Understanding Narrative Transportation in Fantasy Fanfiction

Kelsey Neis
University of Minnesota
College of Science and Engineering
neis@umn.edu

Yu Fang
University of Minnesota
College of Science and Engineering
fangx174@umn.edu

Abstract

Our study aims to use understandings of narrative transportation as defined by social scientists to identify discourse features that could serve as predictors for how transporting a particular text will be to a reader.

1 Motivation

In their seminal work, Green and Brock defined transportation "as absorption into a story" (Green and Brock, 2000). An earlier definition by Gerrig (Gerrig, 1993) gave color to the phenomenon by using the analogy of travel:

> The traveler goes some distance from his or her world of origin, which makes some aspects of the world of origin inaccessible. The traveler returns to the world of origin, somewhat changed by the journey.

This state of mind can momentarily alter a reader's perspective, compel them to take on the beliefs of the protagonist, and make them forget about their current surroundings (Melanie C. Green, 2004).

It's a powerful phenomenon that can bring about positive transformation in an audience's outlook, or conversely have a negative effect on their behavior, as Green and Clark's findings in "Transportation into narrative worlds: implications for entertainment media influences on tobacco use" (Green and Clark, 2013) support a connection between transportation in film and its influence on tobacco use when tobacco is present:

> Transportation effects work through reducing counterarguing, creating connections (identification and liking) with characters and increasing perceptions of realism and emotional involvement.

Thus, the increased influence over an audience that transportation effects has potential for both good and harm.

With the explosion of narrative in all forms in recent years, we need a way to gain a larger picture understanding of the intentions and power of these narratives. We believe transportation theory provides a meaningful framework for understanding the effects of different narratives found in the wild. Our study is an effort towards enabling computational processing of narrative through the transportation theory lens. We attempt to do this by focusing on finding the discourse features of narrative text relevant to predicting transportation.

There are many factors in determining whether or not a particular text will be transporting to a particular audience. Some of them are thought to be dependent on an audience's context, while others are related to the aspects of narrative that are found in the text (Melanie C. Green, 2004). The main question we'll endeavor to answer is: what of those textual characteristics could be used to predict transportation?

2 Background

2.1 Narrative Transportation Theory

Narrative transportation theory comes from the field of psychology, stating that the audience tends to "get lost" in narrative stories, which is reflected in the changes of their emotions and attitudes in real life scenarios as a distinct mental process.(Green and Brock, 2000) Through engagement in narrative, readers may experience a loss of awareness of real-world facts, and researchers have discovered different belief states in readers of two stories.

Transportation into a story world has been conceptualized as a distinct mental process, with the possible mediums being audio, video, and text. In our studies, we mainly focus on the transportation of text-based stories, and we would like to propose our research on finding the correlation between the level of transportation and the selection of extracted features from the stories. Our study aims to produce an analysis pipeline for story elements, as described in one of the predominant role of AI in narratives (Kasunic and Kaufman, 2018).

2.1.1 Measurement of Transportation

Green and Brook proposed the 15-items self-report Transportation Scale(Green and Brock, 2000),(Green et al., 2006) for measuring the level of transportation. Examples as follows:

- When I was reading the narrative, I could easily picture the events in it taking place.
- I could picture myself in the scene of the events described in the narrative.
- I was mentally involved in the narrative while reading it.
- ...

However, this proposed method for measurement of the level of transportation is based on story-wise text; given the size of sample data we are able to collect for our study, we need to break down the metric into a sentence-wise analysis. Our measurement method is described in section 3.2.2.

2.1.2 Forms of Narratives

Kasunic and Kaufman proposed the two categories of narratives: 'listening' and 'agentic' in their work (Kasunic and Kaufman, 2018). 'Listening' form of narrative refers to a passive mode of context; for example, readings and listening to the narrative are both passive roles of receiving narratives. On the other hand, the 'agentic' role is a more interactive approach to receiving the narratives; for example, there are types of narratives that the reader can choose from several options to continue and decide the further story line. The forms of narrative remains a interesting topic to explore with respect to the level of transportation correlated to each types.

2.2 Narrative Arc

Whether the extraction of narrative structure is possible through text analysis methods is an interesting topic that has drawn attention recently. There are 3 key fragments of unfolding a story, which is the "arc of story", suggested by Freytag's framework are 1).Staging, 2). Plot progression, and 3).Cognitive tension.(Boyd et al., 2020)

Staging refers to the propositions and elements which help the narrator to set the scene and background information. Once we have the staging,

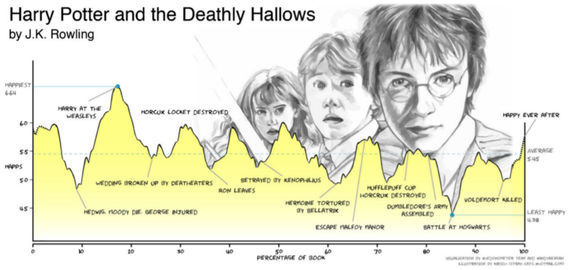


Figure 1: Sample emotion arc for Harry Potter. (Reagan et al., 2016)

plot progression takes over the narrative, which is the movements and interactions between the main characters. From the semantic perspective, the plot progression part will include an abundant amount of verbs, adverbs, and pronouns due to its functionality. As the narrative proceeds, it will reach a climax which is the **cognitive tension** which can appear as some central conflict for the characters to resolve. One approach of capturing these 3 fragments through textual analysis is by identifying function words in the corpus, which can be a building block to our project ideas.

2.3 Emotional Arc

The other tool we could utilize is the emotional arc, which is similar to the sentiment classification task of NLP on the sentence level. The idea is to identify the change of emotion impact through the narrative by sentiment analysis approaches on the breakdown fragments, and putting together we are able to visualize these classification and find a path of emotions, as the example shown in figure 1.

In general, there are 6 core categories for emotion arcs, according to (Reagan et al., 2016), listed below:

- 'Rags to riches' (rise).
- 'Tragedy', or 'Riches to rags' (fall).
- 'Man in a hole' (fall-rise).
- 'Icarus' (rise-fall).
- 'Cinderella' (rise-fall-rise).
- 'Oedipus' (fall-rise-fall).

In the future, we would like to apply NLP classification and data mining related methods and identify these two arcs: narrative arc and emotional arc as a supporting analysis for our topic studies, and we wish to find connections between these components and narrative transportation for the audience in a computational manner.

2.4 Related Work

In "The narrative arc: Revealing core narrative structures through text" analysis (Boyd et al., 2020), Boyd et al. used the presence of function words and cognitive processing words to predict the different stages of stories to generate narrative arcs. Our work does not focus on the narrative arc and is more focused on the reader's experience, but from this work we borrowed the idea that cognitive process words would predict for narrative transportation.

In the study, "Modelling Suspense in Short Stories as Uncertainty Reduction over Neural Representation" (Wilmot and Keller, 2020), Wilmot and Keller modeled suspense using two different mathematical formulas and collected annotations of suspense in stories by having readers simply mark each sentence as increasing, decreasing, or not changing in suspense. Although not specifically about transportation, we adopted this simple annotation setup, as we thought it was a relatively easy task that would not interfere with the readers' engagement with the stories.

Another study, "The Textual Features of Fiction That Appeal to Readers: Emotion and Abstractness" (Maslej et al., 2019) used reviews of short stories online to conduct statistical analysis on surface textual features to find those that had a strong correlation with appeal to readers.

3 Methods

In Green and Brock's influential work, they defined three high level aspects of transportation: ”imagery, affect, and attentional focus” (Green and Brock, 2000). Our first task was to find relevant textual features for these aspects to inform what features to extract from the story text.

We selected four stories from Archive of Our Own[1], a site where writers share their fanfiction, and had volunteers give feedback on how transporting the beginnings of the stories were at the sentence level.

3.1 Feature Extraction

3.1.1 Sentiment

Although neither positive nor negative sentiment alone predicts a reader's transportation (Maslej et al., 2019), we included sentiment analysis as a predictor of affect, because one of the indicators of transportation is whether the reader feels the same way as the characters in the story. For example, if the main character meets misfortune and the reader is transported, they will experience the sense of loss as they are reading.

We calculate the sentiment score by simply extracting the word count portion from the sentence, calculating as the difference of the number of positive sentiment words minus negative sentiment words over the total length of the sentence, The reason for us not to use the language model is that the model trained for the sentiment classification task only produces a soft-max probability of sentiment probability, while this information is not sufficient in our study. Our metric involves word sentiments and the length of the sentence, providing us with more insightful information that we need.

3.1.2 Part-Of-Speech tagging

The Part-Of-Speech-tagging features are derived from NLTK(Bird et al., 2009) package, with some slight modifications. We obtained the POS-tagging using the NLTK library trained on default corpus, with a total of 35 categories of taggings. We then generated both sentence-wise and story-wise tag distributions plot and observed the differences. An example for the analysis is shown in figure 5. In addition, we merged several categories together to reduce the level of chaos. For example, we combined VBD (past tense), VBG (gerund), VBN (past participle), VBZ (3-rd person), and VB(base) as verbs in our distribution analysis. A similar operation is applied to other categories that can be merged together without influencing the analysis of sentences.

3.1.3 Perception words

The Linguistic Inquiry and Word Count (LIWC) is a dictionary and text processing tool commonly used in the social sciences to analyze text. One category we found useful from it is perception words. They break down perception words into the following categories:

- Attention (e.g. look, watch, check)
- Motion (e.g. go, come, went, came)
- Space (e.g. in, out, up, there)
- Visual (e.g. see, look, eye*, saw)

[1]https://archiveofourown.org/

- Auditory (e.g. sound*, heard, hear, music)
- Feeling (e.g. feel, hard, cool, felt)

In narrative texts, these words can relate to the physical experiences of the characters or physical descriptions of the scene. So, they may contribute to the audience's level of presence in the story. Specifically, these words relate both to imagery and affect.

3.1.4 Cognition words

Drawing on the idea that "[t]ransportation draws upon, and perhaps helps develop, individuals' natural tendency toward empathy and perspective-taking" (Melanie C. Green, 2004) together with Boyd et. al's work on the Narrative Arc in describing how a "group of cognitive processing [...] words reflects the sense-making process that people engage in while working through a conflict or challenge in their life" (Boyd et al., 2020), we extracted cognitive processing words from our stories. In narrative, they indicate that a character is experiencing inner conflict or uncertainty. They may also indicate a closer focalization to a character, because more insight is being provided on their inner state. We selected this feature as a possible predictor of increased affect.

Using the Python BookNLP library[2], we obtained the per sentence cognition scores with supersense tagging, which tags text with 41 different lexical categories and scored each sentence based on the presence of verb.cogntition and noun.cognition words.

3.1.5 Concreteness

Taking the idea that concreteness in story text indicates the author's ability to "show not tell" and "using concrete images to convey deeper meanings" (Maslej et al., 2019), we obtained scores of concreteness per sentence using an annotated concreteness dictionary (1 et al., 2014).

3.2 Data Collection

3.2.1 Stories

In order to get a variety of positive and negative feedback from readers, we selected two stories from two different fandoms with high and low kudos, or votes. We took the first 500 or so words to create our survey.

Our choice of stories is based on two assumptions: 1).Since we are using fan-fiction narratives, we would like people to be familiar with the characters; in this way, we don't need to be limited by the introductory part of the stories, and we are able to increase the variety of data by choosing from different part of the narrative following the story arc. The reason we did not do so is that our annotators are volunteers, which was unexpected while designing this survey. 2).We assume there is a positive correlation between the number of kudos and the level of transportation for a narrative. With high kudos, people are more likely to be sympathetic to the narrative, thus being more transported compared with the low kudo ones.

Title	Fandom	Kudos
Oh God Not Again!	Harry Potter	25,145
Remnants	Harry Potter	2
Deku? I think he's some pro...	My Hero Academia	47,995
Aurum: Eyes of Gold	My Hero Academia	4

Table 1: Our data set of four stories selected from fantasy fanfiction.

3.2.2 Annotation

We collected 11 crowdsourced annotations using an anonymous Qualtrics survey. More detail about the survey is in Appendix A. To start, it included some questions to assess the survey takers' exposure to the fandoms, how much they read for pleasure, and their ability to conjure images in their mind.

Then, for each story they were instructed to select sentences and mark them if they increased or decreased their experience of being transported. After each story passage, they were given a subset of questions from Green and Brock's transportation scale. At the end, we asked survey takers to provide feedback on the survey and assess whether the task interfered with their ability to be transported.

Our metric for transportation level is defined in such a labeling go-along-the-way manner because we want to capture the labels as sentence-wise annotations. Also, instead of self-report the general impressions to a narrative/sentence by a multi-option labeling question, we utilize the method of highlighting to reduce the influence of readings to a minimum point.

[2]https://github.com/booknlp/booknlp

3.3 Classification

We format our goal as a set of classification tasks. We want to arrange our extracted features to be classified by a simple SVM classifier in a particular manner. The setting of the classification tasks is to separate the sentences with high and low levels of transportation. The binary labels used in our classification task are determined by collected annotation results weighted by the scale of vividness for a visual imaginary pre-survey check.

3.3.1 Classifier with POS-tagging

We used one-hot encoding to represent the POS tags that appeared in sentence-wise tagging analysis; then, we trained a polynomial kernel SVM to classify such distributions with respect to high/low transporting labels. We wish to know if the distribution of POS-tagging can influence a sentence's transportation level.

3.3.2 Classifier with Extracted Features

Additionally, using the normalized value of our proposed extracted features, we trained another classifier to check if we are able to find a hyperplane that separates the highs and lows.

4 Results & Discussions

We received 11 responses to the survey. We gave weights to the response highlights in proportion to the reader's self reported vividness of visual imagery score to obtain a combined score for each sentence.

4.1 Highlights & Correlation Analysis

Figure 2 shows the combined highlights per sentence, and figure 3 shows the positive highlights minus the negative highlights accumulated over the story. Although readers chose to highlight different sentences, we can observe some clusters of highlights. In future studies, we would like to make an in-depth analysis of these clusters and use story context to inform their interpretation. Figure 6 is an example of one of the positive highlight areas. Figure 3 gives a higher level picture of the readers' transportation. In future studies, we would like to analyze these patterns this over a whole story. More results can be found in our code[3].

The correlation heatmaps shown in Figure 4 highlight one of the challenges of our task. Although there are some correlations in the features we selected to an increase or decrease in transportation, each story appears to have a different correlation pattern. When combined together, the only correlations remaining are that for perception and binary sentiment, which is a 1 if the absolute value of the sentiment score is more than 0 and a 0 otherwise. Those correlations, shown in Table 2, are not strong.

Feature	Correlation with Increase	Correlation with Decrease
Perception	0.26	0.01
Binary Sentiment	0.19	0.03

Table 2: Combined correlations

4.2 POS-tagging & Classification Task Results

Through the plot for story-wise analysis, we found no significant difference between tag distributions. The result of our sentence-wise POS-tagging distribution analysis for the positive and negative rated sentences is shown in figure 5. The positive and negative sentences correspond to the high and low levels of transportation, respectively.

Positive sentences are constructed with less adverb(RB) words by 5%; and more determiner(DT) words by 6%, more noun(NN) words by 5%, along with more proper noun by 2%.

The results for our classification task are shown in table 3. As both the classification accuracy for positive and negative sentences is over 0.9, the POS-tag features are fairly effective. The classification using extracted features for sentence transporting level is not very promising, with an accuracy of around 0.7.

With only 11 survey responses on 4 passages of stories, these results are not in any way conclusive, but we do think that they reveal some areas of interest for this topic.

Task	SVM kernel	Accuracy
Positive sentences	polynomial	0.917
Negative sentences	linear	0.970
Extracted Features	polynomial	0.727

Table 3: Classification results

[3] https://github.com/kelseyneis/narrative_transportation

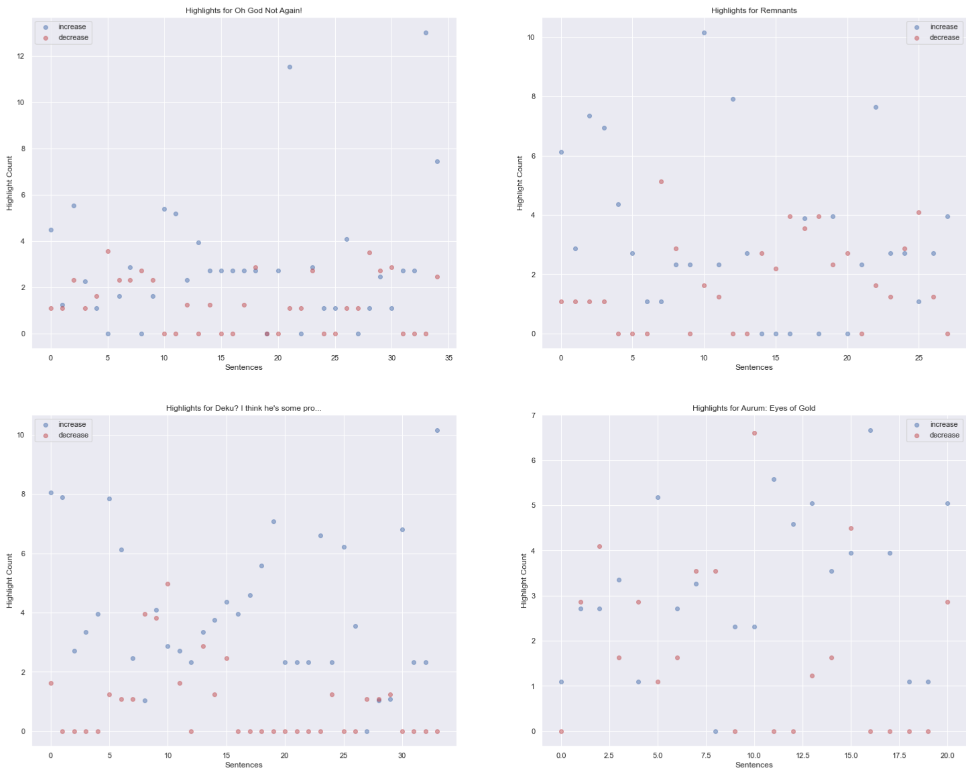


Figure 2: Combined highlights for each story

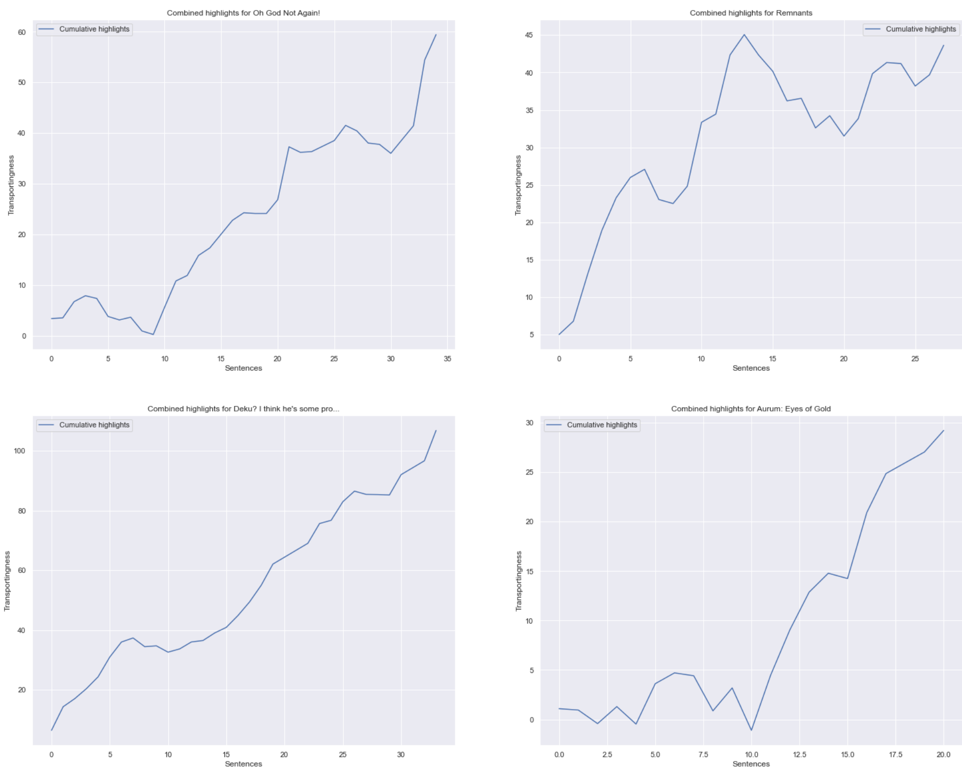


Figure 3: Cumulative highlights for each story

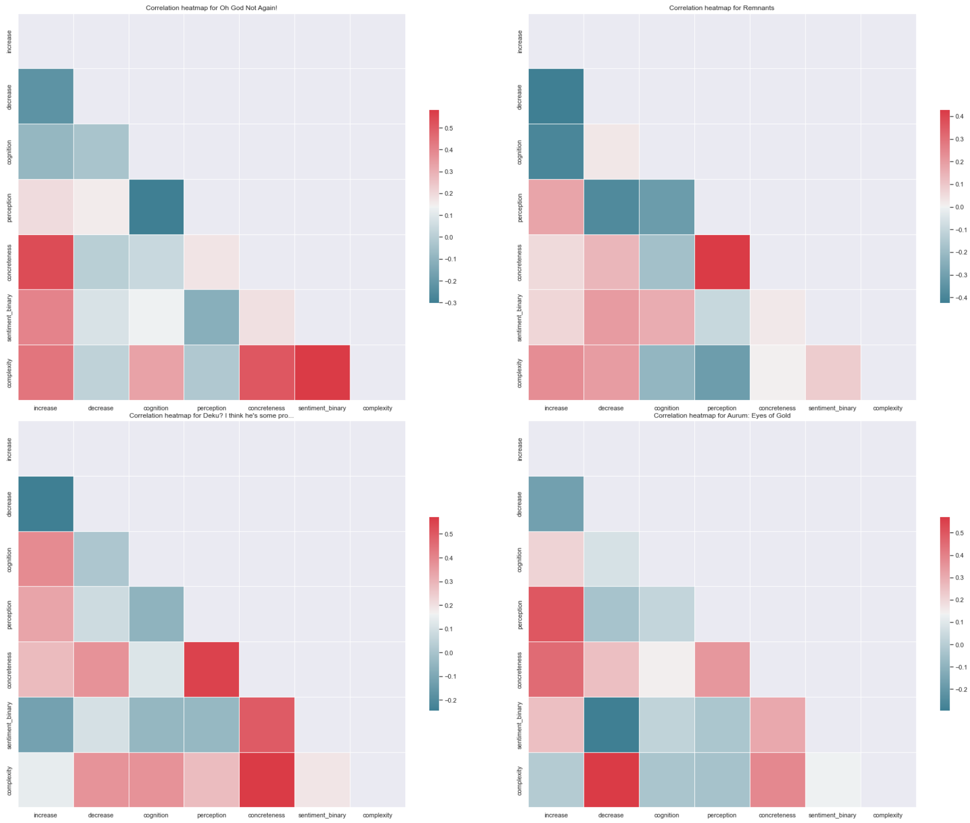


Figure 4: Correlation heat map for each story

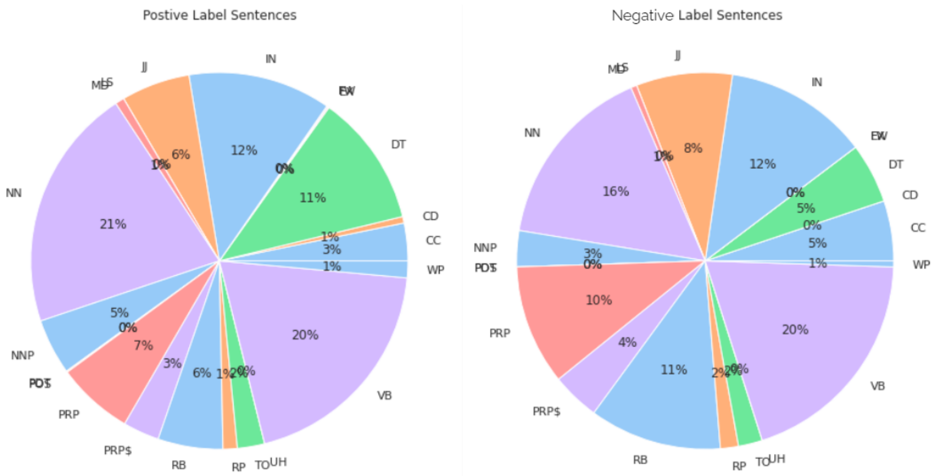


Figure 5: Sentence-wise POS-tag distributions (high & low)

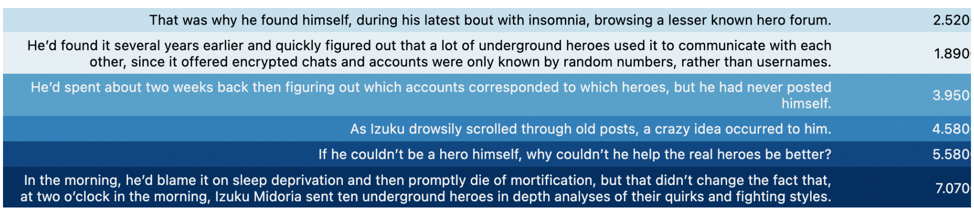


Figure 6: Heatmap of highlighted text for Deku? I think he's some pro...

5 Limitations

As mentioned earlier, the main limitation of this study is that of scale. Although there seem to be some observable patterns in the data, more responses are needed on longer texts.

It should also be acknowledged that this survey captures only a self-reported state of mind, and a couple respondents reported that the survey setup slightly interfered with their ability to get into the story. If we were to conduct this on a larger scale, we would want to improve the design to minimize the interference.

6 Ethics

Any attempt to automate the detection or generation of transporting narrative text poses a risk of misuse. The main claim that Green and Brock supported in their original study (Green and Brock, 2000) is a connection between how transporting a narrative is and how persuasive it is, and because narratives do not rely on any substantive proof for the beliefs they are supporting, it makes them easier to accept. Although this feature of narrative can be used to ease the acceptance of perfectly benign or even helpful knowledge (Dahlstrom, 2014), it should not be the sole source of a person's opinions and beliefs.

7 Conclusion

As a pilot study, we think our preliminary results show some promising patterns that should be tested on a larger scale in order to verify them. The patterns seem to be story-specific, so analysis should happen at the story level or in groups of similar stories rather than at the genre level. Our observations on the contribution of perception words, POS tag distribution, and binary sentiment have shown to be the most consistent across stories, but more data is needed to verify this.

8 Future Work

We see a few directions for scaling up the study and expanding the analysis of narrative features.

8.1 Data Collection

The highlighting task proved sufficiently easy that we think it is a good way of getting more granular feedback from readers on transportation. With more participants, we could present the readers with full short stories or carefully selected passages from longer works and collect feedback on a larger corpus of narrative text. Fan-fiction also provided participants with an engaging experience, so we would continue to draw from that genre.

In addition to our observation that people will like different features in different stories, there is also variance between readers. We used the vividness of visual imagery quiz to weight responses, but more analysis is needed to find the ideal weighting scheme or grouping of participants.

8.2 Feature Extraction

We would like to explore how the use of metaphor impacts transportation, so in a future study, finding a way to detect metaphor, or manually annotating it is something we'd like to pursue.

Since our efforts so far have been focused more on the word-level, we would also like to use embedding space from large scale language models, such as transform-based models like BERT, to examine contextual correlations and apply our existing analysis at different splits, such as clusters of positive and negative highlights.

8.3 Classification

We have demonstrated a baseline for classification tasks, so we'd like to apply this to a larger data set. More work could be done in fine-tuning models on fiction, and perhaps on the particular fandoms we want to use for our study.

Acknowledgements

Thanks to the Blue Lantern writing group for their feedback on the first version of the survey, and to those who participated in the pilot.

References

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.

Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

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).

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.

Michael F. Dahlstrom. 2014. Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences of the United States of America*, 111:13614–13620.

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

Melanie C. Green and Timothy C. Brock. 2000. The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5):701–721.

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.

Melanie C. Green and Jenna L. Clark. 2013. Transportation into narrative worlds: implications for entertainment media influences on tobacco use. *Addiction*, 108(3):477–484.

Melanie C. Green, Sheryl Kass, Jana Carrey, Benjamin Herzig, Ryan Feeney, and John Sabini. 2008. Transportation across media: Repeated exposure to print and film. *Media Psychology*, 11(4):512–539.

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*.

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

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.

Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 298–311, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1).

Tom van Laer, Ko de Ruyter, Luca M. Visconti, and Martin Wetzels. 2014. The extended transportation-imagery model: A meta-analysis of the antecedents and consequences of consumers' narrative transportation. *Journal of Consumer Research*, 40(5):797–817.

David Wilmot and Frank Keller. 2020. Modelling suspense in short stories as uncertainty reduction over neural representation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1763–1788, Online. Association for Computational Linguistics.

A Survey

A.1 Pre-survey Questions

1. What is your favorite genre of fiction?
2. On average, how much do you read for pleasure per week?
3. Have you read Harry Potter or seen the movies?
4. Have you read My Hero Academia (Boku no Hero Academia) manga series or seen the anime?
5. Close your eyes and picture a horse. Then go to the next page.

From vividness of visual imagery quiz

6. Which image above is closest to how well you were able to picture a horse? (scale: 1-6)

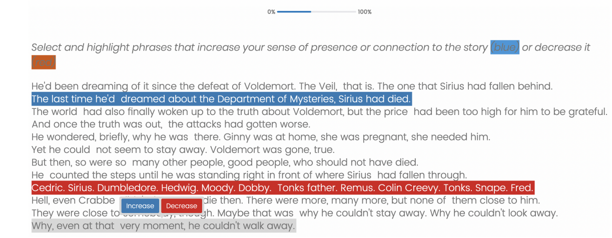
A.2 Highlighting instructions

In this survey, we want to get feedback on how transporting some stories are. You will read 2 short passages from 4 different stories and provide feedback on the text by highlighting it. The highlighted text will help us identify what parts of the story are more transporting and what parts are less so.

As you read each sentence, highlight it in blue if it makes you feel present in the story, able to vividly picture the scene, or connected with the character.

Highlight it in red if it makes you feel less present or less connected to the story.

You will then be asked a few questions about your experience of the passage.



A.3 Transportation scale questions

Please rate your experience of the passage (Options: Strongly disagree, Slightly disagree, Neutral, Slightly agree, Strongly agree)

1. The story affected me emotionally.
2. While reading my body was in the room, but my mind was inside the world created by the story.
3. I had a hard time keeping my mind on the story.
4. The characters were alive in my imagination.
5. I could picture myself in the scene of the events described in the story.

A.4 Survey feedback

1. Did the survey process make it difficult for you to get into the stories?
2. Did you find the instructions clear?
3. Additional feedback?

B Survey Results

Question	Rating
emotional affect	2.2
forgetting surroundings	2.5
distracted	1.8
characters alive	3.3
mental imagery	2.7

Table 4: Ratings for "Oh God Not Again!" Total: 8.9

Question	Rating
emotional affect	1.8
forgetting surroundings	2.6
distracted	2.0
characters alive	2.6
mental imagery	2.6

Table 5: Ratings for "Remnants" Total: 7.6

Question	Rating
emotional affect	2.3
forgetting surroundings	2.5
distracted	1.5
characters alive	2.4
mental imagery	2.6

Table 6: Ratings for "Remnants" Total: 7.6

Question	Rating
emotional affect	1.9
forgetting surroundings	2.2
distracted	2.3
characters alive	2.0
mental imagery	2.4

Table 7: Ratings for "Remnants" Total: 7.6

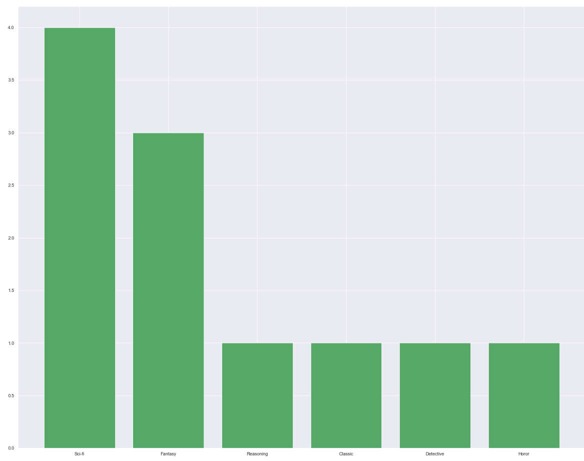


Figure 7: Genre preferences of participants