative storytelling game. In *International Conference on the Foundations of Digital Games*, Georgios N. Yannakakis, Antonios Liapis, Penny Kyburz, Vanessa Volz, Foaad Khosmood, and Phil Lopes (Eds.). ACM, Bugibba, Malta, 13:1–13:4. https://doi.org/10.1145/3402942.3402953

[13] Max Kreminski, Melanie Dickinson, and Noah Wardrip-Fruin. 2019. Felt: A simple story sifter. In *Interactive Storytelling - 12th International Conference on Interactive Digital Storytelling (Lecture Notes in Computer Science, Vol. 11869)*, Rogelio Enrique Cardona-Rivera, Anne Sullivan, and R. Michael Young (Eds.). Springer, Little Cottonwood Canyon, UT, USA, 267–281. https://doi.org/10.1007/978-3-030-33894-7_27

[14] Ben Kybartas and Rafael Bidarra. 2015. A semantic foundation for mixed-initiative computational storytelling. In *Interactive Storytelling - 8th International Conference on Interactive Digital Storytelling (Lecture Notes in Computer Science, Vol. 9445)*, Henrik Schoenau-Fog, Luis Emilio Bruni, Sandy Louchart, and Sarune Baceviciute (Eds.). Springer, Copenhagen, Denmark, 162–169. https://doi.org/10.1007/978-3-319-27036-4_15

[15] Ben Kybartas and Rafael Bidarra. 2017. A survey on story generation techniques for authoring computational narratives. *IEEE Transactions on Computational Intelligence and AI in Games* 9, 3 (2017), 239–253. https://doi.org/10.1109/TCIAIG.2016.2546063

[16] Antonios Liapis. 2019. The Newborn World: Guiding creativity in a competitive storytelling game. In *IEEE Conference on Games*. IEEE, London, United Kingdom, 1–8. https://doi.org/10.1109/CIG.2019.8848116

[17] George A Miller. 1995. WordNet: a lexical database for English. *Commun. ACM* 38, 11 (1995), 39–41.

[18] Olatz Castaño Pérez, Ben Kybartas, and Rafael Bidarra. 2016. TaleBox: A mobile game for mixed-initiative story creation. In *Joint International Conference of Digital Games Research Association and Foundation of Digital Games*, Staffan Björk, Casey O'Donnell, and Rafael Bidarra (Eds.). Digital Games Research Association/Society for the Advancement of the Science of Digital Games, Dundee, Scotland, UK, 1–2. http://www.digra.org/wp-content/uploads/digital-library/paper_482.pdf

[19] Adam Plunkett. 2020. Tall Tales. https://www.kickstarter.com/projects/talltalesgame/tall-tales.

[20] The Noun Project. 2022. The Noun Project: Free Icons & Stock Photos for Everything. https://thenounproject.org.

[21] Melissa Roemmele and Andrew S Gordon. 2015. Creative Help: A story writing assistant. In *Interactive Storytelling - 8th International Conference on Interactive Digital Storytelling (Lecture Notes in Computer Science, Vol. 9445)*, Henrik Schoenau-Fog, Luis Emilio Bruni, Sandy Louchart, and Sarune Baceviciute (Eds.). Springer, Copenhagen, Denmark, 81–92. https://doi.org/10.1007/978-3-319-27036-4_8

[22] Jean-Louis Roubira, Régis Bonnessée, and Marie Cardouat. 2010. Dixit.

[23] Ben Samuel, Michael Mateas, and Noah Wardrip-Fruin. 2016. The Design of Writing Buddy: A mixed-initiative approach towards computational story collaboration. In *Interactive Storytelling - 9th International Conference on Interactive Digital Storytelling (Lecture Notes in Computer Science, Vol. 10045)*, Frank Nack and Andrew S. Gordon (Eds.). Springer, Los Angeles, CA, USA, 388–396. https://doi.org/10.1007/978-3-319-48279-8_34

[24] Reid Swanson and Andrew S Gordon. 2008. Say Anything: A massively collaborative open domain story writing companion. In *Interactive Storytelling - First International Conference on Interactive Digital Storytelling (Lecture Notes in Computer Science, Vol. 5334)*, Ulrike Spierling and Nicolas Szilas (Eds.). Springer, Erfurt, Germany, 32–40. https://doi.org/10.1007/978-3-540-89454-4_5

[25] Reid Swanson and Andrew S Gordon. 2012. Say anything: Using textual case-based reasoning to enable open-domain interactive storytelling. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3 (2012), 1–35.

[26] Lili Yao, Nanyun Peng, Ralph Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. 2019. Plan-and-Write: Towards better automatic storytelling. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (2019), 7378–7385. https://doi.org/10.1609/aaai.v33i01.33017378