

# Actionable Entities Recognition Benchmark for Interactive Fiction

**Alexey Tikhonov**  
Inworld.AI  
Berlin, Germany  
altsoph@gmail.com

**Ivan P. Yamshchikov**  
Max Planck Institute for  
Mathematics in the Sciences  
Leipzig, Germany  
CEMAPRE,  
University of Lisbon, Portugal  
ivan@yamshchikov.info

## Abstract

This paper presents a new natural language processing task - Actionable Entities Recognition (AER) - recognition of entities that protagonists could interact with for further plot development. Though similar to classical Named Entity Recognition (NER), it has profound differences. In particular, it is crucial for interactive fiction, where the agent needs to detect entities that might be useful in the future. We also discuss if AER might be further helpful for the systems dealing with narrative processing since actionable entities profoundly impact the causal relationship in a story. We validate the proposed task on two previously available datasets and present a new benchmark dataset for the AER task that includes 5550 descriptions with one or more actionable entities.

"One must never place a loaded rifle on the stage if it is not going to go off. It is wrong to make promises you do not mean to keep."

A. Chekhov.

## 1 Introduction

One of the bottlenecks that hold modern Natural Language Processing (NLP) from the generation of longer texts is the concept of *narrative* or storyline. There are constant attempts to generate longer blocks of text, such as (Kedzierski, 2019) or (Agafonova et al., 2020). These attempts succeed under certain stylistic and topical constraints that exclude the problem of narrative generation altogether. Although there are several recent results in the areas of suspense generation (Doust and Piwek, 2017), narrative personalization (Wang et al., 2017), and generation of short context-based narratives (Womack and Freeman, 2019), generating long stories is still a challenge (van Stegeren and Theune, 2019).

Though philosophers and linguists have tried to conceptualize the notions of plot, narrative arc, action, and actor for almost a century (Shklovsky, 1925; Propp, 1968; Van Dijk, 1976), few of these

theoretical concepts could be instrumental for modern NLP. Ostermann et al. (2019) present a machine comprehension corpus for the end-to-end evaluation of script knowledge. The authors demonstrate that though the task is not challenging to humans, existing machine comprehension models fail to perform well, even if they make use of a common-sense knowledge base. Despite these discouraging results, there are various attempts to advance narrative generation within the NLP community, see (Fan et al., 2019; Ammanabrolu et al., 2020).

We believe that one of the possible ways to overcome the challenge of narrative generation could be found through more profound insights into a narrative structure. However, there are frustratingly few established NLP tasks that engage at least some aspects of natural language crucial for narrative generation and comprehension. Partly, that state of affairs could be attributed to the void in our conceptual understanding of narrative from the point of view of computer science, see (Yamshchikov and Tikhonov, 2022). This paper tries to amend this and proposes a new NLP task — Actionable Entities Recognition (AER). The idea of the task is based on a simple premise: as the story unfolds reader understands which elements of the story might further affect the storyline. This phenomenon is especially pronounced in Interactive Fiction, where a player is trying to interact with the textually described environment via textual commands. We believe that detecting entities that the protagonists can interact with and that have an impact on the further development of the story might be instrumental for a better understanding of the story’s structures and, ultimately, pave the way to entertaining generative fiction.

The contributions of the paper are four-fold:

- it formalizes the problem of Actionable Entities Recognition;
- it provides a dataset of AEs, collected from

various interactive fiction resources;

- it contrasts AER and NER tasks and demonstrates that modern language models can detect actionable entities in interactive fiction, and their predictions go in line with the decisions of human players;
- finally, it demonstrates that the frequency of AEs in the text corresponds with the macrostructure of a storyline. Thus AER task might be one of the stepping stones toward the generation of long entertaining stories.

## 2 AER Task

Actionable Entities Recognition (AER) is a subtask of information extraction that seeks to locate entities mentioned in an unstructured text that significantly affect the narrative as the story unfolds. Similarly to named entity recognition, AER works with the entities for which one or many strings, such as words or phrases, stand consistently for some referent in line with the so-called concept of *rigid designators*, see (Kripke, 1971) and (Maxwell, 1978). We understand the capacity of an entity to change the structure of the story in terms of the philosophy of action, see (Van Dijk, 1976). The structural analysis of narrative first proposed in (Shklovsky, 1925) and revived in (Propp, 1968) is mostly focused on a characterization of the action and interaction sequences of 'heroes', the protagonists, and antagonists. (Shklovsky, 1925) introduces the concept of 'actor' as an entity that moves the story forward. Thus, AE is understood as the recognition of any 'actor' (no matter hero or not) in the unstructured text of the story.

It is important to note that we suggest understanding such *actionable entity* in the broadest possible terms. For example, if at some point a protagonist opens a window to air the room and the antagonist enters the building through this open window later as the story unfolds, such 'window' could be regarded as a perfect example of an actionable entity. We suggest doing it due to two core reasons. First, this definition allows us to pinpoint AEs as a broad class of entities mentioned in unstructured texts in natural language. However, the current definition allows further fine-graining of the term depending on the use case or research question. Second, this approach is simple in its naivety and thus allows us to avoid a deeper conceptual discussion of action in the narrative context.

We state that if a character interacts with something or someone, we declare it an *actionable entity*. The difference between a named entity and an actionable entity is in the interactive nature of the latter. In the next Section, we describe the dataset for AER that we propose. The examples of the dataset clarify these differences further.

## 3 Data

Interactive Fiction games are text-based simulators where a player uses text commands to change the environment and navigate through the story. Such games represent a unique intersection of natural language processing and sequential decision making, making them extremely valuable in a machine learning context and drawing various researchers' attention. For example, Narasimhan et al. (2015) show that one could learn control policies for text-based games with reinforcement learning methods. Côté et al. (2018) presents TextWorld, a sandbox learning environment for the training and evaluation of reinforcement learning agents on text-based games. Hausknecht et al. (2020) introduce Jericho, a learning environment for human-made interactive fiction games alongside possible benchmarks for language-based autonomous agents. Nelson et al. (2006); Martin et al. (2018); Jain et al. (2020) and many other authors develop various designs of reinforcement learning agents for interactive fiction that play with various types of feedback and demonstrate various levels of generalization.

Large action spaces impede an agent's learning ability, especially when many actions are redundant or irrelevant. This is especially prevalent in interactive fiction that combines natural language understanding with sequential decision-making. Navigating through interactive fiction, the agent must often detect affordances: the set of behaviors enabled by a situation. Affordance detection benefits domains with large action spaces, allowing the agent to prune its search space by avoiding futile behaviors. (Fulda et al., 2017) present such an affordance detection mechanism, and (Haroush et al., 2018) develop the action-elimination architecture that solves quests in the text-based game of Zork, significantly outperforming the baseline agents.

In some sense, Actionable Entities recognition could be linked to affordance detection mentioned above, yet AER is an isolated natural language processing (NLP) task due to its relevance to the structure of the narrative. Further research of AER as

a stand-alone NLP task could shed light on fundamental properties of narrative generation, systematic principles of human attention, and structural conditions that these principles impose on story formation. Further in this paper, we illustrate AER applicability to NLP tasks outside the scope of interactive fiction. However, interactive text games are perfect playgrounds to test and benchmark AER algorithms.

This paper presents a new benchmark — A Benchmark for Actionable Entities Recognition (BAER). The dataset includes 5550 locations with one or more AE in each of them and is based on 995 games. Its total size is 1.2 megabytes. Every Actionable Entity in the text is labeled as shown in Figure 1. The nature of interactive fiction helps to distinguish AER from NER, since there could be several entities in a given location. However, only several of them are actionable and affect the storyline. We publish the resulting BAER dataset<sup>1</sup> with labeled AEs and suggest it for AER algorithms benchmarking.

*The Museum of Illusion was dedicated to the memory of the Great **[Implementers]** - figures who were believed to have created the GUE as an act of pure will. Before you, a row of delicately crafted **[porcelain busts]** of these immortal greats once stood. Legend has it that so lifelike were these **[busts]**, that they would seem to talk among themselves, discussing history, the arts, music, and philosophy, much as those mythic figures did in the Golden Age of Text Adventures. But years of neglect and the ravages of time have toppled most from atop their finely-wrought pedestals. Now, alas, the only two that remain are those of **[Marc Blank]** and **[Mike Berlyn]**, bloodied, but unbent; battered, but unbroken; shaken, but not stirred.*

Figure 1: Text description of a location with highlighted items available for interaction. One can clearly see that AER differs from NER though there is an overlap.

The data consists of three different types of texts from interactive fiction games and is formatted similarly to the example given in Figure 1. The first part of the dataset is based on the Jericho introduced in (Hausknecht et al., 2020). It includes 57

interactive fiction games with lists of all locations and all actionable entities. One could assume that the game’s authors regarded every actionable entity in these descriptions as something that could help the player move forward in plot evolution. Iteratively searching all such actionable entities in every location with a script, we get a list of Actionable Entities.

We have also collected 24 additional interactive fiction games that have published solutions and are not included in Jericho. For every game, we have created a script that can execute the commands described in the solution and controls the location where the command is executed. As an output of this script, we obtain a list of locations on the critical path that leads to the completion of the game. For every location, we also store a prefix of commands that leads the character to this particular location. Then we brute force the labels of AEs in the following manner. For every location, we reset the game and entered the prefix that brings us into this location. We store the description of the location and start iterating over all objects in this description using the command `examine <obj>`. If interaction with one of the entities gives a non-trivial reaction from the game, we label this entity as an AE. Most of IF have a standard response to inconsequential actions of the player. If the game’s response is different, we assume that the entity is actionable.

Finally, we include 1500 other interactive fiction games for which we found no working solution. Instead of a solution-based strategy, we implemented a random walk for labeling Actionable Entities in these games. In every game, we did 2500 random steps and implemented the same logic we used for the solution but in a limited number of locations that we were lucky to obtain.

## 4 Task validation

We propose several ways to approach the AER problem using the BAER dataset and other datasets available in the literature.

### 4.1 Contrasting Actionable and Named Entities

Let us discuss how AER differs from NER. To do that, we use a T-NER model developed by Ushio and Camacho-Collados (2021). The T-NER real-

<sup>1</sup><https://github.com/altsoph/BAER>

	Accuracy	F1-score
Pre-trained T-NER	0.05	0.00
T-NER fine-tuned on BAER	0.50	0.51

Table 1: Pre-trained XLN-RoBERTa T-NER hardly detects Actionable Entities out of the box. Yet after fine-tuning on BAER for entity span prediction the quality significantly improves.

ization based on the XLM-RoBERTa<sup>2</sup> works in the entity span prediction regime marking the beginning and the end of a named entity in the text. The transformer-based architectures significantly outperform classical NLP models on NER tasks, so we see no reason why the situation would differ for AER. Thus, we validate the proposed task regarding the state-of-the-art NER solution. Table 1 shows the results of the pre-trained T-NER of the BAER dataset along with results after fine-tuning.

Table 1 illustrates that Actionable Entities profoundly differ from Named Entities. Fine-tuning a pre-trained T-NER model on the BAER dataset boosts the F-1 score from virtually zero to around one-half, yet there is room for progress. It is important to note that BAER is far smaller than the NER datasets that are commonly used in the literature. This experiment allows us to conclude that AEs form a distinct category of entities that is out of the scope of current NER: some named entities are actionable, yet many Actionable Entities are NOT named.

Table 2 provides some qualitative understanding of the contrast between named entities and AEs. T-NER classifies detected entities into several categories. After fine-tuning XLM-RoBERTa-based T-NER on BAER dataset, we apply T-NER and our AER model to four different datasets and look for the T-NER categories that AEs dominate. The BAER column stands for the validation part of BAER. There are only three T-NER categories in which Actionable Entities are a majority. These are "product" with only one NER that is not an AE per every eighteen AEs, "person" with one NER that is not an AE per 2.6 AEs, and "work of art" with one NER that is not an AE per two AEs.

Yao et al. (2020) present a ClubFloyd dataset crawled from the ClubFloyd website<sup>3</sup> and contains 426 human gameplay transcripts, which cover 590 text-based games of diverse genres and

<sup>2</sup><https://huggingface.co/asahi417/tner-xml-roberta-large-all-english>

<sup>3</sup>[http://www.allthingsjacq.com/interactive\\_fiction.html](http://www.allthingsjacq.com/interactive_fiction.html)

styles. The data consists of 223,527 context-action pairs in the format [CLS] observation [SEP] action [SEP] next observation [SEP] next action [SEP]. Every context entry observation is accompanied by variants of human attempts to interact with objects and characters in the given context. Table 2 shows that for these context descriptions, T-NER categories "product," "person," and "work of art" are dominated by AEs as well as in BAER. However, there are other AE-dominated categories, such as "chemical."

Malysheva et al. (2021) present the TVMAZE<sup>4</sup> dataset. The dataset consists of 13 000 texts that describe plots of TV series split into short episode annotations. WikiPlots<sup>5</sup> is a collection of 112,936 story plots extracted from the English Wikipedia. These two datasets are significantly larger than datasets of interactive fiction. Table 2 demonstrates that once again, categories "person" and "product" are dominated by AEs, but such categories as "corporation," "organization," and "group" add up themselves to the picture.

An Actionable Entity is a type of entity that plays a causal role in developing a narrative. Since some of such entities are named, modern NER models could detect some percentage of AEs in a text. These would typically be single characters (say, Thor), groups or some organizations (say, Asgard), or objects that are "branded" in some sense (i.e., Mjölner, Obedience Potion, or statue of Loki). Along with AEs that are named, a large portion of AEs is not detected within the framework of NER. In this subsection, we have provided a basic benchmark for AER via fine-tuning XLM-RoBERTa based T-NER model and have shown that the model trained on BAER does not have to be only used for interactive fiction but can be applied to other NLP datasets.

## 4.2 Actionable Entities and Human World Knowledge

Humans tend to understand causal relations between entities in a given text intuitively. If a player wants to leave the location, she might try to use the door that was mentioned in the description. If the door is closed, to open it, the player could look for a key, etc. This given understanding of the "world" is crucial for interactive fiction, yet its importance is not limited to textual gameplay. Models endowed

<sup>4</sup><https://www.tvmaze.com/>

<sup>5</sup><https://github.com/markriedl/WikiPlots>



BAER 1 117 texts		ClubFloyd 43 795 texts		TV-MAZE 299 197 texts		WikiPlots 2 070 449 texts	
T-NER category	$\frac{\nu_{AER}}{\nu_{ner}}$	T-NER category	$\frac{\nu_{AER}}{\nu_{ner}}$	T-NER category	$\frac{\nu_{AER}}{\nu_{ner}}$	T-NER category	$\frac{\nu_{AER}}{\nu_{ner}}$
product	18	chemical	7.5	person	36.5	product	31.2
person	2.6	product	7.3	product	12.2	corporation	22.7
work of art	2.0	person	5.5	corporation	3.9	person	20.1
organization	0.7	other	2.1	organization	3.3	group	5.4
location	0.5	work of art	1.3	chemical	3.1	chemical	4.9

Table 2: Lists of T-NER categories in which Actionable Entities tend to have higher relative frequencies across four different datasets. For every T-NER category, the number of Actionable Entities that fall into this category is divided over T-NER recognized entities that are not Actionable Entities (provided this number is not zero). The table shows five categories in which Actionable Entities have the highest relative frequency.

with such understanding would be far more usable across the board of NLP tasks. Let us show that AER is a cornerstone for the acquisition of such knowledge.

In Subsection 4.1, we have already introduced the dataset presented in (Yao et al., 2020) and used the game context descriptions to explore the differences between NER and AER. Now let us focus on the actions that this dataset contains. These are actual attempts by human players to interact within the proposed environment. One action typically consists of a verb (i.e., go, take, open, etc.) followed by a description of an entity with which the player is trying to interact. We call these entities Action Targets (AT). Since the dataset contains game logs of several players, some ATs could be mentioned several times. These would be the entities that are perceived by humans as more interesting to explore and use in the game. Table 3 summarizes what share of ATs could be labeled by an AER model.

AER model predicts if an entity is an AE with some probability. Varying this probability threshold would change the prediction of the model. With a higher threshold, the list of AEs would naturally be shorter. Table 3 shows that the share of AEs that overlap with some human action targets grows if the threshold is higher. At the same time, if the model has a lower threshold, it predicts more than 80% of all the entities that humans try to interact with.

## 5 Actionable Entities and the structure of narrative

Papalampidi et al. (2019) introduce a TRIPOD

dataset<sup>6</sup> that includes plot synopses with turning point annotation. They also suggest looking at five turning points in the script since this variant is commonly employed by screenwriters as a practical guide for producing successful screenplays. These five turning points are:

- Opportunity — the introductory event that occurs after the presentation of the setting and the background of the main characters;
- Change of Plans — the event where the main goal of the story is defined;
- Point of No Return — the event that pushes the main character(s) to commit to their goal fully;
- Major Setback — the event where everything falls apart;
- Climax — the final event of the main story.

Figure 2 shows how the average number of AEs per sentence varies in these turning points. Since we are interested in relative dynamics within a story, we calculate the average number of AEs per sentence and see its difference with an average number of AEs across all five turning points. Since the number of AEs per sentence might be higher for longer sentences, we also calculate how the average number of words per sentence differs at every turning point.

Turning point number one — Opportunity — tends to have more words per sentence and more Actionable Entities in them. As the plot thickens in Change of Plans and Point of No Return, sentences get shorter. However, the number of Actionable

<sup>6</sup><https://github.com/ppapalampidi/TRIPOD>

AER model threshold	All payer action targets (AT)				Unique ATs			
	p > 0.5	p > 0.65	p > 0.8	p > 0.95	p > 0.5	p > 0.65	p > 0.8	p > 0.95
Share of AEs that occur in AT list	0.38	0.43	0.48	<b>0.57</b>	0.22	0.25	0.29	0.36
Share of ATs labelled as AEs	<b>0.84</b>	0.65	0.30	0.05	<b>0.84</b>	0.65	0.30	0.04

Table 3: XLN-RoBERTa T-NER fine-tuned on BAER predicts which action targets from the Club Floyd dataset a player would try to interact with. The entities with higher AE probability estimates almost surely end up on the list of action targets. If the entities with lower AE probability are included, they cover up to 84% of all action targets used by the human players. Since the Club Floyd dataset contains raw interactive fiction data, there are entities with which several players try to interact. The table provides both the results for the full list of action targets as well as for the list of unique action targets.

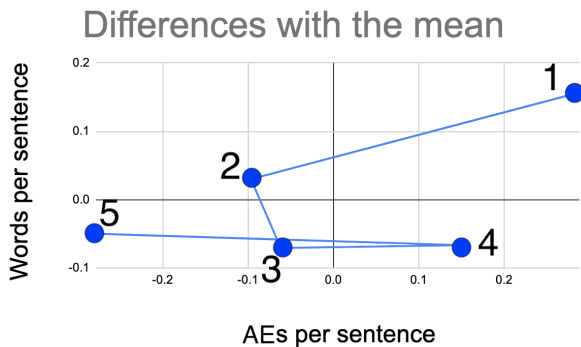


Figure 2: Average number of AEs per sentence and the average number of words in a sentence differ in different turning points of the story. The axis shows the difference between these values calculated at the corresponding turning point and the story at large. The first turning point of a story has more words and AEs per sentence than usual. Then the sentences tend to get shorter. The story’s second and third turning points have fewer Actionable Entities per sentence, while at the fourth turning point, the number of AEs per sentence peaks.

Entities per sentence gets lower as well. In his letter to Lazarev, Anton Chekhov famously wrote: "One must never place a loaded rifle on the stage if it is not going to go off. It is wrong to make promises you do not mean to keep." Rephrasing Chekhov’s quote, one could say that some of the "loaded rifles" are already placed on the stage, and we are waiting for the shots. Indeed, in Major Setback number of AEs per sentence jumps, while the average number of words per sentence remains low. In Climax, the number of AEs drops to its lowest.

Table 4 shows how AEs first occur and reoccur in the story depending on the turning point. Indeed, almost one-third of AEs first emerge in the Opportunity part of the story. In every later turning point, five to seven percent of all AEs in a story are the

Turning Point #	1	2	3	4	5
1	32.2	7.2	6.3	6.8	5.4
2	0.0	21.7	4.0	4.0	3.3
3	0.0	0.0	17.3	2.5	2.6
4	0.0	0.0	0.0	16.0	2.8
5	0.0	0.0	0.0	0.0	12.8

Table 4: Percentage of AEs in every turning point of a story. Every row represents AEs that first occurred in the corresponding turning point and then reoccurred in later parts of the story that stand in the corresponding column. The diagonal sums to 100% representing all first occurrences.

ones that we first meet in the Opportunity part. The percentage of first occurrences consistently drops from 32.2% in the first to 12.8% in the last turning point.

This is a qualitative picture that illustrates the potential of AER as a stand-alone NLP task: since Actionable Entities affect causal relations within the narrative, they could be useful for the overall analysis of the narrative structure.

## 6 Discussion

In recent years we have seen various exciting results in the area of Natural Language Generation (NLG). One line of research addresses the generation of semi-structured texts varying from dialogue responses, see (Li et al., 2016, 2017, 2019), to traditional Chinese poetry, see (Zhang and Lapata, 2014; Wang et al., 2016). Another line of research addresses the generation of stylized texts, see (Tikhonov and Yamshchikov, 2018; Yang et al., 2018). There have been recent results that try to generate longer blocks of text, such as (Kedziorski, 2019) or (Agafonova et al., 2020), yet the genera-

tion of longer texts is only possible under certain stylistic and topical constraints that exclude the problem of narrative generation altogether.

We believe understanding narrative and its key principles is paramount to further advances in NLG. However, the narrative is fundamentally causal. Thus, a deeper understanding of causality in NLP should provide new insights into the structure of the narrative. Currently, our understanding is lacking due to a variety of reasons:

- human cognition is often verbal and narrative-based, which makes attempts to conceptualize narrative implode on themselves;
- narration is not only a linguistic but a cultural act that fundamentally affects humans as 'cultural animals'. This influence of narrative hinders its conceptualization in rigorous mathematical terms;
- narrative is centered around long-term dependencies that could be formally characterized as a critical behavior of language, see (Lin and Tegmark, 2017).

These very general issues need particular attention, but we believe the AER is an exemplary step to address these problems. First, AER allows addressing the notion of 'actor' in line with (Shklovsky, 1925), broadening it from a person to any entity that interacts with a storyline. This ability to personalize various entities could be regarded as a cornerstone associative mechanism that might be a basis for creative cognition (Mednick, 1962). Secondly, as with any NLP task, AER could be developed for every culture or language. Further development of AER could provide general cross-cultural insights into the process of story formation. Finally, AEs significantly contribute to the criticality of language in line with ideas expressed in (Lin and Tegmark, 2017). AEs reoccur in the story, and understanding which object is pivotal for the narrative to unfold is one of the key bottlenecks to automated narrative generation.

In an interactive setting, trying to interact with the world can help to learn AEs. At the same time, the notion of AE can help models discover affordances in games. One could also look into how large language models perform in a zero-shot manner on AER tasks. We believe that all these factors make AER an NLP task that could bring us further toward narrative generation and provide insights into interactive learning.

## 7 Conclusion

This paper presents a new causal natural language processing task — Actionable Entities recognition (AER). It formalizes the notion of an Actionable Entity as any entity that could be engaged by one of the protagonists and thus affect the story's development in line with the structuralist sense. It provides a corpus of labeled texts to introduce the benchmark for AER. To illustrate the differences between AER and NER, it validates the proposed task on four different datasets. The paper demonstrates that the AER model predicts human attempts to interact with entities within interactive fiction. Finally, it shows how the AER classifier could be applied to an external dataset to obtain insights into the structure of the narrative.

## 8 Acknowledgments

Ivan Yamshchikov obtained some of the results while working at LEYA Laboratory by Yandex and Higher School of Economics in St. Petersburg. His work is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

## References

- Yana Agafonova, Alexey Tikhonov, and Ivan P Yamshchikov. 2020. Paranoid transformer: Reading narrative of madness as computational approach to creativity. *arXiv preprint arXiv:2007.06290*.
- 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7375–7382.
- Marc-Alexandre Côté, Ákos Kádár, Xingdi Yuan, Ben Kybartas, Tavian Barnes, Emery Fine, James Moore, Matthew Hausknecht, Layla El Asri, Mahmoud Adada, et al. 2018. Textworld: A learning environment for text-based games. In *Workshop on Computer Games*, pages 41–75. Springer.
- Richard Doust and Paul Piwek. 2017. A model of suspense for narrative generation. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 178–187.
- Angela Fan, Mike Lewis, and Yann Dauphin. 2019. Strategies for structuring story generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2650–2660.

- Nancy Fulda, Daniel Ricks, Ben Murdoch, and David Wingate. 2017. What can you do with a rock? affordance extraction via word embeddings. *arXiv preprint arXiv:1703.03429*.
- Matan Haroush, Tom Zahavy, Daniel J Mankowitz, and Shie Mannor. 2018. Learning how not to act in text-based games.
- Matthew J Hausknecht, Prithviraj Ammanabrolu, Marc-Alexandre Côté, and Xingdi Yuan. 2020. Interactive fiction games: A colossal adventure. In *AAAI*, pages 7903–7910.
- Vishal Jain, William Fedus, Hugo Larochelle, Doina Precup, and Marc G Bellemare. 2020. Algorithmic improvements for deep reinforcement learning applied to interactive fiction. In *AAAI*, pages 4328–4336.
- Richard Koncel Kedzior. 2019. *Understanding and Generating Multi-Sentence Texts*. Ph.D. thesis.
- Saul Kripke. 1971. Identity and necessity. *Identity and individuation*, 1971:135–64.
- Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao. 2016. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202.
- Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, and Dan Jurafsky. 2017. Adversarial learning for neural dialogue generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2157–2169.
- Ziming Li, Julia Kiseleva, and Maarten de Rijke. 2019. Dialogue generation: From imitation learning to inverse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6722–6729.
- Henry W Lin and Max Tegmark. 2017. Critical behavior in physics and probabilistic formal languages. *Entropy*, 19(7):299.
- Anastasia Malysheva, Alexey Tikhonov, and Ivan P Yamshchikov. 2021. Dyplocod: Dynamic plots for document classification. *arXiv preprint arXiv:2107.12226*.
- Lara J Martin, Srikanth Sood, and Mark Riedl. 2018. Dungeons and dqn: Toward reinforcement learning agents that play tabletop roleplaying games. In *INT/WICED@ AIIDE*.
- Grover Maxwell. 1978. Rigid designators and mind-brain identity.
- Sarnoff Mednick. 1962. The associative basis of the creative process. *Psychological review*, 69(3):220.
- Karthik Narasimhan, Tejas Kulkarni, and Regina Barzilay. 2015. Language understanding for text-based games using deep reinforcement learning. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1–11.
- Mark J Nelson, Michael Mateas, David L Roberts, and Charles L Isbell. 2006. Declarative optimization-based drama management in interactive fiction. *IEEE Computer Graphics and Applications*, 26(3):32–41.
- Simon Ostermann, Michael Roth, and Manfred Pinkal. 2019. Mscript2. 0: A machine comprehension corpus focused on script events and participants. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (\*SEM 2019)*, pages 103–117.
- Pinelopi Papalampidi, Frank Keller, and Mirella Lapata. 2019. Movie plot analysis via turning point identification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1707–1717.
- Vladimir Propp. 1968. Morphology of the folktale, trans. *Louis Wagner, 2d. ed.*
- Viktor Shklovsky. 1925. Theory of prose (b. sher, trans.). *Champaign, IL: Dalkey Archive Press. Original work published.*
- Alexey Tikhonov and Ivan P Yamshchikov. 2018. Guess who? multilingual approach for the automated generation of author-stylized poetry. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 787–794. IEEE.
- Asahi Ushio and Jose Camacho-Collados. 2021. T-ner: An all-round python library for transformer-based named entity recognition. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 53–62.
- Teun A Van Dijk. 1976. Philosophy of action and theory of narrative. *Poetics*, 5(4):287–338.
- Judith van Stegeren and Mariët Theune. 2019. Narrative generation in the wild: Methods from nanogenmo. In *Proceedings of the Second Workshop on Storytelling*, pages 65–74.
- Daisy Zhe Wang, Wei He, Hua Wu, Haiyang Wu, Wei Li, Haifeng Wang, and Enhong Chen. 2016. Chinese poetry generation with planning based neural network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1051–1060.
- Pengcheng Wang, Jonathan P Rowe, Wookhee Min, Bradford W Mott, and James C Lester. 2017. Interactive narrative personalization with deep reinforcement learning. In *IJCAI*, pages 3852–3858.



- Jon Womack and William Freeman. 2019. Interactive narrative generation using location and genre specific context. In *International Conference on Interactive Digital Storytelling*, pages 343–347. Springer.
- Ivan P. Yamshchikov and Alexey Tikhonov. 2022. What is wrong with language models that can not tell a story? *arXiv preprint arXiv:2211.05044*.
- Cheng Yang, Maosong Sun, Xiaoyuan Yi, and Wenhao Li. 2018. Stylistic chinese poetry generation via unsupervised style disentanglement. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3960–3969.
- Shunyu Yao, Rohan Rao, Matthew Hausknecht, and Karthik Narasimhan. 2020. Keep calm and explore: Language models for action generation in text-based games. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8736–8754.
- Xingxing Zhang and Mirella Lapata. 2014. Chinese poetry generation with recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 670–680.