A model of suspense for narrative generation

Richard Doust
Open University, UK
richard.doust@gmail.com

Paul Piwek
Open University, UK
paul.piwek@open.ac.uk

Abstract

Most work on automatic generation of narratives, and more specifically suspenseful narrative, has focused on detailed domain-specific modelling of character psychology and plot structure. Recent work on the automatic learning of narrative schemas suggests an alternative approach that exploits such schemas for modelling and measuring suspense. We propose a domain-independent model for tracking suspense in a story which can be used to predict the audience's suspense response on a sentence-by-sentence basis at the content determination stage of narrative generation. The model lends itself as the theoretical foundation for a suspense module that is compatible with alternative narrative generation theories. The proposal is evaluated by human judges' normalised average scores correlate strongly with predicted values.

1 Introduction

Research on computational models of narrative has a long tradition, with important contributions from researchers in natural language generation, such as the AUTHOR system (Callaway and Lester, 2002), which provides the blueprint for a prose generation architecture including a narrative planner and organiser with traditional NLG pipeline components (sentence planner, realiser) and a revisor.

Two features are found in much of such research. Firstly, work on content determination and organisation has often centred on use of detailed representations for character goals and plans – see for example Cavazza et al. (2002) and Cavazza and Charles (2005). According to such approaches, a good and, more specifically, a suspenseful story should arise out of the complex interaction of the system components. It is often hard to separate the different contributions of the system choices from the quality of the underlying domain-specific plans and story templates (Concepción et al., 2016).

Secondly, existing approaches to suspense often interlock with the concept of a story protagonist under some kind of threat. For example, the suspense modelling in the SUSPENSER system Cheong and Young (2015) - which aims to enable choices that can maximise suspense in narrative generation - is entirely based on Gerrig and Bernardo (1994)'s definition, according which suspense varies inversely with the number of potential actions of the central protagonist which could allow him or her to escape a threat. Similarly, Zillman's definition (Zillmann, 1996) links suspense to the reader's fearful apprehension of a story event that threatens a liked protagonist. Delatorre et al. (2016) proposed a computational model based on Zillman's definition from which they derived the use of emotional valence, empathy and arousal as the key components of suspense. Interestingly, their experimental results led the authors to question the usefulness of empathy for measuring suspense.

In this paper, we approach the problem of suspenseful story generation from a different angle. Our main contribution is a domain-independent model of suspense together with a method for measuring the suspensefulness of simple chronological narratives. Thus we identify a separately testable measure of story suspensefulness. By separating out emotional salience and character empathy considerations from informational and attentional processes at the heart of the suspense reaction, we construct a modular definition of suspense that could in theory encompass the definitions cited above. The method builds on the psychological model of narrative proposed by Brewer

Proceedings of The 10th International Natural Language Generation conference, pages 178–187,
Santiago de Compostela, Spain, September 4-7 2017.

and Lichtenstein (1982), and was first developed in Doust (2015).

Inspired by Brewer and Lichtenstein's informal model, we build a formal model in which the concept of a *narrative thread* plays a pivotal role. Narrative threads model the reader's expectations about what might happen next in a given story. As a story is told, narrative threads are activated and de-activated. Different threads may point to conflicting events that are situated in the future. As more of the story is revealed, the moment of resolution of the conflict may appear more or less proximal in time. We capture this by formally defining the concept of *Imminence*. Imminence is based on the potential for upcoming storyworld events to conflict with one another and on the narrative proximity with these conflicts. It is the key factor in what we call *conflict-based suspense*. Additionally, as the story is told, conflicting *interpretations* about certain events in the story may prevail. This leads us to define a distinct second type of suspense which we call *revelatory suspense*.

Our approach has a number of advantages. Firstly, by disentangling suspense from story protagonist-based modelling, our approach can deal with scenes that have no discernable human-like protagonist. An example could be an ice floe slowly splitting up or a ball slowly rolling towards the edge of a table[1]. Thus the empirical coverage of the theory is extended.

Secondly, the model lends itself as the theoretical foundation for a suspense module that is compatible with alternative narrative generation theories. Such a module could be used to evaluate story variants being considered in the search space of a narrative generation program. This is possible because the model operates at a level of abstraction where character motivations are subsumed by higher level properties of an unfolding story, such as imminence.

Thirdly, the underlying world knowledge that our model relies on, makes use of a format, narrative threads, which meshes with recent work in computational linguistics on the automatic learning of narrative schemas as proposed in Chambers and Jurafsky (2009) and Chambers and Jurafsky (2010).

The remainder of this paper is organised as follows. In Section 2 we discuss previous primarily computational work on suspense. The section concludes with a description of Brewer and Lichtenstein's psychological theory of suspense, which forms the basis for our model. Section 3.1 introduces our formal model. Section 3.2 presents our model of the reader's response to a suspenseful narrative and the algorithm for calculating suspense. Section 3.3 presents an overview of *revelatory suspense*. Section 4 reports on the evaluation of the model by human judges. Finally, in Section 5 we present our conclusions and avenues for further research.

2 Related work on computational models of narrative and suspense

In computational models of narrative, a common approach is to determine some basic element, which, when manipulated in certain ways, will produce a skeletal story-line or plot. For example, TALE-SPIN (Meehan, 1977) uses the characters' *goals*, whereas MINSTREL (Turner, 1992) uses *both* authorial and character goals. MEXICA (Pérez y Pérez and Sharples, 2001) uses a *tension curve* to represent *love*, *emotion* and *danger* in order to drive the generation process. Riedl and Young (2010) and later the GLAIVE narrative planner (Ware and Young, 2014) introduce a novel refinement search planning algorithm that combines intentional and causal links and can reason about character intentionality.

The focus in the aforementioned work is on the *global* story-modelling task and the automatic generation of new narratives. Suspense is seen as one of a set of by-products of story generation. There is no re-usable model of what makes a suspenseful story.

Other approaches have been more specifically aimed at generating *suspenseful* stories. Cheong and Young (2015)'s SUSPENSER and O'Neill and Riedl (2014)'s DRAMATIS propose cognitively motivated heuristics for suspense using the planning paradigm. Characters have goals and corresponding plans, and suspense levels are calculated as a function of these. The measurement of suspense in these algorithms was evaluated using alternate versions of a story. In particular, O'Neill and Riedl (2014) asked human judges to make a decision about which story was more suspenseful and then compared these results to the story their system identified as most suspenseful.

Several psychological theories of narrative under-

[1] One could postulate the existence of imaginary protagonists for such scenarios, but this seems unnecessarily complex.

standing have attempted to approach suspense modelling. For example, Kintsch (1980) considers the schemata and frames that readers call upon to actually learn from the text they are reading. They examine the additional focus generated by expectation violations, i.e., surprise rather than suspense.

Our suspense measurement method is grounded in Brewer and Lichtenstein (1982)'s psychological theory of narrative understanding. They suggest that three major discourse structures account for the 'enjoyment' of a large number of stories: surprise, curiosity and suspense. This approach is based on the existence of Initiating Events (*IE*) and Outcome Events (*OE*) in a given narrative.

For suspense, an *IE* is presented which triggers the prediction of an *OE* which corresponds to a significant change in the state of the storyworld. The reader feels concern about this outcome event, and if this state is maintained over time, the feeling of suspense will arise. Such a change in the state of the storyworld can have a positive or negative valence for the reader, and may often be significant because it concerns the fate of a central character in the story. However, this link to a character's fate is not a requirement of our model.

Also, as Brewer & Lichtenstein say: 'often additional discourse material is placed between the initiating event and the outcome event, to encourage the build up of suspense' (Brewer and Lichtenstein, 1982, p. 17). Thus, to produce suspense, the *IE* and *OE* are ordered chronologically and other events are placed between them.

We propose a computational extension of this model based on what we call *narrative threads*, a concept grounded in psychological research such as Zwaan et al. (1995), the constructionist and prediction-sustantiation models of narrative comprehension (Graesser et al., 1994) and scripts (Lebowitz, 1985; Schank and Abelson, 1977). Our narrative threads include both causal and intentional links as do for example Ware and Young (2014) and also include the concept of *recency* (see for example Jones et al. (2006)).

The research reported in this paper differs in two key respects from the computational approaches described above.

Firstly, it eschews a planning approach to story generation that makes use of detailed modelling of character intentions and goals. In our view, suspense is *not dependent* on the existence of characters' goals: we can experience suspense about a piece of string breaking under the strain of a weight. Nor does our approach require the existence of a central protagonist and his or her predicament.

Secondly, our model tracks suspense throughout the telling of a story. Unlike much previous work, rather than evaluate only the predicted *overall* suspense level of a story, in our evaluation we compare predicted suspense levels with human judgements at *multiple steps* throughout the telling of the story. This provides us with a much more fine-grained evaluation of our suspense measurement method than has hitherto been used.

Brewer and Lichtenstein's work has been the basis of further work such as Hoeken and van Vliet (2000) and Albuquerque et al. (2011). The former found that 'suspense is evoked even when the reader knows how the story will end' whereas the latter explored story-line variation to evoke suspense, surprise and curiosity. However, neither presents a model of how suspense fluctuates during the telling of a story nor any empirical evaluation of such a model.

In the following section, we present our computational model of suspense that extends that of Brewer and Lichtenstein (1982).

3 Formalism and Algorithm for Suspense Generation

3.1 Formalism

A *story* can be considered the work of a hypothetical author, who first chooses some events from a *storyworld* and orders them into a *fabula* (this includes all relevant events from the storyworld, not just those that get told). The author then chooses events and orderings of events from this fabula to create a story designed to trigger specific reactions from its readers.

3.1.1 A storyworld

A storyworld $\mathbb{W} = (\mathbb{E}, \mathbb{N}, \mathbb{D})$ is made up of the following elements:

- $\mathbb{E}$, the set of possible events,
- $\mathbb{N}$, the set of narrative threads. Each narrative thread $Z \in \mathbb{N}$ consists of a fixed sequence of distinct events chosen from the set $\mathbb{E}$ and an *Importance* value, $\mathrm{Value}(Z)$,

- $\mathbb{D}$, the set of ordered pairs (a, b) of disallowing events where $a, b \in \mathbb{E}$ and a disallows b.

We will be dealing in this research only with *chronological stories*. For a given set of narrative threads, a story will satisfy the chronological constraint if and only if:

> For all pairs of events a and b where a precedes b in the story, if there *are* any narrative threads in which both a and b occur, then in at least *one* of these threads a precedes b.

Using the chronological qualities of narrative threads, we can now define a *fabula* as a chronologically ordered list of *n* events chosen from $\mathbb{E}$, the set of possible events in the storyworld $\mathbb{W}$. We define a *story* as an ordered list of events chosen from a given *fabula*. In the general case, a story for a fabula can reorder, repeat or skip any of the fabula's events. Because we are only dealing with chronological stories, in our current model, the only allowable difference between a fabula and a story is that some elements of the fabula can be skipped.

We now give two constraints on fabulas for a given storyworld $\mathbb{W} = (\mathbb{E}, \mathbb{N}, \mathbb{D})$. An (optional) completeness relation between the set of events, $\mathbb{E}$, and the set of narrative threads, $\mathbb{N}$, is a useful constraint to include in most storyworlds. It excludes the possibility that an event in a fabula has no narrative thread which contains it, thus avoiding the situation where an event is 'uninterpretable' in storyworld terms.

Concerning $\mathbb{D}$, the set of disallowing event-pairs, $(a, b) \in \mathbb{D}$ means that if *a* is told, then *b* is predicted *not* to occur in storyworld $\mathbb{W}$, or we can also say, *b should not* be one of the subsequent events to be told. We will therefore require that no event that is a member of a fabula disallows any other[2].

3.1.2 Telling the story

Telling a story is equivalent to going through an ordered list of events one by one. To 'tell an event in the story', we take the next event from a list of *Untold* events and add it to the tail of a list of *Told* events. Each new told event may have an effect on one or more narrative threads.

[2]This is in fact a transposition of the constraints used in the GLAIVE narrative planner (Ware and Young, 2014).

Each narrative thread also has a *Conveyed* and *Unconveyed* event list. Events in a thread become *conveyed* in two cases: when they are *told* in a story, or when they are *presumed* to have occurred in the storyworld because in some thread they *precede* an event which has been told (see also 3.1.3).

If the new story event matches a member of the *Unconveyed* list of any narrative thread, then we move it (and all the events before it) into the thread's *Conveyed* list. Additionally, the thread also becomes active (if previously, it was not).

Finally, certain threads may be deactivated by the new story event. Any active narrative thread with an event α in its Unconveyed list will be deactivated if an event γ is told in the story and $(\gamma, \alpha) \in \mathbb{D}$.

For each thread Z, we designate $\mathrm{state}(Z)$ which indicates both whether Z is *active* or *inactive*, which events in Z have been conveyed, and which are as yet unconveyed. Before the story starts to be told, all narrative threads are *inactive* and all their events are in their respective *Unconveyed* lists. *Inactive* threads always have this form and have no effect on suspense calculations. When the last event in a narrative thread Z gets conveyed in the story, we can say: 'Z succeeds'.

3.1.3 Implicated events

An *implicated prior event* is any event in the Conveyed list of some active narrative thread that has *not* been told in the story, but is 'presumed to have occurred'. If α and γ are implicated prior events (in different active threads), and $(\gamma, \alpha) \in \mathbb{D}$, then α is a *conflicted* implicated prior event. In a similar way to implicated *upcoming* events, implicated prior events in different threads may remain in conflict with each other over several story steps. A *conflicted thread* is a thread whose Conveyed list contains at least one conflicted implicated prior event.

An *implicated upcoming event* is just any member of the *Unconveyed* list of an active thread. Such an event is predicted to be told in the current story with a confidence level that depends on the confidence we have in the narrative thread of which it is a member. It is conflicts between implicated upcoming events that create suspense.

3.1.4 Confirmed and unconfirmed threads

Active threads may be confirmed or unconfirmed. An active *confirmed* thread is any thread whose *Conveyed* list contains at least one told event. Active *unconfirmed* threads with *no told events* are important in our system because they allow for a degree of flexibility in the linking together of different narrative threads. Thus, an inactive thread which shares at least one event with some other active thread can become *active* but *unconfirmed* for the purposes of suspense calculation. For example, a set of threads which detail the different things that someone might do when they get home can under this rule be activated before the story narrates the moment when they open their front door. We can formalise this in the following way:

> An inactive thread Z can become an *active unconfirmed thread* if any of its (unconveyed) events appears in the Unconveyed list of some other *active* confirmed thread (and as long as it has no event that is disallowed by some told event).

Thus, in such a case, an inactive (and thus unconfirmed) thread Z becomes active even though none of its events have yet been told in the story. We can say that 'the confirmation of thread Z is predicted'.

3.2 Modelling the reader's predicted reactions: the suspense algorithm

For each narrative thread, we first determine the following intermediate values: *Imminence, Importance, Foregroundedness*, and *Confidence*. We then combine these values to calculate the suspense contribution from each *individual* narrative thread. Finally we propose a heuristic to combine all these individual narrative thread suspense values and produce the *global* suspense level for each moment in the story.

3.2.1 Imminence

Each active narrative thread Z generates two values for Imminence. Completion Imminence is related to the number of events in Z still to be conveyed for it to be completed or to 'succeed'. Figure 1 shows a thread with a Completion Imminence number of 4.

Interruption Imminence is related to the smallest number of events still to be conveyed in some other thread Y before an event is told which can interrupt Z by disallowing one of its events. In the case where no thread can interrupt Z, the Interruption Imminence of Z is zero. In Figure 2, thread A has an Interruption Imminence number of 3 due to thread B.

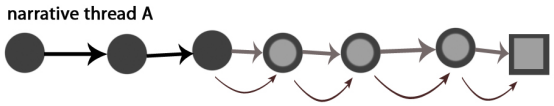

Figure 1: Completion Imminence
Conveyed events are black, *unconveyed* grey.

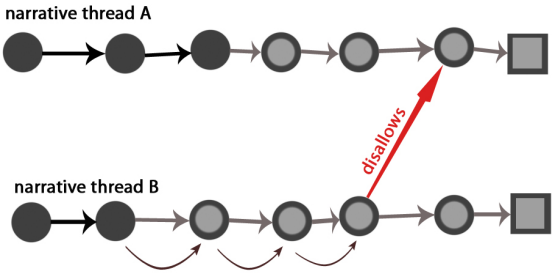

Figure 2: Interruption Imminence

If a *large* number of events must be told for a thread to be completed, the Imminence is *low*, and vice versa. To model this behaviour, we adopted the ratio $1/x$. Also, to enable exploration of the relative effects of Completion and Interruption Imminence on this measure, we used a factor ρ to vary the relative weighting of these two effects.

We can now give a first definition of the Total Imminence$_n(Z)$ of a narrative thread Z after the n^{th} event in the story.

$$\text{Total Imminence } = \rho\frac{1}{H} + (1-\rho)\frac{1}{R} \qquad (1)$$

where H is the number of events to the completion of Z and R is the minimum number of events before an event in some other narrative thread could be told which would disallow some unconveyed event in Z.

Experimentation with the implementation of our model led us to choose $\rho = 0.7$, in effect boosting the relative effect of *Completion imminence*.

3.2.2 Foregroundedness

We use a parameter called *Foregroundedness* to represent how present a given narrative thread is in the reader's mind. This is similar to the concept of

recency in the psychological literature[3]. The Foregroundedness of each narrative thread changes with each new event in the story and varies between 0 and 1. Threads which *contain* the current story event are considered to be very present and get ascribed the maximum level of Foregroundedness, that is 1.

In our current model, the Foregroundedness of all *other* narrative threads is simply set to decrease at each story step due to the following decay function:

$$\text{decayFunction}(x) = \beta x, \text{where } 0 < \beta < 1 \quad (2)$$

Experimentation led us to use $\beta = 0.88$.

3.2.3 Confidence

Depending on the number of its conflicted prior events, a thread will have varying degrees of *Confidence* as the story progresses. In Figure 3, we show narrative threads *A* and *B* that share an event, the event which has just been told in the story.

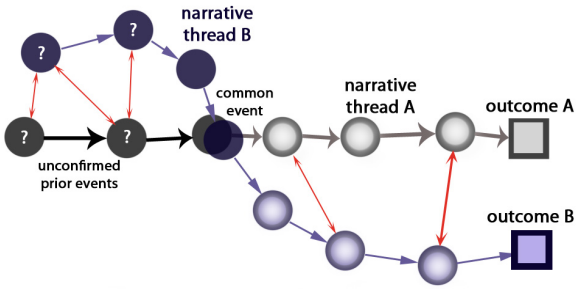

Figure 3: Threads with a shared event. *Implicated prior events* have a question mark. Bidirectional arrows show *mutual disallowing* relations.

We see that two implicated prior events in *A* are in conflict with implicated prior events in *B*. Overall then, thread *A* has *two* conflicted prior events and *one* confirmed event. Narrative threads with many *conflicted* prior events will have a low Confidence level, reducing their potential effect on suspense. We define the $\text{Confidence}_n(Z)$ of a narrative thread Z after the n^{th} event in the story as follows:

$$\text{Confidence} = \frac{1}{(1 + \frac{\phi Q}{P})} \text{where } \phi = 1.5 \quad (3)$$

where P is the (non-zero) number of *told events* and Q the number of *conflicted implicated prior events* in Z and $0 \leq \text{Confidence} \leq 1$. Note that if the threads containing events conflicting with Z get deactivated, then Z may come to no longer have *any* conflicted prior events. In such a case, as long as Z has at least *one* confirmed event, its confidence level would reach the maximum value of 1. In other words, if $P > 0, Q = 0$, then Confidence $= 1$. Empirical work on our implementation led us to use a 'conflicted-to-told ratio' of $\phi = 1.5$.

3.2.4 Importance

We next define $\text{Value}(Z)$ as the measure of the *Importance* of a narrative thread Z.

> $\text{Value}(Z)$ = the predicted degree of positive or negative appraisal of the storyworld situation that the reader would have, were Z to succeed.

In our model, we use the range $(-10, +10)$ for this value, where -10 and $+10$ correspond to events about which the reader is very negative (sad, dissatisfied) and very positive (happy, satisfied) respectively.

3.2.5 Our suspense algorithm

After the telling of the n^{th} story event, we calculate the $\textit{Imminence}_n(Z)$, $\textit{Foregroundedness}_n(Z)$, $\textit{Confidence}_n(Z)$ and $\textit{Importance}_n(Z)$ for each active narrative thread Z. For the general case, we assumed that all four variables were independent and chose *multiplication* to combine them and create a measure of the suspense contribution of each narrative thread after story event n:

$$\begin{aligned} \text{Suspense}_n(Z) = \quad & \text{Imminence}_n(Z) \\ & \times \text{Importance}_n(Z) \\ & \times \text{Foregroundedness}_n(Z) \\ & \times \text{Confidence}_n(Z) \end{aligned} \quad (4)$$

Once we have the suspense level of each active thread, we assume that the thread with the *highest suspense value* is the one that will be responsible for the story's evoked suspense at that point. We therefore define this thread's suspense value as equivalent to the suspense level of the narrative as a whole at that point in the story.

3.3 Revelatory suspense

As well as the general case of conflict-based suspense described above, our thread-based model allows us

[3] See for example, Jones et al. (2006).

to deal with a type of suspense we call *revelatory suspense*, or *curiosity-based* suspense. This kind of suspense is linked to the potential disambiguation of a story event that belongs to several narrative threads. There is suspense about which thread will provide the 'correct' interpretation of the event.

To understand this, we can imagine that in a given storyworld, event δ is present in several different threads. When δ is told in the story, *several* threads become activated as candidates to uniquely 'explain' it. Subsequent story events may disallow some candidate threads. Exactly which thread turns out to be the correct 'explanation' of δ in the storyworld will be determined by the rest of the story. In Figure 3 already mentioned, there is therefore revelatory suspense about which of threads A and B will remain active to explain the shared event. Revelatory suspense is potentially present as soon as the storyworld has threads with shared events.

We can contrast this disambiguation process with Cheong and Young (2015)'s SUSPENSER system, where suspense varies inversely with the number of possible actions of a central protagonist. Similarly, because the decrease in conflicted events boosts the thread's Confidence level, in our model, the suspenseful effect of a thread will go up as ambiguity is reduced. However, in the SUSPENSER system, the suspense depends on the *protagonist's* options, whereas in our model, it depends on the reduced number of options for the *reader*.

4 Evaluation

We implemented our formal model and suspense metric computationally. To test our implementation, we designed and wrote a short suspenseful story, the *Mafia story*, where an important judge drives towards his home with a bomb ticking in his car. The story was inspired by the story used in Brewer and Lichtenstein (1982)'s experiment.[4]

4.1 Original story and storyworld calibration

First we used Zwaan's protocol (Zwaan et al., 1995) to split the story into separate events each time there was a significant change in either time, space, interaction, subject, cause or goal.

The next step was to create the storyworld information. Our model is designed to rely on information in a form which could be generated automatically from real-world data or corpora. The actual generation of this information lay however outside the scope of this research. We therefore created the events, $\mathbb{E}$, the narrative threads $\mathbb{N}$, their importance values $\mathrm{Value}(Z)$ and the set of disallowing events $\mathbb{D}$ by hand, partly modelling our work on the event chains described in Chambers and Jurafsky (2009).

We then conducted an experiment to calibrate the importance values and check the validity of the events in our hand-made narrative threads. The online interface created for the experiment presented a warm-up story and then the *Mafia story* to the participants (N=40 for 33 story steps, 1320 individual judgements), recording step-by-step self-reported suspense ratings using magnitude estimation (see for example Bard et al. (1996)). The raw suspense ratings were converted to normalised z-scores.

Next, we used our suspense algorithm to produce suspense level predictions for all the steps in the Mafia story. Once we had obtained both predicted and experimental values for suspense levels in the Mafia story, we examined their degree of match and mismatch for different sections of the story. We then adjusted some importance values and made minor modifications to some narrative threads.

4.2 Variant of the original story

Next, we created the *Mafia-late* story variant, which differed only from the original (henceforth) *Mafia-early* story in that the vital information suggesting the presence of a bomb in the judge's car is revealed at a later point in the story. Apart from this change in the event order, we strove to create as realistic a story as possible that used exactly the same events.

With the calibrated importance values from the first study, we then used our implementation to create new predictions for this story variant. For comparison, we show these together with the original *Mafia-early* predictions in Figure 4[5].

We then collected human suspense level judgements (N=46 for 31 steps, 1426 individual judge-

[4]Both the PROLOG implementation and story variants are available at `https://doi.org/10.6084/m9.figshare.5208862`.

[5]The *Mafia-late* story has two events fewer than the *Mafia-early* story. To facilitate comparison, we have aligned the *Mafia-late* and *Mafia-early* results so that as far as possible the same events occur at the same point on the x-axis.

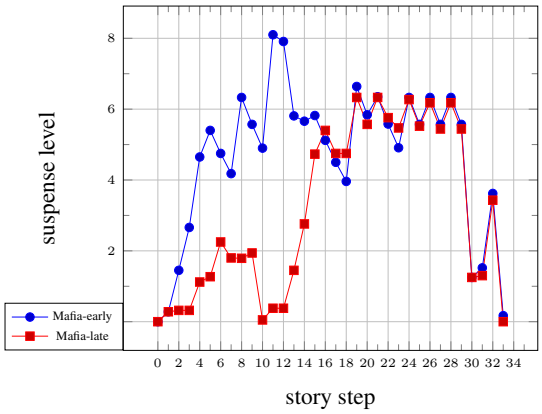

Figure 4: Predicted suspense for the story variants. The Mafia-early and Mafia-late stories mention the bomb at steps 4 and 15 respectively.

ments) for the *Mafia-late* story. We predicted that the suspense levels calculated by our calibrated model for the *Mafia-late* story variant would agree with the step-by-step averaged z-scores of ratings for this story given by a new set of participants.

4.3 Results and Statistical Analysis

The magnitude estimation ratings obtained for each participant were first converted to z-scores. For each story step, we then calculated the mean and standard deviation of the z-scores for all participants which, for comparison, we present together with the predicted values in Figure 5.

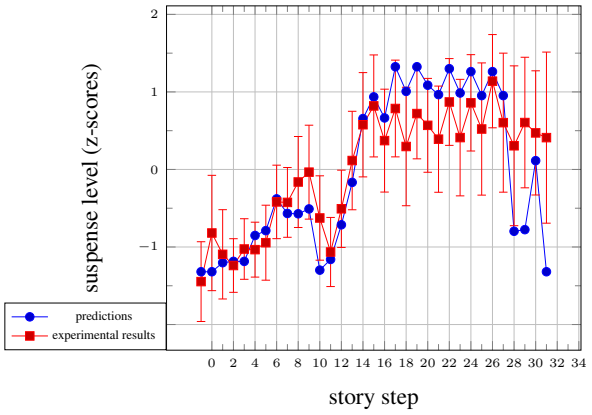

Figure 5: Experimental and predicted suspense for the Mafia-late story

For these two curves, the *Pearson Correlation Coefficient* is 0.8234 and the *Spearman's Rho Coefficient* is 0.794, both values indicating a strong positive correlation. However, the vertical standard deviation values for the z-scores are large, suggesting large variations between participants' responses. To check inter-coder reliability, we performed Fleiss' Kappa test and achieved a value of 0.485. Landis and Koch (1977) interprets this value as signifying *moderate* inter-coder agreement. Levels of agreement as measured by Fisher's test between predicted and experimental suspense levels for our story-variant show *highly significant* success in prediction (P=0.002).

5 Conclusions and further work

We believe that narrative and more specifically suspense is an important topic of study. As discussed in Delatorre et al. (2016), suspense is a pervasive narrative phenomenon that is associated with greater enjoyment and emotional engagement.

In this paper, we describe a formal model of suspense based on four variables: Imminence, Importance, Foregroundedness and Confidence together with a method for measuring suspense as a story unfolds. The model enabled us to predict step-by-step fluctuations in suspensefulness for a short story which correlate well with average self-reported human suspense judgements.

Our method for obtaining suspense judgements was intrusive and may have created some interference with the reading process. Ideally, future research should explore less intrusive methods that use direct physiological measurements of the participants. So far, the search for measurement methods that correspond with perceived suspense has been unsuccessful (Cheong and Young, 2015).

A key difference with previous work is that our model predicts and evaluates suspense at multiple stages *within* the same story. This is why we focussed on variations in one storyworld instead of a large corpora of stories. Indeed, our goal is to create the first model of the suspense evoked as a narrative is being received, and not just a single overall suspense rating. Also, instead of starting from character goals and plans, our basic construct is the narrative thread which is akin to narrative schemas that can be harvested automatically (Chambers and Jurafsky, 2009). In future work, we aim to apply our method to such automatically harvested schemas and extend our model to different storyworlds and story variants.

References

Alexandre C Albuquerque, Cesar Tadeu Pozzer, and Angelo EM Ciarlini. 2011. The usage of the structural-affect theory of stories for narrative generation. In *Games and Digital Entertainment (SBGAMES), 2011 Brazilian Symposium on Games and Digital Entertainment*, pages 250–259. IEEE.

Ellen Gurman Bard, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language*, 72(1):32–68.

W.F. Brewer and E.H. Lichtenstein. 1982. Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics*, 6(5-6):473–486.

Charles B Callaway and James C Lester. 2002. Narrative prose generation. *Artificial Intelligence*, 139(2):213–252.

Marc Cavazza and Fred Charles. 2005. Dialogue Generation in Character-based Interactive Storytelling. *AAAI, AIIDE2005,*, pages 21–26.

Marc Cavazza, Fred Charles, and Steven J Mead. 2002. Character-Based Interactive Storytelling. *IEEE Intelligent Systems,*, 17(4):17–24.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, volume 2, pages 602–610. Association for Computational Linguistics.

Nathanael Chambers and Daniel Jurafsky. 2010. A database of narrative schemas. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA)*.

Yun-Gyung Cheong and R Michael Young. 2015. Suspenser: A story generation system for suspense. *IEEE Transactions on Computational Intelligence and AI in Games*, 7(1):39–52.

Eugenio Concepción, Gonzalo Méndez, Pablo Gervás, and Carlos León. 2016. A Challenge Proposal for Narrative Generation Using CNLs. In *9th International Natural Language Generation Conference, INLG*, pages 171–173, Edinburgh, Scotland, UK.

Pablo Delatorre, Barbara Arfe, Pablo Gervás, and Manuel Palomo-Duarte. 2016. A component-based architecture for suspense modelling. *Proceedings of AISB, 3rd International on Computational Creativity (CC2016)*, pages 32–39.

Richard Doust. 2015. *A domain-independent model of suspense in narrative*. Ph.D. thesis, The Open University.

R.J. Gerrig and A.B.I. Bernardo. 1994. Readers as problem-solvers in the experience of suspense. *Poetics*, 22(6):459–472.

Arthur C Graesser, Murray Singer, and Tom Trabasso. 1994. Constructing inferences during narrative text comprehension. *Psychological review*, 101(3):371.

H. Hoeken and M. van Vliet. 2000. Suspense, curiosity, and surprise: How discourse structure influences the affective and cognitive processing of a story. *Poetics*, 27(4):277–286.

Matt Jones, Bradley C Love, and W Todd Maddox. 2006. Recency effects as a window to generalization: separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2):316–332.

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

J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Michael Lebowitz. 1985. Story-telling as planning and learning. *Poetics*, 14(6):483–502.

J.R. Meehan. 1977. Tale-spin, an interactive program that writes stories. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence (IJCAI), Cambridge, MA, USA*, volume 1, pages 91–98.

Brian O'Neill and Mark Riedl. 2014. Dramatis: A computational model of suspense. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence, Québec City, Québec, Canada*, pages 944–950.

Rafael Pérez y Pérez and Mike Sharples. 2001. Mexica: A computer model of a cognitive account of creative writing. *Journal of Experimental and Theoretical Artificial Intelligence*, 13(2):119–139.

Mark O Riedl and Robert Michael Young. 2010. Narrative planning: balancing plot and character. *Journal of Artificial Intelligence Research*, 39(1):217–268.

Roger C Schank and Robert P Abelson. 1977. *Scripts, plans, goals and understanding: an inquiry into human knowledge structures*. Lawrence Erlbaum Associates Publishers, Hillsdale, NJ.

S.R. Turner. 1992. *MINSTREL: a computer model of creativity and storytelling*. University of California at Los Angeles, CA, USA.

Stephen G Ware and R Michael Young. 2014. Glaive: A state-space narrative planner supporting intentionality and conflict. In *Proceedings of the 10th International Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE2014*, pages 80–86, North Carolina State University, Raleigh, NC USA.

Dolf Zillmann, 1996. *The psychology of suspense in dramatic exposition*, pages 199–231. Vorderer, P, Wulff, HJ and Friedrichsen, M.

Rolf A Zwaan, Mark C Langston, and Arthur C Graesser. 1995. The construction of situation models in narrative

comprehension: An event-indexing model. *Psychological Science*, 6(5):292–297.