Computational Detection of Narrativity: A Comparison Using Textual Features and Reader Response

Max Steg and **Karlo Slot**[*] and **Federico Pianzola**
University of Groningen / Oude Kijk in Het Jatstraat 26, 9712 EK Groningen, Netherlands
`m.steg@student.rug.nl, k.h.r.slot@student.rug.nl, f.pianzola@rug.nl`

Abstract

The task of computational textual narrative detection focuses on detecting the presence of narrative parts, or the degree of narrativity in texts. In this work, we focus on detecting the local degree of narrativity in texts, using short text passages. We performed a human annotation experiment on 325 English texts ranging across 20 genres to capture readers' perception by means of three cognitive aspects: suspense, curiosity, and surprise. We then employed a linear regression model to predict narrativity scores for 17,372 texts. When comparing our average annotation scores to similar annotation experiments with different cognitive aspects, we found that Pearson's r ranges from .63 to .75. When looking at the calculated narrative probabilities, Pearson's r is .91. We found that it is possible to use suspense, curiosity and surprise to detect narrativity. However, there are still differences between methods. This does not imply that there are inherently correct methods, but rather suggests that the underlying definition of narrativity is a determining factor for the results of the computational models employed.

1 Introduction

Storytelling is and has been a big part of the daily lives of many. People learn from stories, get inspired by stories, relate to stories and *feel* stories. We as humans can be seen as story-interpreting machines. However, how we perceive and interpret storytelling elements has been a point of discussion in the field of narratology for a number of years. Previous research has attempted to define narrativity by means of cognitive processes (Genette and Levonas, 1976; Herman, 2009; Willis, 2021). A notable way to define narrativity is discussed extensively by Herman (2009), who uses four perceptive elements to define what makes a text narrative: situatedness, event sequencing, world-making and, as he describes it, "feltness".

Another important change in narratology concerns the shift from an idea of narrativity as a binary class, i.e. a text is a narrative or not, to narrativity as "a local, multidimensional scalar property" (Piper et al., 2021; Sternberg, 2001). Accordingly, Piper et al. (2021) have attempted to use the definition of narrativity as a scalar property to computationally detect narrativity, emulating human judgement using statements regarding the elements defined by Herman (2009).

However, there are still a lot of aspects of narrativity detection left unexplored. Alternative definitions heavily focus on readers' perception of texts (Sternberg, 2003, 2011; Passalacqua and Pianzola, 2016). For instance, Sternberg (2011) illustrates that narrativity can be defined by inherent human interpretation, which he refers to as the three "narrative universals": suspense, curiosity, and surprise. This research will attempt to explore the possibility of using this form of readers' perception as a way to detect narrativity, and will thus answer the questions:

To what extent can suspense, surprise and curiosity as a form of readers' perception be employed to detect narrativity?

Is there a relation between textual features associated to narrativity and reader response identified by narrative universals?

To answer these questions, we improve a text-based approach to detect the local narrativity of documents (Piper et al., 2021), provide new reader-response-based annotations for an existing corpus (Piper and Bagga, 2022), and discuss the results and implications of detecting narrativity using two different theoretical frameworks.

2 Related works

The goal to define the concept of narrativity in order to detect said concept within literary texts has

---

[*]The first two authors contributed equally.

been a main drive within the study of narratology. One of the fundamental definitions of narrative has been proposed by Genette and Levonas (1976), stating that the minimum requirements of a narrative should be that they represent a sequence of events by means of one or more characters.

Over time, researchers have elaborated on the aforementioned definition utilising two paradigms related to narrativity, as described by Herman (2009): "etic" and "emic". Etic approaches to narrativity regard the concept as definable and detectable by means of textual and structural elements. Whereas emic approaches utilise cognitive processes to classify texts by means of their degree of narrativity. A similar dichotomy has been described by Passalacqua and Pianzola (2016) by comparing objectivist and constructivist paradigms in the context of narrative theory. Passalacqua and Pianzola (2016) describe the objectivist paradigm as viewing textual aspects, such as semantics and syntactics, as defining features of a narrative. Constructivist theory, however, regards the relation between the audience and the text as a way to detect narrativity, which Passalacqua and Pianzola (2016) described as being "the result of a process of construction and combination of certain processes and properties whose specificity is also dependent on extra-objectual factors". One approach grounded within constructivist narrative theory, that can thus be considered an emic approach, is the concept of "readers' perception".

Readers' perception, or often called "reader response" and "reception" in literary studies, depicts the way a reader interprets textual elements on an emotional or cognitive level (Willis, 2021). Several researchers have attempted to define narrativity by means of readers' perception. These definitions have been employed by other research to detect narrativity in a handful of manners, including computational methods.

Herman (2009) suggests that the degree of narrativity in stories can be defined by the means of four perceptive elements:

1. *Situatedness*. In what context the narrative is presented.

2. *Event sequencing*. How events within a narrative are ordered, i.e. in a temporal fashion.

3. *World-making*. How the narrative presents a fictional or realistic world.

4. *"Feltness"*. How the reader is affected by the experiences presented in texts.

Metilli et al. (2019) have used the perceptive elements by Herman (2009)—mainly event sequencing—to propose a framework with technological challenges and requirements regarding the extraction of narratives from text. This framework consists of a combination of techniques to detect events using textual elements, such as temporal and named entity recognition, human annotation, and deep learning. Metilli et al. (2019) defined events as being "set in space and time, endowed with factual components" and having "semantic relations" with each other. Based on this definition, human annotators were assigned to identify sentences as being events or not. Simultaneously, a similar framework was proposed by Rodrigues et al. (2019), who provided guidelines to be used when aiming to visualise narratives by means of spatio-temporal relations. This vision described the concept of time and space in storytelling as something intuitive and inherently human. Both approaches by Metilli et al. (2019) and Rodrigues et al. (2019) indicate that human interpretation, or readers' perception, as described by Herman (2009) can be utilised to detect and visualise narrativity within texts by means of annotation and construction of classification models.

This idea has been successfully implemented by Piper et al. (2021) as well. While utilising etic, objectivist approaches to detect narrativity across long time scales, such as extracting narrativity by means of lexical and syntactic features, Piper et al. (2021) also utilized emic, constructivist approach similar to Metilli et al. (2019) and Rodrigues et al. (2019) using the perceptive elements by Herman (2009). Piper et al. (2021) assembled a team of three trained annotators to annotate 5-sentence-long passages spanning across four "discursive domains": (1) non-fiction, (2) fiction, (3) poetry and (4) science. The annotators were assigned to rate the passages on a 5-point Likert scale based on three elements: (1) "feltness", reworded as "agency", (2) event sequencing and (3) world-making. These elements have been explained in a codebook including annotation guidelines. While their approach to quantify readers' perception by means of Herman's elements and human annotation as a way to detect narrativity within texts is valid, their data lacks linguistic and stylistic differences. Piper and Bagga (2022) uses an expanded dataset

spanning across 19 genres, ranging from legal documents to Aesop's fables, and from 19th century literature to Reddit stories. This extensive dataset is extremely helpful for narrativity detection, due to its divergent nature, and can also be used to detect narrativity by means of different forms of readers' perception.

One other theory, also mentioned by Piper et al. (2021), which acknowledges the human interpretative aspects of narratives, has evolved over several years (Sternberg, 2011). Sternberg (2011) views narrativity as having the possibility to be defined by the effect that it can have on the reader, a view grounded within readers' perception. Sternberg states that the degree of narrativity of literary texts can be defined through three "master effects", or "universals": (1) suspense, (2) curiosity, and (3) surprise:

- *Suspense*. An event can be experienced as having suspense when the reader is presented with information that can eventually guide the reader to a sense of closure, or fulfilment (Sternberg, 2003). An example of this can be that the reader of a story is given the information that character A has a knife behind their back, but character B, with whom character A is having a conversation, is not aware of it; the reader lacks information about what will happen in the future.

- *Curiosity*. Curiosity can occur when the reader is presented with information about the present, also eliciting a desire for information about the past (Sternberg, 2003). For example, the reader is told that character B is found with several stab wounds, but is not told how they have gotten said wounds.

- *Surprise*. An event can be surprising if the readers' idea of the event being described is challenged through new information (Sternberg, 2003). For example, the reader knows that character A and character B are having a "normal" conversation, but suddenly, the reader is presented with the information that character A has stabbed character B. This could result in an "error" in the mental organization of the information previously acquired by the reader via the story, sparking a feeling of surprise.

Former research has used Sternberg's constructive approach to narrativity to construct computa-

tional models using some of the three universals. Doust and Piwek (2017) developed a computational model to detect suspense and found that, when compared with human judgements, their model predicted suspense rather well. However, their approach only used human judgements as a way to compare results, rather than attempting to use human judgements to train the model. A similar approach has also been used by Wilmot and Keller (2020).

There is a lack of research where all three universals are utilised to detect narrativity. While researchers have attempted to model one of the three universals by means of textual elements, they used human judgements as a way of evaluation, rather than as an approach to model suspense. Detection of narrativity by means of human judgements of suspense, curiosity and surprise can provide an insight into the relation between a story and its reader. Therefore, the question which will be tested is whether it is possible to detect narrativity utilising all three of Sternberg's universals. The approach taken will utilise human annotation to train a machine learning model and calculate narrativity scores based on readers' perception.

3 Data

The data used within this research was a corpus constructed and provided by Piper and Bagga (2022), consisting of 17,706 documents, ranging across 20 unique genres (Table 1). Examples of the genres within this corpus are historical non-fiction, fairy tales, documents from the Supreme Court of the United States, scientific abstract, and flash fiction. These documents contain short textual fragments of roughly 5 sentences, hereafter referred to as "passages". Out of the 17,706 passages, 334 have been manually annotated and used to build a model for detecting and predicting narrativity (Piper and Bagga, 2022). We reused the same annotated dataset for our own annotations and to build a new predictive model.

4 Methods

4.1 Annotation

The original dataset contains annotations referring to textual features identified on the basis of the following 3 statements (Piper et al., 2021):

- Agency: "This passage foregrounds the lived experience of particular agents."

- Event sequencing: "This passage is organized around sequences of events that occur over time."

- World-making: "This passage creates a world that I can see and feel."

Additionally, we annotated the data based on 3 statements regarding readers' perception (Sternberg, 2003):

- Suspense: "This passage presents information indicative of future events and postpones a feeling of resolution."

- Curiosity: "This passage presents information indicative of past events and leaves me wondering about missing information."

- Surprise: "This passage presents information, which I experience as unexpected, about an event."

The annotators expressed their agreement with the statements by means of a five-point Likert scale (Strongly disagree, Somewhat disagree, Unsure, Somewhat agree, Strongly agree). The choice to use a scalar rather than a categorical approach for annotation is in accordance with the theoretical framework adopted, namely that these cognitive aspects can be experienced as a spectrum. A passage can not only be defined as being suspenseful or not, some passages can be more or less suspenseful than others. Similarly, Piper et al. (2021)'s objectivist features can be experienced with different degrees of intensity in a text.

To test whether the selection of the dataset by Piper and Bagga (2022) was suitable for the annotation of the narrative universals defined by Sternberg (2003), one annotator annotated the 20 passages with the lowest and highest narrative probability, according to Piper and Bagga (2022), thus 40 passages in total. Since we are looking at *degrees* of narrativity, we used Kendall rank correlation coefficient (τ) to compare the ranking of the average annotated narrativity with the ranking of Piper's narrative probability scores. Kendall's τ for the annotations by Piper and Bagga and Piper's narrative probability scores was 0.385 (p < .001), while Kendall's τ for our initial annotations and Piper's narrative probability scores was 0.377 (p < .001). These values are close and indicate a strong rank correlation (> 0.3), hence this data set is suitable for annotation using Sternberg's universals.

Table 1: Finalised distribution of genres in the corpus and annotated dataset by Piper and Bagga (2022). Percentages within brackets refer to the ratio with respect to the total number of texts in the corpus and in the annotated dataset respectively.

Genre	Corpus	Annotated
ABSTRACT	993 (5.6%)	26 (7.8%)
APHORISM	486 (2.7%)	19 (5.7%)
BIO	990 (5.6%)	17 (5.1%)
BREVIEW	864 (4.9%)	-
FABLE	273 (1.5%)	10 (3%)
FAIRY	784 (4.4%)	20 (6%)
FLASH	889 (5%)	12 (3.6%)
HIST	1075 (6%)	25 (7.5%)
LEGAL	1115 (6.3%)	31 (9.3%)
LITSTUDY	544 (3.1%)	16 (4.8%)
MIXED NONFIC	1000 (5.6%)	-
NOVEL-CONT	974 (5.5%)	23 (6.9%)
NOVEL19C	1050 (5.9%)	25 (7.5%)
OPINION	1611 (9.1%)	-
PHIL	558 (3.1%)	20 (6%)
POETRY	1000 (5.6%)	-
REDDIT	1000 (5.6%)	20 (6%)
ROC	1000 (5.6%)	23 (6.9%)
SCOTUS	1000 (5.6%)	35 (10.5%)
SHORT	500 (2.8%)	12 (3.6%)
Total	17,706	334

Out of the initial experiment, 9 passages were extracted to exemplify each universal: 3 passages per universal, having low, medium, and high degree of each universal. We prepared an annotation guidebook with instructions and the commented examples, also including a brief background of the research goal, the theoretical framework, and an explanation of common annotation pitfalls, so that these can be avoided.[1]

We did a first round of annotation to check whether the constructed guidebook was clear and instructive enough. A total of 7 annotators (6 Dutch Information Science students and one Italian professor of computational humanities) annotated approximately 20 passages each, randomly assigned from the data set. Each passage was annotated by 3 annotators. Thus, this round yielded 47 annotated passages. We calculated Inter-rater reliability (IRR) using the average deviation index (ADI), as discussed by Burke et al. (1999). Since we used a 5-point scale, the ADI should not exceed the threshold value of 0.5. The final ADI score of the first round was 0.35, thus the guidelines did not need further improvements. Based on feedback from the annotators, we only added that the annotation should take into account the entirety of the passage, rather than just a part of it. However, we now acknowledge that this specification may be misleading in some cases, namely when the need for information related to curiosity or suspense is triggered and fulfilled in the same passage (see examples in Appendix).

In the second and final round 6 annotators annotated the remaining 278 passages, with approximately 138 passages per annotator. The ADI score of the second round was 0.37 and the ADI of both rounds combined was 0.36. Since both values are below 0.5, it can be concluded that there is a reasonable level of agreement between annotators.

Once we had annotated all passages, we compared them to Piper et al. (2021)'s annotations. We used Pearson's r and Kendall's τ to calculate the correlation between the results.

4.2 Models

Despite the theoretical framework adopted for the annotation, conceiving narrativity as a scalar property, Piper et al. (2021) eventually worked with computational models whose main goal is to classify texts into discrete categories (*Logistic Regression*, *Random Forest* and *Support Vector Machine*). To train a machine learning classifier, they used values ranging from 1 to 5 (average annotation on the Likert scale) but they also created an additional variable called "reader predicted label": if the average annotation value was higher than 2.5, it got the POS label, else, it got the NEG label. The resulting predicted narrativity for the whole corpus is thus the probability of either being a narrative or not. They also tested the performance of their classifiers using this 2-classes predictions.

Alternatively, we decided to implement a modelling approach consistent with our theoretical framework and predict the *degree* of narrativity of a text using linear regression models. Hence, we used both Piper and Bagga (2022)'s and our annotation to train several models (*Linear Regression*, *Lasso*, *Ridge*, *ElasticNet*, and *Theil-Sen*), predict narrativity scores ranging from 0 to 1, and test the models' performance on these continuous values. We did not try any neural approach because we wanted to be able to identify in a straightforward way the predictive power of various features.

For the selection of the textual features to supply to the model when training, we relied on Piper et al. (2021) but also tried a few other features that we thought could perform well. The features we used from Piper et al. (2021) are *unigrams*, *tense*, *mood*, *voice*. The latter three are composite features computed with the Python package *BookNLP*[2]. We also adapted the *concreteness* score (Brysbaert et al., 2014) by extending it with the lexicon developed by Muraki et al. (2022), which consists of 62 thousand English multiword expressions. The concreteness score of a document is the sum of all concreteness scores for all expressions in a document, divided by the total number of words. To explore the relation between semantics and narrativity, other features that we used are *Tf-idf* and Doc2Vec (Le and Mikolov, 2014). We tried different combinations of the selected features to train and test our models, focusing on predictions that correlate more strongly with the annotator scores. Due to limited size of our annotated data set, we used 5-fold cross-validation (train/test: 80/20). After having determined the best model, we trained it again twice using all the annotated data for each method (text-based and reader-based), and predicted narrativity scores for the complete corpus.

---

[1]https://github.com/maxsteg/Computationally-Narrativity-Detection

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

5 Results and Discussion

The best predictive model for both types of narrativity (text-based and reader-based) is Theil-Sen Regressor (TSR) with two features: Tf-idf and concreteness (Table 2). However, given that the model using only Tf-idf explains almost the same amount of variance (.01 difference), for the sake of interpretability we decided to use this simpler model for the prediction of narrative probability. Interestingly, the words contributing the most to predicting narrativity are not all the same for the two theoretical frameworks. When detecting narrativity based on textual features, third person pronouns seem more relevant, but first person pronouns are better predictors of narrativity based on readers' perception (Table 3).

Table 2: Coefficient of determination (R^2) of the annotators' scores and various features when predicting the narrativity of texts using the Theil-Sen model.

Features	R^2	
	Piper	*Univ.*
unigrams	.50	.33
tfidf	**.68**	**.59**
doc2vec	-1.53	-1.99
concreteness	.34	.35
doc2vec concr	-1.48	-1.78
tfidf concr	**.69**	**.60**
doc2vec ttr concr	-1.52	-1.71
tense mood voice	.65	.50
tfidf doc2vec	.02	.07
tfidf doc2vec concr	.11	.13
tfidf doc2vec unigrams	.51	.38
tfidf unigrams concr	.51	.38
unigrams doc2vec concr	.51	.38
unigrams doc2vec tfidf concr	.51	.38

Before evaluating the predicted values, we looked at the annotations done using two different theoretical frameworks to define narrativity. In Table 4, it can be seen that there is a positive correlation for all statement pairs between Piper's framework and Sternberg's universals. However, two of these are only moderate: between "Event sequencing" and "Curiosity" ($r = .63$), and between "Event sequencing" and "Surprise" ($r = .66$). These results show that the textual and cognitive dimensions covered by the two theoretical frameworks do not completely overlap.

These findings are also supported by the total annotation averages: there is a strong positive cor-

Table 3: Top words positively and negatively associated with narrativity. Computed with the Python package ELI5. See Appendix for a longer list

Positive		Negative	
Piper	*Universals*	*Piper*	*Universals*
he	out	is	of
my	me	of	by
was	was	which	is
him	door	or	or
had	different	2d	for
derrick	woman	this	can
his	my	agreement	as
day	plane	even	may

Table 4: Correlations (Pearson's *r*) between the average annotators' scores for each statement based on Piper's framework and Sternberg's universals.

	Suspense	**Curiosity**	**Surprise**
Agency	.75	.72	.70
Event	.70	.63	.66
World	.76	.72	.73

relation between the total annotation averages ($r = .77$), but there is some unexplained variance. Moreover, Kendall's τ shows a very weak correlation between the actual ranking of documents ($\tau = .03$). This indicates that, even though there is a strong correlation between the predicted values, the way in which the documents are ranked by means of the predicted narrativity scores differ strongly between models. This does not imply that one of these theories is inherently correct, but that they are different ways of viewing narrativity. Notably, the distribution of annotations shows that text-based annotation led to a majority of passages with the highest narrativity score, whereas reader-based annotation led to the opposite result: the majority of passages has been assigned the lowest narrativity score (Figure 1).

If we look at values predicted for the whole corpus, we see that they have an even stronger correlation ($r = .91$) and span the whole narrativity spectrum in a more even way (although still skewed) than the annotated passages, confirming that high or low narrativity texts are not more frequent than those with a moderate degree of narrativity (Figure 2). Conversely, Piper and Bagga (2022)'s use of a binary classifier (Logistic Regression) pushed the predictions towards extreme values, biasing the interpretation.
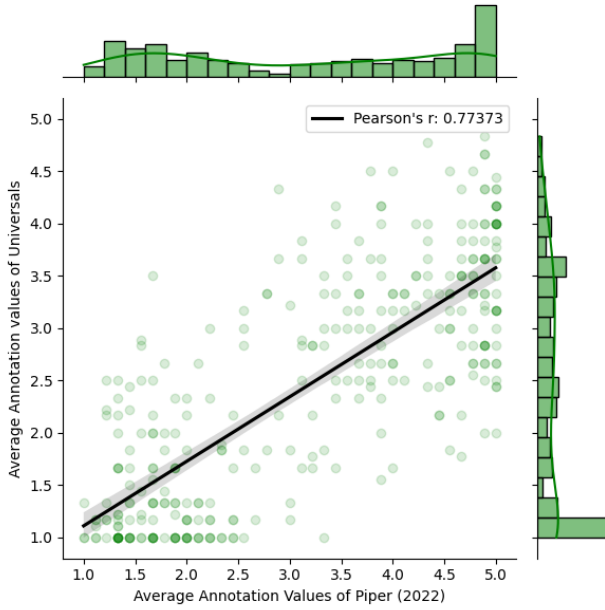
Figure 1: Scatterplot of Annotation Piper (2022) vs. Annotation Universals (*n* = 325)

We also looked at the average predicted probabilities for each genre (Table 5) and they are in line with what could be expected. For example, legal documents (LEGAL, SCOTUS) have low narrativity, between .06 and .3. This applies to academic texts (LITSTUDY, ABSTRACT), philosophy (PHIL), and book reviews (BREVIEWS) as well. All other genres have relatively high narrativity scores, with Reddit posts having the highest degree. Regardless of the theoretical framework employed, the variation in the degree of narrativity between genres is similar. However, the values are remarkably lower when narrativity is computed based on the narrative universals (mean = .39 vs. mean = .57). This result may be due to the different distribution of the annotated passages and the consequent unbalanced training sets, biased in different ways: towards high narrativity scores for text-based annotation and towards low narrativity scores for reader-based annotation. Follow-up research should aim for a larger and more balanced annotated dataset.

6   Conclusion

As narratology moved towards the idea that narrativity is a local, scalar, and multidimensional characteristic of texts, this research aimed to answer the questions *"To what extent can suspense, surprise and curiosity as a form of readers' perception be employed to detect narrativity?"* and *"Is there a*
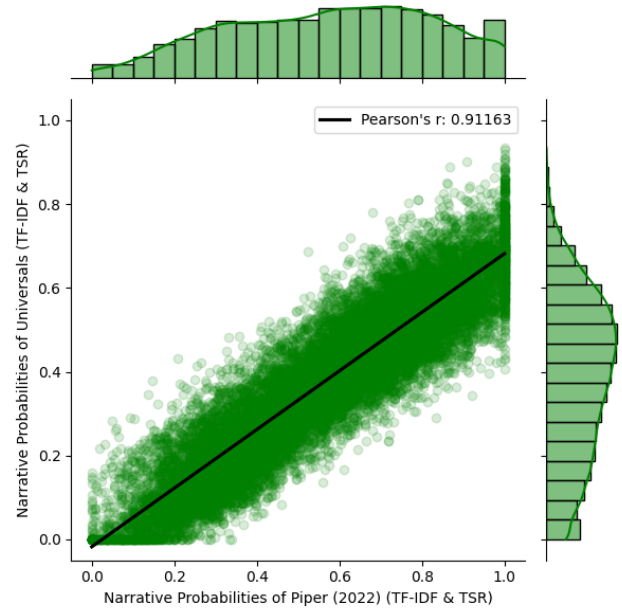

Figure 2: Scatterplot of narrativity scores predicted using a Theil-Sen regression model based on Piper and Bagga (2022) annotation vs. based on annotation related to Sternberg (2003) Universals (*n* = 17,372)

Table 5: Average predicted degree of narrativity per genre (TF-IDF & TSR model)

Genre	Avg. narrativity	
	Piper	*Universals*
ABSTRACT	.31	.13
APHORISM	.30	.27
BIO	.68	.43
BREVIEW	.45	.31
FABLE	.77	.51
FAIRY	.79	.54
FLASH	.75	.53
HIST	.60	.37
LEGAL	.17	.06
LITSTUDY	.37	.26
MIXED-NONFIC	.60	.40
NOVEL-CONT	.77	.56
NOVEL19C	.71	.50
OPINION	.53	.34
PHIL	.32	.24
POETRY	.61	.45
REDDIT	.82	.61
ROC	.81	.50
SCOTUS	.30	.15
SHORT	.78	.56

relation between textual features associated to narrativity and reader response identified by narrative universals?".

To accomplish this, we performed multiple rounds of annotation to quantify readers' perception by means of three cognitive effects of narrative (suspense, curiosity, surprise). We found that the annotators generally agree with each other and there are moderate to strong correlations between text-based and reader-based narrative dimensions.

As for the computational aspect, we trained a quite accurate regression model on a single highly predictive feature (*Tf-idf*). Comparing our results to the results by (Piper and Bagga, 2022), we found that there is a strong positive correlation between their and our approach to defining narrativity as a scalar property of texts. However, we also found that text-based and reader-based conceptions of narrativity do not completely overlap.

While this research has been able to capture narrativity by means of readers' perception as defined by Sternberg (2011), it is not able to capture narrativity as a whole. As mentioned before, narrativity can be defined in various ways and thus there are many plausible ways to detect it. Future research could combine both text-based and reader-based approaches to better grasp the complex, multidimensional nature of narrative. For instance, principal component analysis could help identify dimensions that could be conflated and dimensions that are irreducible to objective and textual properties.

Acknowledgments

References

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.

Michael J Burke, Lisa M Finkelstein, and Michelle S Dusig. 1999. On average deviation indices for estimating interrater agreement. *Organizational Research Methods*, 2(1):49–68.

Richard Doust and Paul Piwek. 2017. A model of suspense for narrative generation. Association for Computational Linguistics.

Gérard Genette and Ann Levonas. 1976. Boundaries of narrative. *New Literary History*, 8(1):1–13.

David Herman. 2009. *Basic elements of narrative*. John Wiley & Sons.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.

Daniele Metilli, Valentina Bartalesi, and Carlo Meghini. 2019. Steps towards a system to extract formal narratives from text. In *Text2Story@ ECIR*, pages 53–61.

Emiko J Muraki, Summer Abdalla, Marc Brysbaert, and Penny M Pexman. 2022. Concreteness ratings for 62 thousand english multiword expressions.

Franco Passalacqua and Federico Pianzola. 2016. Epistemological problems in narrative theory: Objectivist vs. constructivist paradigm. *Narrative Sequence in Contemporary Narratology*, pages 195–217.

Andrew Piper and Sunyam Bagga. 2022. Towards a data-driven theory of narrativity. *New Literary History*, (forthcoming).

Andrew Piper, Sunyam Bagga, Laura Monteiro, Andrew Yang, Marie Labrosse, and Yu Lu Liu. 2021. Detecting narrativity across long time scales. *Proceedings http://ceur-ws. org ISSN*, 1613:0073.

Sara Rodrigues, Ana Figueiras, and Ilo Alexandre. 2019. Once upon a time in a land far away: guidelines for spatio-temporal narrative visualization. In *2019 23rd International Conference Information Visualisation (IV)*, pages 44–49. IEEE.

Meir Sternberg. 2001. How narrativity makes a difference. *Narrative*, 9(2):115–122.

Meir Sternberg. 2003. Universals of narrative and their cognitivist fortunes (ii). *Poetics today*, 24(3):517–638.

Meir Sternberg. 2011. Reconceptualizing narratology. arguments for a functionalist and constructivist approach to narrative. *Enthymema*, (4):35–50.

Ika Willis. 2021. Reception theory, reception history, reception studies. In *Oxford Research Encyclopedia of Literature*.

David Wilmot and Frank Keller. 2020. Modelling suspense in short stories as uncertainty reduction over neural representation. *arXiv preprint arXiv:2004.14905*.

Appendix

In this Appendix there is a list of the words contributing the most to predicting low narrativity (red) and high narrativity (green), and examples of passages for which the two models disagree the most about the degree of narrativity (Figures 4 to 8).

Weight?	Feature	Weight?	Feature
+0.568	he	+0.466	out
+0.483	<BIAS>	+0.405	me
+0.426	my	+0.378	was
+0.418	was	+0.318	door
+0.411	him	+0.304	<BIAS>
+0.376	had	+0.295	different
+0.362	derrick	+0.294	woman
+0.338	his	+0.288	my
+0.337	day	+0.283	plane
+0.306	her	+0.268	while
+0.281	out	+0.264	too
+0.273	while	+0.241	man
+0.250	they	+0.234	back
+0.245	she	+0.233	think
+0.243	down	+0.232	his
+0.243	me	+0.225	later
+0.242	them	+0.225	said
+0.242	plane	+0.223	she
+0.240	would	+0.216	us
+0.236	house	+0.215	know
+0.236	maddie	+0.214	him
+0.234	said	+0.213	one
+0.231	we	+0.201	honest
+0.215	again	+0.199	did
+0.214	staring	+0.199	boat
+0.213	freezer	+0.199	fairies
+0.205	do	+0.196	night
+0.202	army	+0.193	stood
+0.202	children	+0.185	head
+0.200	stood	+0.184	second
+0.197	kittens	+0.183	ali
+0.193	honest	+0.178	hallo
+0.191	recital	+0.178	traveller
+0.190	back	+0.178	facing
+0.189	thinking	+0.178	wider
+0.185	night	+0.178	you
+0.185	minutes	+0.177	would
+0.182	called	+0.175	stand
+0.181	more	+0.174	he
+0.179	really	+0.173	dad
+0.178	then	+0.169	clinton
+0.178	sam	+0.169	saw
+0.175	middle	+0.168	house
+0.175	benton	+0.168	sent
+0.175	clay	+0.167	down
+0.174	make	+0.167	tree
+0.174	stopped	+0.166	someone
+0.172	road	+0.165	going
+0.169	few	+0.164	fire
+0.169	red	+0.163	them
+0.167	living	+0.161	leonard
... 3352 more positive ...		... 3249 more positive ...	
... 4549 more negative ...		... 4582 more negative ...	
-0.152	forth	-0.115	ii
-0.152	meaning	-0.115	public
-0.153	religion	-0.116	themselves
-0.156	letter	-0.116	novel
-0.156	such	-0.116	work
-0.156	anything	-0.120	developer
-0.158	who	-0.122	lad
-0.159	pleasure	-0.123	lucid
-0.160	client	-0.124	less
-0.161	case	-0.126	such
-0.163	human	-0.127	caulaincourt
-0.164	states	-0.127	community
-0.166	platform	-0.127	securities
-0.166	subsided	-0.128	known
-0.166	grace	-0.128	thrombocytopenia
-0.166	delirium	-0.128	caterpillar
-0.166	yarmouth	-0.131	use
-0.166	civ	-0.133	itz
-0.166	leonora	-0.134	than
-0.167	may	-0.135	fee
-0.167	activity	-0.136	case
-0.176	set	-0.136	farmer
-0.177	itz	-0.137	liability
-0.178	are	-0.138	only
-0.179	term	-0.141	maddie
-0.181	makes	-0.143	board
-0.184	knowledge	-0.143	were
-0.185	germany	-0.143	better
-0.186	shall	-0.144	agreement
-0.190	fees	-0.145	finally
-0.191	ice	-0.145	makes
-0.194	dog	-0.147	shall
-0.197	mother	-0.150	control
-0.198	co	-0.151	john
-0.201	be	-0.151	fees
-0.202	longer	-0.154	are
-0.208	servant	-0.155	2d
-0.210	john	-0.155	sheriff
-0.215	by	-0.155	wages
-0.218	can	-0.155	law
-0.223	everything	-0.157	party
-0.229	in	-0.157	state
-0.233	even	-0.158	may
-0.262	agreement	-0.180	as
-0.264	this	-0.183	can
-0.274	2d	-0.190	for
-0.336	or	-0.200	or
-0.476	which	-0.253	is
-0.623	of	-0.291	by
-0.667	is	-0.359	of

Figure 3: Words positively and negatively associated with narrativity for the two models: text-based on the left (Piper) and reader-based on the right (Universals). Computed with the Python package ELI5

Piper | 1

tom had the day off from work . he wanted to just lounge around all day . he realized how messy his apartment was . tom decided to spend the day cleaning instead . he was a bit tired but felt accomplished .

Universals | .43

tom had the day off from work . he wanted to just lounge around all day . he realized how messy his apartment was . tom decided to spend the day cleaning instead . he was a bit tired but felt accomplished .

Figure 4: Passage id: f1979b1e-dd8f-4c91-b9f2-ade48d8b3596; genre: ROC

Piper | 1

the kids were at a petting zoo . they saw a very funny llama . they kept trying to pet it , but he would run away ! they spent all day trying to play with the llama . finally he let them pet his ear .

Universals | .43

the kids were at a petting zoo . they saw a very funny llama . they kept trying to pet it , but he would run away ! they spent all day trying to play with the llama . finally he let them pet his ear .

Figure 5: Passage id: 0f7c0e3f-3974-4271-af05-0d39cf3fba2e; genre: ROC

Piper | 1

barry was on his computer all day long . he worked writing a blog for a local newspaper . when barry would have breaks he would smoke cigarettes . barry tried to quit one day and struggled immensely . he was finally able to stop by substituting snacks for cigarettes .

Universals | .41

barry was on his computer all day long . he worked writing a blog for a local newspaper . when barry would have breaks he would smoke cigarettes . barry tried to quit one day and struggled immensely . he was finally able to stop by substituting snacks for cigarettes .

Figure 6: Passage id: ab7e5971-590b-4e59-8e47-ffd4f4c00e91; genre: ROC

Piper | .99

nick found $ 1000 . he had been poor all his life . he spent it all on lottery tickets . he lost all of it . he wished he had put it in the bank instead .

Universals | .42

nick found $ 1000 . he had been poor all his life . he spent it all on lottery tickets . he lost all of it . he wished he had put it in the bank instead .

Figure 7: Passage id: b49fdceb-29b8-4e83-9cf2-b3abd7b34aa5; genre: ROC

Piper |.92

he had been diagnosed with melanoma , and it had metastasized to his brain . he told her that he had only three months to live . she spent the rest of that fall flying back and forth from new haven to hilo , a journey of more than twelve hours . chunks of time spent at her father ' s bedside were interspersed with hours on the phone with cate . each day , cate would send her a new electron-density map by fax or email , and they would talk about ways to interpret it .

Universals | .34

he had been diagnosed with melanoma , and it had metastasized to his brain . he told her that he had only three months to live . she spent the rest of that fall flying back and forth from new haven to hilo , a journey of more than twelve hours . chunks of time spent at her father ' s bedside were interspersed with hours on the phone with cate . each day , cate would send her a new electron-density map by fax or email , and they would talk about ways to interpret it .

Figure 8: Passage id: Code-Breaker-The-Walter-Isaacson; genre: BIO