Investigating the Use of Word Embeddings to Estimate Cognitive Interest in Stories

Morteza Behrooz
morteza@ucsc.edu
University of California Santa Cruz
1156 High St. Santa Cruz, California 95064

Justus Robertson
jjrobert@ncsu.edu
North Carolina State University
Raleigh, NC 27695

Arnav Jhala
ahjhala@ncsu.edu
North Carolina State University
Raleigh, NC 27695

Abstract

Narrative processing is an important skill to model both from a cognitive science perspective and a computational modeling perspective which applies to intelligent agents. Communication between humans often involves storytelling patterns that make the mundane exchange of information more interesting and with proper emphasis on important communicative goals. Current narrative generation models evaluate their generations based on either a priori domain semantics (e.g. game state for an in-game conversation with player agents) or generic text quality measures (e.g. coherence). However, in utilizing storytelling as a communicative tool for real-world interactions, domain-specific approaches fail to generalize and text quality measures fail to ensure that the narrative is perceived as *interesting*. Hence, such generation needs to consider the cognitive processes involved in the perception of narrative. Using theories of cognitive interest, we present results of an investigation of whether word embeddings (e.g. GloVe (Pennington, Socher, & Manning, 2014)) could be used to model and estimate cognitive interestingness in stories.

Introduction and Background

In computational narrative generation, the communication context for which the narratives are generated plays an integral role in determining both the method constraints during the generation and the evaluation metrics for the resulting narratives. Not all approaches to narrative generation are compatible with all narrative communication paradigms, because they result in vastly different qualities in the generated narratives and also differ in their assumptions and constraints.

Moreover, no single set of evaluation or optimization metrics can ensure the success of a narrative generator across multiple paradigms. Such "success" is usually dependent upon being received positively by the audience and achieving any potential communicative or social goals. In simpler terms, a "good" narrative has to be interesting to the audience.

Entertainment, and games in particular, have been a prominent context for narrative generation and communication. Many games change the events that are not (at least directly) in control of the player, or affect what the players say (in voice or text), in order to create the "best" storyline possible with a goal of maximal immersion and character believability (Mateas & Stern, 2003; McCoy, Treanor, Samuel, Mateas, & Wardrip-Fruin, 2011; Ryan, Mateas, & Wardrip-Fruin, 2016). Other games can involve an interactive settings, where the player can influence the progression of the story through making choices (Riedl & Bulitko, 2012).

In such game-related use cases, it is often possible to infer the quality or interestingness of the generated story using known domain semantics. For instance, if a simple generator is making a story about chess, it is easy to know which sequence of events or moves are worthy of being recited as a story, since we know the significance of every move, or sequence of moves, to the game progression or to the winning chances of each side. Similar inferences about event sequences can be made about more complex games as well, given that some game semantics are available. Moreover, even when games are not involved, many story modeling and narrative generation approaches rely on a semantic model of a particular domain (e.g. characters, goals, entity relationships, etc.) which allows the derivation of a sequence of events and ultimately a narrative, such as in (Elson, 2012a). The same is true about classic story generation systems that while inspiring, rely on a bank of previous stories and their assumed structures to generate new ones with a measure of interestingness or success, such as Minstrel (Turner, 1994) and Mexica (PÉrez & Sharples, 2001).

Other narrative generation approaches are less dependent on a particular context of communication and use case, and consequently, do not depend on a priori semantic models. Instead, they attempt to generate narrative of stories that make general sense (as a sequence of events) and contain correct sentences (if presented in text). Thus, in order to assess the quality of the generated story, such approaches often focus on the general properties and qualities of the generated text, such as coherence or the causal plausibility of the sentence ordering (Papineni, Roukos, Ward, & Zhu, 2002). This way of generating narrative is sometimes referred to as *open story generation* (Martin et al., 2017; Swanson & Gordon, 2008).

Improving on generic text-based evaluation metrics, in (Purdy, Wang, He, & Riedl, 2018), a set of proxy measures are introduced to assess the "story quality" in an open story generation task. These measures are shown in (Purdy et al., 2018) to correlate with human judgment of story quality; hence, they can be used towards a better evaluation of the generated narrative and an easier and faster fine-tuning of many generative models, such as Recurrent Neural Networks (RNNs). They include:

- Correct grammar use ("grammaticality"),
- Complexity of used language ("narrative productivity"),
- Similarity of adjacent sentences ("local contextuality"),
- Level of adherence to the usual ordering of events in most

stories, e.g. "eat" comes after "order" ("temporal ordering").

Humans possess an intuitive evaluation metric for stories, one that goes beyond linguistic measures. Expert human storytellers are not considered experts merely because of the quality of their use of language (however sophisticated it may be), but also because of their ability to tell stories that seem interesting to a large number of audience. Such experts master narrative authorship techniques and can recognize the processes involved in human's cognitive perception of narrative. In other words, they tell stories in ways that are informed by their understanding of how human perception of narrative works.

To that end, proxy measures introduced above are a useful start to assessing narrative quality when it is not tied to a specific domain of semantics. However, an important aspect of story quality, i.e. "how good a story is", depends on more complex evaluations metrics than language use, local contextuality, or the normality of the event orderings. While those measures are relevant, they do not inform the generation process about the perception of narrative. Ideally, a generator should also optimize for its generated narratives to be perceived as interesting. Moreover, as mentioned above, a computational generation of narrative heavily depends on the communication context in which it operates. A particular reason why a focus on narrative perception is imperative is the rapid evolution of such contexts, which will increasingly include interactive and sociable agents (e.g. embodied or virtual agents (Goodrich, Schultz, et al., 2008; Fong, Nourbakhsh, & Dautenhahn, 2003) or conversational agents (NPR, 2017)).

Story Interestingness

Storytelling, as an intuitive, natural and commonplace human behavior, seems deceptively simple to judge in terms of "interestingness". However, similar to some other intuitive and natural behaviors, such as nodding and gazing, it is extremely complicated to predict or reconstruct a story's interestingness. This perceived interest can be subjective, is often cultural and it can also change over time (e.g. a popular movie's narrative becomes less popular among a new generation). Moreover, the subtleties and arts of authorship makes the ways in which a narrative can seem interesting incredibly diverse, subtle and nuanced. Despite such difficulties, there are ways in which we can start understanding this phenomenon and begin developing proxy measures for perceived story interestingness, to be used in generative models. To this end, the related work in the field of cognitive science is a great resource to draw from.

While various types of interest can be established in a story, many researchers have broadly categorized these interests in two main groups. Under various names, such as *individual* and *situational* (Hidi & Baird, 1986), or *cognitive* and *emotional* (Kintsch, 1980), researchers have focused on the source of interest to make such categorization. "Cognitive" interests are largely the properties of the narrative (or authorship techniques) and "emotional" interests are largely rooted in an audience's predispositions. The latter group is more subjective, and can consist of instinctive "absolute" (Schank, 1979) interests (e.g. danger, power, sex), or "topic interests" (Campion, Martins, & Wilhelm, 2009).

While it is plausible to assume that all kinds of interest affect each other when it comes to perception, cognitive interests are categorized as the less subjective factors, ones that have a larger focus on the stimuli: the properties of the narrative. Many researchers have developed theories of the mechanisms that lead to the establishment of cognitive interest in stories. Notable theories include: unexpectedness (Schank, 1979), the interaction between background knowledge, uncertainty and postdictability (Kintsch, 1980), incongruity (Mandler, 1982), change in one's belief (Frick, 1992), generation of inference (Kim, 1999), and the generation of predictive inference (Campion et al., 2009).

Many of these theories above are conceptually close to and can overlap with each other. In this paper, we focus on two of these theories that represent familiar notions: **unexpectedness** (closely related to surprise) and **predictive inference** (closely related to foreshadowing).

A detailed overview of the theories of story interestingness is provided in (Behrooz, Mobramaein, Jhala, & Whitehead, 2018).

Search for Specificities

Another reason for creating proxy measures for story interestingness is the potential roles of such measures in choosing an appropriate set of specificities in a narrative.

Picking the Right Specificity in a Situated Context If a narrative generation system, for instance one used by an agent operating in the real world, attempts to build a narrative from events that have previously happened, there would be a search problem involved to choose which observations, details or specificities (if any) should be included in the story. At a minimum, a sequence of events can be described as a mundane narrative that minimally describes the story's events. However, the inclusion of certain specificities about the elements in the story is usually what allows for authorship skills.

The "Chekhov's Gun" principle says: "every element in a story must be necessary, and irrelevant elements should be removed." On the other hand, many seemingly unnecessary parts of a telling of a story serve the particular purpose of making the narrative more interesting (e.g. through foreshadowing or red herring techniques). For instance, specifying that "the moon was shining bright" a few events before two characters (that the audience may suspect are in love) kiss for the first time, asserts a property of the moon that is (most likely) inconsequential to what happens in the story, but is nonetheless a part of what makes the telling of it interesting.

Thus, while completely irrelevant details and specificities can violate Chekhov's Gun principle, some details and specificities, when chosen and employed in an informed and artistic way, can contribute to the interestingness of narrative when perceived by an audience.

Complementing Approaches That Involve Generalization of Concepts This search problem can also arise when generative neural networks (such as RNNs) are used to generate stories. In order to increase the chances of convergence in such models, researchers sometimes replace verbs and words in a story corpus that is used to train the model with generalized concepts (Martin et al., 2017) using semantic word networks such as VerbNet and WordNet (Schuler, 2005; Miller, 1995). This would result in the replacement of both of the words "car" and "automobile"with the semantic label "self-propelled vehicle.n.01", and consequently, it becomes easier for the model to find event patterns involving either of these words. However, the narratives generated using such models would then also include the generalized concepts, and hence, they can be more mundane and less specific as a result. Having proxy measures to find the more interesting specificities may offer a solution to this problem. In particular, word vectors can help with choosing a specific instance of a semantic label. This lack of specificity can occur in any generative method for open story generation that involves generalization of concepts or events, and consequently results in mundane generated stories, such as in (Li, Lee-Urban, Johnston, & Riedl, 2013).

Cognitive Interest as a Proxy Measure

In the absence of a domain's semantic model (as explained in previous sections), we explore the idea of using word embedding vectors with the goal of developing proxy measures for story interestingness. Word vectors introduce a way to estimate the semantic similarity and relationships between words, largely based on co-occurrence. The rapid improvements in deep learning have greatly contributed to the quality of word embeddings and they have seen much success in many computational linguistic tasks. In this paper, we investigate the use of word embeddings to estimate the cognitive interest in stories.

Foreshadowing

As briefly reviewed before, one of the main causes of the establishment of cognitive interest in stories is predictive inference by the audience (Campion et al., 2009). Among the diverse set of reasons why and ways in which a reader may try to infer what will occur in the continuation of a story, we focus on a common way in which authors attempt to intentionally cause such inference in the reader. Commonly known as *foreshadowing* (Chatman, 1980), this authorship techniques involves giving readers implicit hints that can, in various ways, provide clues about the upcoming noteworthy events in the story. Foreshadowing can have various degrees of subtly. In some cases, it can create a vivid question mark in user's mind about why a particular point is mentioned in the story (e.g. "the road seemed scary and dark, with no barriers in the middle of it"). In such cases, foreshadowing is more likely to lead to predictive inference. At other times, what is also recognized as foreshadowing may be too subtle of a hint to drive predictive inference and may not pose a question mark to the user until a later event reveals a rather cryptic connection. In both cases, the goal is for the reader to realize this connection and make sense of a "coherent macrostructure" of the story in retrospect; a notion called postdictability by Kintsch (Kintsch, 1980).

There have been a few notable attempts to generate foreshadowing in stories. Minstrel (Turner, 1994), relying on a bank of stories that it has seen before and knows about, attempts to foreshadow those upcoming events that are uncommon and hence unexpected. In (Bae & Young, 2008), another planning-based system provides solutions for generating foreshadowing and flashbacks for events that are found to be surprising. Suspenser (Cheong & Young, 2006) uses similar approaches to generate suspense in a planning-based story generation system. While our focus on cognitive interest and foreshadowing is not part of a story generation system, it can be used in one and the aforementioned system are a great source of inspiration for our work. However, as explained earlier, our focus is on systems that cannot assume the levels of semantics needed for use in planning-based systems.

Using Word Vectors to Find Foreshadowing

Estimating the presence of foreshadowing, without a semantic model of the domain, is a complicated task. Foreshadowing can take many different shapes, be causal or non-causal, and can depend on domain-specific clues. However, certain cases of foreshadowing involve usage of words that co-occur in many contexts and hence, are likely to have similar word vectors in an embedding space. This is the main intuition behind our approach.

Obtaining the Story Keywords Consider the example story in Table. 1. It contains a case of foreshadowing with a potential to cause predictive inference in the reader: event 5 (waiter is distracted and tired) foreshadows event 7 (food is wrong, waiter apologizes). Treating all the words in the story as a bag-of-words, we first remove stop words (e.g. "the", "is"), and then further narrow down our selection of words using part-of-speech tags. In order to focus on the words that capture most of the events and descriptions in the story, we select verbs, nouns and adjectives. Specifically, for verbs we use verb roots extracted via VerbNet (Schuler, 2005) and for nouns we exclude named-entities such as "Sam". It is worth noting that the current target for state-of-the-art open story generation approaches is short stories that are 6-10 sentences (Purdy et al., 2018).

Table 1: An example story which contains a case of foreshadowing. The numbers on the left are story event indexes.

1	`Sam and Judy went out for dinner at their favorite restaurant.`
2	`While driving to the restaurant, Judy's favorite song played on the radio.`
3	`Sam found a parking space at the very front of the restaurant.`
4	`Sam and Judy were seated immediately and ordered their favorite food to the waiter.`
5	`The waiter looked distracted and tired but was polite while taking their order.`
6	`Sam's favorite song played on the radio while they waited for their food.`
7	`When the waiter returned with their food it was all wrong! The waiter apologized and returned a few minutes later with the correct order.`
8	`Sam and Judy enjoyed their meal.`
9	`They paid their tab, left a tip for the waiter, and drove back home.`

Table 2 shows the keywords extracted as above for the story in Table. 1 (using Stanford CoreNLP (Manning et al., 2014) for part-of-speech tags).

Table 2: Extracted keywords from the story in Table. 1.

`waiter, return, pay, song, seat, order, radio, look, go, apologize, dinner, take, home, wrong, favorite, find, space, leave, minutes, restaurant, food, enjoy, parking, tired, drive, distracted, front, correct, meal, tip, tab, play, wait`

Vectorizing and Visualizing the Story Keywords We used GloVe embeddings, pre-trained on Wikipedia articles, in order to obtain a set of vectors that represent the words in Table 2. Hence, this set of vectors represent the major occurrences and descriptions in the story, as they map onto the embedding space at use. Moreover, by extension, these vectors can also represent major groups of concepts that are perceived by the audience when reading the story.

The original embedding space used is 300-dimensional. In order to visualize the word vectors, we used the T-SNE algorithm (Maaten & Hinton, 2008) to yield a 2-D representation of them. The results can be seen in Fig. 1.

Interpreting the Vector Space The T-SNE visualization shows to us that certain clusters of words can be distinguishable from others. These clusters can semantically categorize the contents of the story without any semantic models of the domain, such that each focus on a particular aspect of the story, involving its own events, objects and specificities. Attempting to extract some of the clusters seen in Fig. 1, we notice the following by grouping the words that are reasonably close to each other:

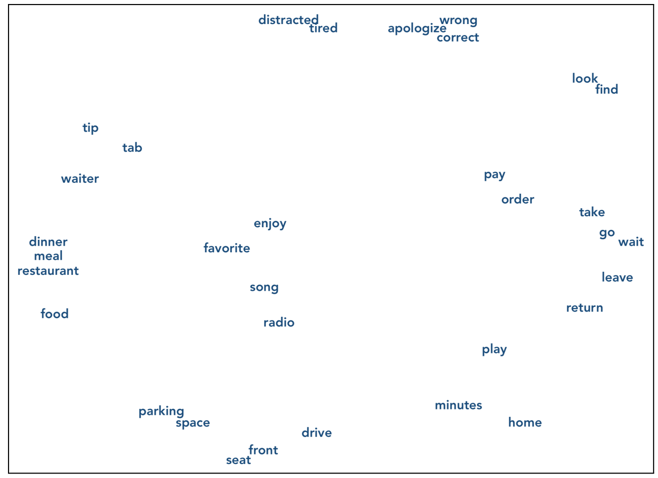

Figure 1: 2-dimensional T-SNE visualization of the GloVe vectors representing the keywords in Table 2.

- **dining:** `waiter, restaurant, dinner, meal, food, tip, tab`
- **car:** `parking, front, seat, space, drive`
- **logistics:** `return, take, wait, go, leave`
- **music:** `song, radio, enjoy, favorite`
- **cashier:** `pay, order`
- **searching:** `look, find`
- **mistake-recovery:** `distracted, tired, apologize, wrong, correct`
- **play-minute-home:** `play, minute, home`

It is also worth noting that other unsupervised clustering approaches, such as K-means, would lead to very similar clusters. We used T-SNE for this analysis because K-means proved less deterministic and could yield less predictable results depending on its starting state; however, the distance between two given pairs of word vectors is constant, hence, T-SNE depicts an appropriate representation of those constant distances.

Finding the "Key Event" Usually, a key event in a short story (or a segment of a long one) is the target of foreshadowing. In classic dramatic structures, such event can play the role of the story "climax" (Elson, 2012b). Alternatively, an "inciting incident" in the story (McKee, 1997) can become the subject of foreshadowing. Such events are often followed by a resolution (e.g. the correct food order is then brought, in our example story). Usually, this key event is unexpected, surprising, or otherwise interesting to the audience, such that it would justify the telling of the story in the first place. Finding this key event without semantic models of the story's domain is not an easy task. Most techniques employed for this purpose depend on irregularities and unexpectedness in a story. In order to find irregularities, one would need to first develop an understanding of regular progressions of the story first (without relying on a priori semantics about them). In (Behrooz, Swanson, & Jhala, 2015), for instance, sequence modeling is employed to build a model of regular event sequence in a domain, and subsequently, irregular progressions of the story and the events that cause them are identified.

In this paper, we use the *cosine similarity* of vectors representing all of the verbs in the story in order to find the most anomalous verb. Based on the above, this verb has the highest chance of being part of the key event. In Table 3, all of the roots of the verbs in the story in Table 1 are listed along with the cosine similarity metric between *each verb root vector* and the *mean of all verb root vector* in the story. This measure can indicate how close or far each verb vector is from the rest of the verbs in the story, and hence, how semantically related or unrelated.

Table 3: Verb roots of all of the verbs in the story in Table 1 (excluding stop words), along with a cosine similarity distance between each verb root vector and the mean of all verb root vectors. Verb roots are obtained using VerbNet (Schuler, 2005), and word vectors using a pre-trained GloVe model (Pennington et al., 2014).

Verb root	Cosine similarity
`go`	0.838
`drive`	0.517
`play`	0.609
`find`	0.734
`seat`	0.416
`order`	0.566
`look`	0.698
`take`	0.839
`favorite`	0.458
`wait`	0.697
`return`	0.697
`apologize`	**0.335**
`enjoy`	0.57
`pay`	0.631
`leave`	0.744

As we can see in the Table 3, the verb `apologize` is the most anomalous verb in our example story, since it has the lowest cosine similarity score with the mean of all verb root vectors. We identify this verb as the *key verb* in the story, and since the key verb is mentioned in event 7 (in Table 1), we also identify that event as the *key event* in the story.

Finding the Foreshadowing Cluster Given the key event and key verb, as described above, we can use the keyword clustering of the story, seen in 1, to find out if there exists a cluster whose constituent keywords:

1. play a role in the key event and include the key verb, and,
2. play a role in one other preceding event (or sentence) in the story.

With such constraints considered, we can see that the **mistake-recovery** can be the *foreshadowing cluster*; a cluster that includes the words involving the foreshadowing in the story.

Finding the Foreshadowing The preceding event or sentence in the story in Table 1, in which the foreshadowing cluster plays a role, is event 5. Hence, we can guess that event 5 has a chance of foreshadowing our key event, 7. Moreover, as a whole, these steps can result in an estimate of the presence of foreshadowing in the story.

Unexpectedness

As mentioned before, many approaches to open story generation focus on finding the usual progressions of events in the story. Among such approaches are story scripts (Schank & Abelson, 2013) which argue that plots about many domains of storytelling usually follow a similar general pattern. Another example are Plot Graphs (Li et al., 2013), which use crowdsourcing to build networks of usual progressions and precedence rules of events (e.g. a graph covering many of the usual paths that a "dining at a restaurant" story would cover). In (Purdy et al., 2018), using a corpus of movie plot summaries, a temporal ordering network is created to capture the common ordering of verbs in stories. The resulting proxy measure, introduced earlier as "temporal ordering", is then used to find the extent to which a new sequence of events adheres to the common ordering of events in stories.

While such adherence would help estimate a correct causal chain of events or logical precedence between them, it is noteworthy that one of main reasons for cognitive interest in stories is the *unexpectedness* of events (Schank, 1979). Hence, as a story generator would benefit from a proxy measure for correct temporal ordering of events, it may also benefit from one that rewards it for having some unexpected event.

"The Inverted-U Function" Kintsch (Kintsch, 1980) argues that cognitive interest can be an "inverted-U" function of knowledge and uncertainty about the story. Simply described, this view argues that if a story creates too many or too few question marks in user's mind, it is less likely to be perceived as interesting. This guides us towards a proxy measure that can have a higher value if a story deviates in small amounts from the usual ordering of events, and a lower

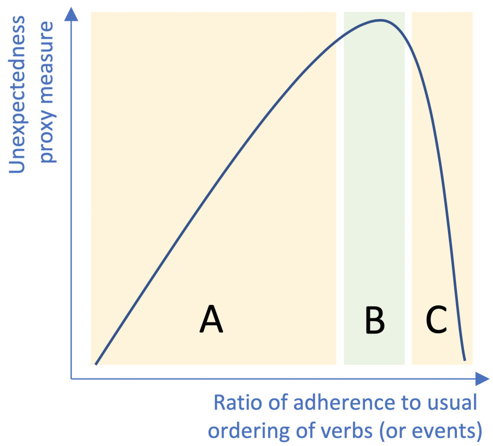

Figure 2: An illustration of a cognitive interest proxy measure based on unexpectedness and inspired by Kintsch arguments (Kintsch, 1980). The area marked as *A* denotes a story that does not sufficiently adhere to the usual ordering of verbs (or events). *C* shows an area where there is no or too little deviation from the usual ordering for the story to cause cognitive interest. *B* shows an area indicating that the story generally adheres to the usual ordering, but contains enough deviations and hence may cause cognitive interest.

value if it deviates too much from (or does adheres at all to) the usual ordering. Using a temporal ordering network, for instance, an unexpectedness proxy measure can have its highest value if most but not all (e.g. 90%) of the pairs of verbs in the story adhere to the network's order. The proxy measure would sharply decrease if this adherence ratio is much less, or approaches 1. An illustration of such proxy measure function can be seen in Fig. 2.

Unexpectedness and Word Vectors Using a vector space that represents verbs (or sentences (Pagliardini, Gupta, & Jaggi, 2017)) in a story, the distance between each vector and the average of all vectors belonging to a story (similar to Table 3) can estimate how unexpectedly each verb is perceived compared to the rest of the story. Hence, in order to follow an inverted-U pattern, a proxy measure of unexpectedness can have the highest value when most entries in Table 3 have large values, but at least one entry has a much lower value than others.

Conclusion

Communication context is a consequential factor in narrative generation, in terms of approach, constraints, and evaluation criteria. Certain narrative generation approaches are tied to a specific communication context (e.g. games) and depend on that context's a priori semantics to evaluate how good a generated story is. Other approaches are not bound to a specific context (called *open story generation*) and often use generic text quality measures to assess the quality of the story. Given the importance of narrative perception in real-world use cases of such story generation (e.g. by an intelligent agent), we draw from theories of cognitive interest and investigate the use of word embeddings vectors to find how interesting a generated narrative is. Specifically, we assess the existence of predictive inference (through foreshadowing) and unexpectedness in stories, using GloVe word vectors (Pennington et al., 2014). We plan to evaluate this approach in a situated scenario and seek to find correlations between proxy measures of cognitive interest and judgments of human subjects.

References

Bae, B.-C., & Young, R. M. (2008). A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In *Joint international conference on interactive digital storytelling* (pp. 156–167).

Behrooz, M., Mobramaein, A., Jhala, A., & Whitehead, J. (2018). Cognitive and experiential interestingness in abstract visual narrative. *Cognitive Science Society (CogSci)*.

Behrooz, M., Swanson, R., & Jhala, A. (2015). Remember that time? telling interesting stories from past interactions. In *Interactive storytelling*. Springer.

Campion, N., Martins, D., & Wilhelm, A. (2009). Contradictions and predictions: Two sources of uncertainty that raise the cognitive interest of readers. *Discourse Processes*, *46*(4), 341–368.

Chatman, S. B. (1980). *Story and discourse: Narrative structure in fiction and film*. Cornell University Press.

Cheong, Y.-G., & Young, R. M. (2006). A computational model of narrative generation for suspense. In *Aaai* (pp. 1906–1907).

Elson, D. K. (2012a). Detecting story analogies from annotations of time, action and agency. *Proceedings of the Third Workshop on Computational Models of Narrative*(1981), 91-99.

Elson, D. K. (2012b). *Modeling narrative discourse*.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3-4), 143–166.

Frick, R. W. (1992). Interestingness. *British Journal of Psychology*, *83*(1), 113–128.

Goodrich, M. A., Schultz, A. C., et al. (2008). Human–robot interaction: a survey. *Foundations and Trends® in Human–Computer Interaction*, *1*(3), 203–275.

Hidi, S., & Baird, W. (1986). Interestingnessa neglected variable in discourse processing. *Cognitive Science*, *10*(2), 179–194.

Kim, S.-i. (1999). Causal bridging inference: A cause of story interestingness. *British Journal of Psychology*, *90*(1), 57–71.

Kintsch, W. (1980). Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics*, 9(1-3), 87–98.

Li, B., Lee-Urban, S., Johnston, G., & Riedl, M. (2013). Story generation with crowdsourced plot graphs. In *Aaai.*

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, *9*(Nov), 2579–2605.

Mandler, G. (1982). The structure of value: Accounting for taste. *Center for Human Information Processing Report*, *101*.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55–60).

Martin, L. J., Ammanabrolu, P., Wang, X., Hancock, W., Singh, S., Harrison, B., & Riedl, M. O. (2017). Event representations for automated story generation with deep neural nets. *arXiv preprint arXiv:1706.01331*.

Mateas, M., & Stern, A. (2003). Façade: An experiment in building a fully-realized interactive drama. In *Game developers conference* (Vol. 2).

McCoy, J., Treanor, M., Samuel, B., Mateas, M., & Wardrip-Fruin, N. (2011). Prom week: social physics as gameplay. In *Proceedings of the 6th international conference on foundations of digital games* (pp. 319–321).

McKee, R. (1997). *Substance, structure, style, and the principles of screenwriting*. New York: HarperCollins.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41.

NPR, E. R. (2017). The smart audio report [Computer software manual]. Retrieved from `https://www.nationalpublicmedia.com/smart-audio-report/` (Accessed: 2018-07-20)

Pagliardini, M., Gupta, P., & Jaggi, M. (2017). Unsupervised learning of sentence embeddings using compositional n-gram features. *arXiv preprint arXiv:1703.02507*.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 311–318).

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).

PÉrez, R. P. Ý., & Sharples, M. (2001). Mexica: A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence*, *13*(2), 119–139.

Purdy, C., Wang, X., He, L., & Riedl, M. (2018). Predicting generated story quality with quantitative measures.

Riedl, M. O., & Bulitko, V. (2012). Interactive narrative: An intelligent systems approach. *AI Magazine*, *34*(1), 67.

Ryan, J., Mateas, M., & Wardrip-Fruin, N. (2016). Characters who speak their minds: Dialogue generation in talk of the town. *Proc. AIIDE*.

Schank, R. C. (1979). Interestingness: controlling inferences. *Artificial intelligence*, *12*(3), 273–297.

Schank, R. C., & Abelson, R. P. (2013). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press.

Schuler, K. K. (2005). Verbnet: A broad-coverage, comprehensive verb lexicon.

Swanson, R., & Gordon, A. S. (2008). Say anything: A massively collaborative open domain story writing companion. In *Interactive storytelling* (pp. 32–40). Springer.

Turner, S. R. (1994). Minstrel: A computer model of creativity and storytelling.