Instructions for *ACL Proceedings

First Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Second Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@domain

Abstract

Capturing readers' engagement in fiction is a challenging but important aspect of narrative understanding. In this study, we collected 23 readers' reactions to 2 short stories through eye tracking, sentence-level annotations, and an overall engagement scale survey. Our aim is to analyze the significance of various qualities of the text in predicting how engaging a reader is likely to find it. As enjoyment of fiction is highly contextual, we will also investigate individual differences in our data. Furthering our understanding of what captivates readers in fiction will help better inform models used in creative narrative generation and collaborative writing tools.

1 Introduction

The question of reader engagement in fiction has been studied in the psychology field for decades, with some of the foundational theoretical work from Iser (citation needed) and Gerrig paving the way for more recent theoretical frameworks and experimental setups, which started in the early 2000s with the work by Green (citation) in the field of Media Psychology. In the past ten years, this question has been picked up in more fields, such as cognitive science, psycholinguistics, and computational narrative understanding.

However, as Jacobs emphasized (2018), the samples normally collected are small and not enough to compensate for individual differences in reading patterns due to reader context, among other factors. In order to help close the experimental gap, one contribution of this study is to provide the computational community with a data set of reader reactions to literary short fiction, which Jacobs refers to as "hot" experimental research. This is opposed to setups in which the stories are contrived for the purpose of the study.

Green's 2006 study narrowed down the salient aspects of reader engagement to create narrative engagement scale, which we have modified slightly to gage overall interest in the story. In addition, in order to obtain more granular information, we used these aspects to design an annotation task that would provide sentence-level feedback.

2 Related Research

[See Table 1 for comparison matrix. Brief description of other experimental setups and conclusions]

3 Research questions

3.1 RQ1: Does absorption in a story lead to longer gaze durations?

Subquestion: do "Present" highlights (i.e. transported) correlate with faster reading and "Connected" and "Curious" highlights with slower reading as the Jacobs model hypothesized?

3.2 RQ2: (move to future work) Does gaze duration increase in foregrounding passages and decrease in backgrounding passages?

Subquestion: is absorption greater in foregrounding passages?

3.3 RQ3: How much is engagement dependent on reader context vs. linguistic (discourse) features?

What are the overlapping cases (sentences) when those two sets of features agree? What are the diverging cases? Can discourse features be used as a proxy for predicting eye-tracking features?

3.4 RQ4: Are eye-tracking patterns consistent across users?

If they are consistent, can we build a model using all of the users' data to predict the engagement label or their future time-series data? If not, can we build separate models for each user? Analyses on diverging cases between the users: why are they

	Ours	Kunz et al.	Mangen et al.	Hsu et al.	Mak et al.	Maslej et al.
Data gathered						
Eye tracking	x	x	x		x	
Saccade angle		x	x			
fMRI				x		
Engagement survey	x	x	x		x	x
Engagement annotation	x			x		
Textual features extracted						
Emotional arc	x					
Lexical categories	x			x		x
Description category			x			

Table 1: Comparison between our study and other similar experiments.

diverging? Perhaps based on background, personal experience, interests, etc.

3.5 RQ5: Is there a correlation between a sentence having low average word valence and reader's absorption?

4 Methods

4.1 Participant study

The study asked 31 English speakers (17 female, 11 male, 3 other, 23 native English speakers, average age: 26) were asked to read two short stories by Anton Chekhov while their eyes were tracked, and then answer an engagement scale survey. After reading through both stories, they completed a highlighting exercise where they highlighted areas according to the following categories:

- Present: Able to vividly picture the scene in the story
- Confused
- Curious: Curious about what will happen next
- Connected: Connected to the character; able to identify with them or feel their emotions
- Other: Enjoyed it for a different reason

Due to poor calibration, 8 samples were discarded, leaving 23 (13 female, 8 male, 2 other, 17 native English speakers, average age: 28).

The eye tracking results were drift corrected and interest area reports were exported using words as interest areas. Outliers for dwell time were removed using inner quartile range (1.7% of the data). The data was aggregated to the sentence level and dwell time values were averaged by character count. Due to some missing data, null values for the eye tracking features were filled with the average of the 5 nearest sentences (5.7% of all sentences read across participants).

4.2 Linguistic and discourse features

We extracted the following features from the stories to create sentence-level predictors: sentiment scores using the RoBERTa sentiment base model (footnote with link to huggingface), emotion categories using the DistilRoBERTa emotion base model, concreteness from the Brysbaert (brysbaert2014) corpus, valence and arousal from the NRC-VAD corpus, word frequency from the subtlex corpus, and average word length.

5 Limitations

There is an imbalance in how many participants read each story first (Expensive Lessons: 16, Schoolmistress: 7), which may affect level of attention for the second story. In addition, the stories did not receive very high scores on average in the engagement survey. On a scale from 0-4, Expensive Lessons got an average of 2.09 and Schoolmistress got 1.92.

6 Results

Other studies have shown that valence and arousal play an important role in predicting interest in a story (Maslej, Jacobs 2015, HSU201596) and Jacobs emphasized the importance of affective and emotional processes. We used linear mixed models to fit predictions of the proportion of the sentence highlighted and gaze duration, with random effects of participant (n=23) and story (n=2). For predicting the proportion of a sentence highlighted,

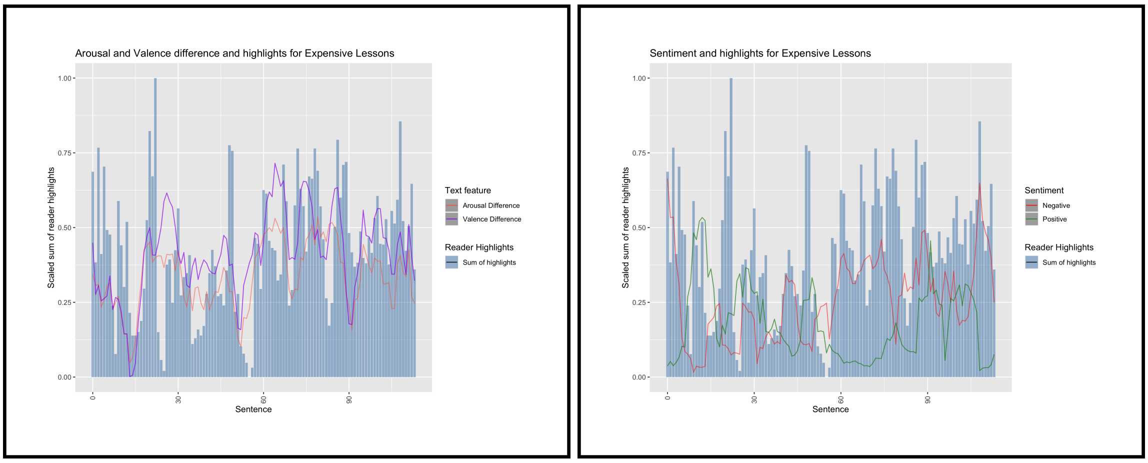

Figure 1: Highlights and features.

a proxy for a higher level of engagement, our results support a slight significance of valence mean (p=0.0049), similar to Hsu.

Unlike in other studies, we found that arousal mean had no significance (p=0.598). However, like Hsu, there was a higher significance for valence-span (p=0.00003) - the difference between valence max and valence min and arousal-span (p=0.00029) - the difference between arousal max and arousal min. This suggests that the reader was more engaged in sentences with a higher range of valence and arousal. Other features that had an effect were negative sentiment score (p=8.74e-07) and character count (p=3.97e-05). This model explains 22.9% of the variance in the proportion of a sentence highlighted with random effects and 3.6% without.

In predicting gaze duration, valence mean had a slight significance (p=0.0181) and arousal mean had none (p=0.365) and valence-span (p=0.0336) had a slight significance while arousal-span was found to be significant, but with a negative slope (p=0.00047). These findings do not support Jacobs's proposed framework, which states that passages that engage our emotions would likely result in slower reading. However, we found a positive relationship with negative sentiment (p=0.0048), which may partially support his proposal that the affective mode of reading is more readily triggered by negative emotion. Length of the sentence (p=0.000012) and word frequency (p=0.000048) were the most significant predictors of dwell time. This model explains 30.15% of the variance in the proportion of a sentence highlighted with random effects and 0.7% without.

7 Future Work

Investigate: Does gaze duration increase in foregrounding passages and decrease in backgrounding passages?

8 Conclusion

8.1 References

References

Exploring the shapes of stories using python and sentiment apis. https://indicodata.ai/blog/plotlines/. Accessed: 2022-12-21.

2015. *Cognitive Neuroscience of Natural Language Use*. Cambridge University Press.

Marc Brysbaert 1, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, 46(3):904–911.

Stefan Bastholm Andrade and Ditte Andersen. 2020. Digital story grammar: a quantitative methodology for narrative analysis. *International Journal of Social Research Methodology*, 23(4):405–421.

Arash Aryani, Markus Conrad, and Arthur Jacobs. 2013. Extracting salient sublexical units from written texts: "emophon," a corpus-based approach to phonological iconicity. *Frontiers in Psychology*, 4.

Ryan Boyd, Ashwini Ashokkumar, Sarah Seraj, and James Pennebaker. 2022. The development and psychometric properties of liwc-22.

Ryan L. Boyd, Kate G. Blackburn, and James W. Pennebaker. 2020. The narrative arc: Revealing core narrative structures through text analysis. *Science Advances*, 6(32).

Rick Busselle and Helena Bilandzic. 2009. Measuring narrative engagement. *Media Psychology*, 12(4):321–347.

	Estimate	Std. Error	df	t value	$Pr(>\|t\|)$	Sig.
(Intercept)	0.01910	0.07439	220.85084	0.257	0.797621	
character count	0.15837	0.03851	7070.78993	4.112	3.97×10^{-5}	$***$
word frequency avg.	0.06215	0.09899	7094.55339	0.628	0.530152	
positive	0.03362	0.02244	7120.97997	1.498	0.134154	
negative	0.10018	0.02035	7112.05570	4.923	8.74e-07	$***$
concreteness	0.01150	0.01361	6942.09621	0.845	0.398257	
valence avg.	0.12612	0.04488	7120.08454	2.810	0.004964	$**$
arousal avg.	-0.03312	0.05385	7118.47514	-0.615	0.538580	
valence diff.	0.10635	0.02634	7118.61978	4.037	5.47×10^{-5}	$***$
arousal diff.	0.11636	0.03165	7064.83819	3.677	0.000238	$***$

Table 2: Fixed Effects for predicting proportion of sentence highlighted

	Estimate	Std. Error	df	t value	$Pr(>\|t\|)$	Sig.
(Intercept)	0.1255	0.0233	91.28	5.396	5.34e-07	$***$
word frequency avg.	0.5548	0.0297	7117	18.656	$< 2 \times 10^{-16}$	$***$
positive	0.0251	0.0068	7117	3.707	0.000211	$**$
negative	0.0191	0.0061	7116	3.121	0.001813	$**$
concreteness	0.0146	0.0041	7086	3.568	0.000362	$**$
valence avg.	-0.0896	0.0135	7117	-6.655	3.03×10^{-11}	$***$
arousal avg.	0.0147	0.0161	7117	0.914	0.360620	
valence diff.	-0.0396	0.0075	7116	-5.306	1.16×10^{-7}	$***$
arousal diff.	-0.0801	0.0089	7117	-9.026	$< 2 \times 10^{-16}$	$***$
proportion	-0.0021	0.0036	7141	-0.598	0.550015	

Table 3: Fixed Effects for predicting gaze duration

Ashley Chung-fat yim, Elena Cilento, Ewelina Piotrowska, and Raymond A. Mar. 2019. Are stories just as transporting when not in your native tongue? *Language and Cognition*, 11(2):285–309.

Gianluca Consoli. 2018. Preliminary steps towards a cognitive theory of fiction and its effects. *Journal of Cultural Cognitive Science*, 2:85–100.

Pablo Delatorre, Alberto Salguero, Carlos León, and Alan Tapscott. 2019. The impact of context on affective norms: A case of study with suspense. *Frontiers in Psychology*, 10.

Mariana Ferreira and Hugo Gonçalo Oliveira. 2018. Seeking the ideal narrative model for computer-generated narratives. In *Proceedings of the 3rd Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2018)*, pages 5–10, Tilburg, the Netherlands. Association for Computational Linguistics.

Melanie C. Green, Timothy C. Brock, and Geoff F. Kaufman. 2006. Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, 14(4):311–327.

Chun-Ting Hsu, Arthur M. Jacobs, Francesca M.M. Citron, and Markus Conrad. 2015. The emotion potential of words and passages in reading harry potter – an fmri study. *Brain and Language*, 142:96–114.

Arthur M. Jacobs. 2017. Quantifying the beauty of words: A neurocognitive poetics perspective. *Frontiers in Human Neuroscience*, 11.

Arthur M. Jacobs and Roel M. Willems. 2018. The fictive brain: Neurocognitive correlates of engagement in literature. *Review of General Psychology*, 22(2):147–160.

Anna Kasunic and Geoff Kaufman. 2018. Learning to listen: Critically considering the role of ai in human storytelling and character creation. *Proceedings of the First Workshop on Storytelling*.

Kai Kunze, Susana Sanchez, Tilman Dingler, Olivier Augereau, Koichi Kise, Masahiko Inami, and Terada Tsutomu. 2015. The augmented narrative: Toward estimating reader engagement. In *Proceedings of the 6th Augmented Human International Conference*, AH '15, page 163–164, New York, NY, USA. Association for Computing Machinery.

Anežka Kuzmičová. 2014. Literary narrative and mental imagery: A view from embodied cognition. *Style*, 48(3):275–293.

Marloes Mak and Roel M. Willems. 2019. Mental simulation during literary reading: Individual differences revealed with eye-tracking. *Language, Cognition and Neuroscience*, 34(4):511–535.

Marta M Maslej, Raymond A. Mar, and Victor Kuperman. 2019. The textual features of fiction that appeal to readers: Emotion and abstractness. *Psychology of Aesthetics, Creativity, and the Arts*.

Geoff F. Kaufman Melanie C. Green, Timothy C. Brock. 2004. Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, 14(4):311–327.

Hannah Rashkin, Sameer Singh, and Yejin Choi. 2016. Connotation frames: A data-driven investigation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 311–321, Berlin, Germany. Association for Computational Linguistics.

P. Stockwell. 2002. *Cognitive Poetics: An Introduction*. Routledge.

Acknowledgements

We would like to thank the Text Group for initial feedback on experiment setup and the Blue Lantern Writing Group for additional feedback.