An Analysis of Reader Engagement in Literary Fiction through Eye Tracking

Rose Neis
University of Minnesota
neis@umn.edu

Karin de Langis
University of Minnesota
dento019@umn.edu

Zae Myung Kim
University of Minnesota
kim01756@umn.edu

Dongyeop Kang
University of Minnesota
dongyeop@umn.edu

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 Gerrig on Transportation Theory (Gerrig, 1993) paving the way for more recent theoretical frameworks and experimental setups, which started in the early 2000s with the work by Green et al. interested in the role of transportation on the persuasiveness of narratives (Melanie C. Green, 2004; Green et al., 2006). In the past ten years, this question has been picked up in more fields, such as cognitive science, psycholinguistics, and computational narrative understanding, so we have more theoretical and experimental work to draw from.

However, as Arthur Jacobs emphasized in his article, "Towards a neurocognitive poetics model of literary reading", the samples normally collected are small and not enough to compensate for individual differences in reading patterns due to reader context and other situational factors (Willems, 2015). 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 natural stories, which Jacobs refers to as "hot" experimental research.

In a 2009 study, a pair of researchers narrowed down the salient aspects of reader engagement, building off of Green's transportation framework to create narrative engagement scale (Busselle and Bilandzic, 2009), 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. We use linear mixed models to discover textual features that have an impact on engagement and dwell time across readers.

2 Related Research

There have been several eye tracking and fMRI studies in the area of reader interest (a few are shown in Table 1). We will compare our findings with a select few existing experiments as well as Jacobs' proposed framework. One 13-participant study showed that words in enactive passages had on average longer fixation durations and dwell times (Magyari et al., 2020). Based on survey responses, the authors hypothesize that in the enactive texts, the ease of imagery contributes to greater involvement in imagination and results in an overall slower reading speed. An fMRI study found valence and arousal scores as good predictors of overall emotional experience of the reader (Hsu et al., 2015).

3 Research questions

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

To answer this question, we will look at how well "Present" highlights (i.e. transported) correlate with faster reading and "Connected" and "Curious" highlights with slower reading to see if we find a relationship between dwell time and different modes of reading – one being immersed and the other more reflective. We will also look at whether

	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.

features related to a more affective reading mode lead to higher dwell times as Jacobs predicts.

3.2 RQ2: How much is engagement dependent on reader context vs. linguistic features?

We will attempt to answer this question by seeing how well the features we extracted can predict whether a sentence is highlighted by readers.

3.3 RQ3: Are dwell time patterns consistent across readers?

In order to answer this question, we will scale dwell times per participant and evaluate the pattern over the story.

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 calibration issues, 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 normalized by sentence character count. To handle 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 [1], emotion categories using the DistilRoBERTa emotion base model [2], concreteness scores from the Brysbaert corpus (Brysbaert et al., 2014), valence and arousal from the NRC-VAD corpus (Warriner et al., 2013), word frequency from the subtlex corpus (Brysbaert, 2015), and average word length.

5 Limitations

There are a few issues with the data that should be mentioned. Since the participants were asked to read two stories in a row, it is best to make sure there is a balance in which story is read first. However, our data ended up with a skew towards one story (Expensive Lessons: 16, Schoolmistress: 7), which may affect level of attention for the second

[1]https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

[2]https://huggingface.co/j-hartmann/emotion-english-distilroberta-base

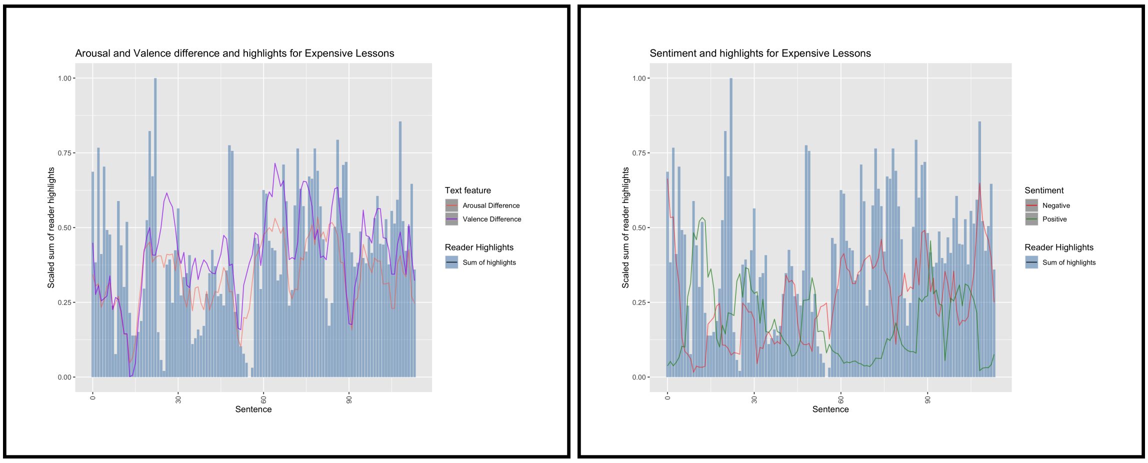

Figure 1: Highlights and features.

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 averaged 1.92. Ideally, stories used for such studies should be more widely popular in order to make absorption more likely. Perhaps in part due to the low average score, the highlighting data is sparse, making it difficult to find relationships between dwell time and engagement categories.

6 Results

Other studies have shown that valence and arousal play an important role in predicting interest in a story (Maslej et al., 2019; Hsu et al., 2015) and Jacobs emphasized the importance of affective and emotional processes in his framework (Willems, 2015). In order to determine the importance of these values for our data, 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, an indication of a higher level of engagement, our results support a slight significance of valence mean (p=0.01), similar to Hsu et al (Hsu et al., 2015). Unlike in other studies, we found that arousal mean had no significance (p=0.686). However, similar to Hsu et al., there was a higher significance for valence-span (p=0.00003) — the difference between valence max and valence min and arousal-span (p=0.00071) — 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.

We also experimented with including emotion categories, and surprise was found to be a significant effect (p=0.001). Other features that had an impact were negative sentiment score (p=5.84e-06) and character count (p=.000014). These findings partially align with Maslej et al. study (Maslej et al., 2019), where negative emotion predicted higher story ratings, although unlike their findings, there was no relationship between concreteness and engagement.

This model explains 3.7% of the variance without random effects and 23% with. So, with respect to RQ2, the reader context is important in elucidating the relationships of the fixed effects with engagement.

In predicting gaze duration, valence mean had a slight significance (p=0.00001) and arousal mean had none (p=0.349) and valence-span (p=0.0029) had some significance while arousal-span was found to be very significant (p=2.19e-12). Related to RQ1, the negative relationship between valence mean and dwell time supports part of Jacobs's proposed framework, which states that passages that engage our emotions, particularly negative valence, would likely result in slower reading. There was no relationship between highlights and dwell time, however, so we were not able to confirm whether the different categories of engagement correlated with different modes of reading.

There was also a positive relationship between concreteness and dwell time (p=0.0002). According to the prevaling theory in neuroscience, "words referring to easily perceptible entities coactivate the brain regions involved in the perception of those entities" (Brysbaert et al., 2014). This observation may indicate that this leads to longer

	Estimate	Std. Error	df	t value	$Pr(> \|t\|)$	Sig.
(Intercept)	-0.05352	0.07761	296.542	-0.690	0.49102	
character count	0.16092	0.03710	7023.920	4.337	1.46×10^{-5}	$***$
word frequency	0.07002	0.04102	6769.055	1.707	0.08786	.
positive	0.03786	0.02241	7118.596	1.690	0.09116	.
negative	0.09271	0.02044	7085.737	4.536	5.84×10^{-6}	$***$
concreteness	0.01938	0.01366	6380.945	1.419	0.15598	
valence mean	0.11482	0.04536	7118.133	2.531	0.01139	$*$
arousal mean	-0.02189	0.05424	7118.401	-0.404	0.68656	
valence span	0.10988	0.02647	7117.824	4.151	3.35×10^{-5}	$***$
arousal span	0.10874	0.03210	7057.269	3.388	0.00071	$***$
surprise	0.08417	0.02558	6987.261	3.290	0.00101	$**$
disgust	0.02947	0.01563	7115.526	1.885	0.05949	.

Table 2: Fixed Effects for predicting proportion of sentence highlighted

	Estimate	Std. Error	df	t value	$Pr(> \|t\|)$	Sig.
(Intercept)	0.1010	0.0245	84.24	4.129	8.53×10^{-5}	$***$
word frequency	0.1877	0.0124	7105	15.086	$< 2 \times 10^{-16}$	$***$
positive	0.01352	0.006773	7120	1.996	0.045920	$*$
negative	0.009938	0.006108	7120	1.627	0.103740	
concreteness	0.01515	0.004112	7077	3.683	0.000232	$***$
valence mean	-0.06013	0.01367	7120	-4.399	1.10×10^{-5}	$***$
arousal mean	-0.01520	0.01625	7120	-0.935	0.349762	
valence span	-0.02226	0.007489	7120	-2.972	0.002967	$**$
arousal span	-0.06345	0.009020	7120	-7.034	2.19×10^{-12}	$***$
surprise	-0.03633	0.007631	7113	-4.761	1.96×10^{-6}	$***$

Table 3: Fixed Effects for predicting gaze duration

processing times. This may also align with the findings of Maslej et al., where enactive passages had higher dwell times. Word frequency (p<2e-16) and arousal span were the most significant fixed effects for predicting dwell time.

To evaluate how consistent dwell time patterns were across readers (RQ3), we examined the dwell time graphs of participants to see if there was a similar pattern. We noticed an especially striking similarity in patterns amongst readers who were highly engaged (see Figure 2).

7 Conclusion

7.1 References

References

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

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.

Marc Brysbaert. 2015. Subtlex us word frequency database.

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

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

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

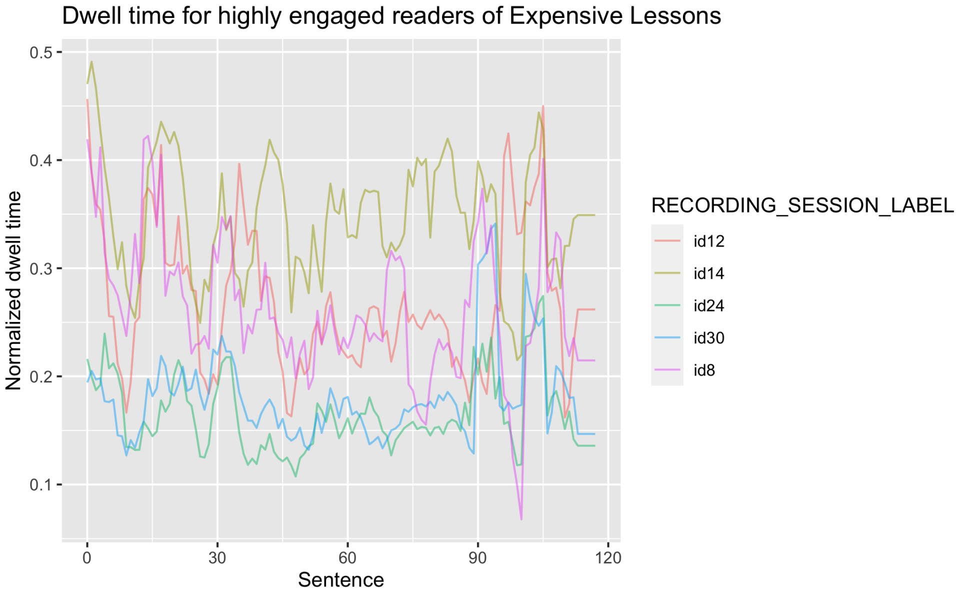

Figure 2: Dwell time for engaged readers

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.

Richard J. Gerrig. 1993. *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Yale University Press.

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.

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.

Lilla Magyari, Anne Mangen, Anežka Kuzmičová, Arthur M. Jacobs, and Jana Lüdtke. 2020. Eye movements and mental imagery during reading of literary texts with different narrative styles. *Journal of Eye Movement Research*, 13(3):10.16910/jemr.13.3.3.

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.

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

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods*, 45(4):1191–1207.

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

Acknowledgements

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