

Detecting Narrativity Across Long Time Scales

Andrew Piper, Sunyam Bagga, Laura Monteiro, Andrew Yang, Marie Labrosse and Yu Lu Liu

McGill University, 688 Sherbrooke St, H2J3B2 Montreal, Canada

Abstract

Storytelling is a universal human practice that serves as a key site of education, collective memory, fostering social belief systems, and furthering human creativity. It can occur in different discursive domains for different social purposes with differing degrees of intensity. In this project, we develop computational methods for measuring the degree of narrativity in over 335,000 text passages distributed across two- to three-hundred years of history and four separate discursive domains (fiction, non-fiction, science, and poetry). We show how these domains are strongly differentiated according to their degree of narrative communication and, second, how truth-based discourse has declined considerably in its utilization of narrative communication. These findings suggest that there has been a long-term historical differentiation between the practices of knowing and telling, which raises important questions with respect to the social acceptance of both science and the arts.

Keywords

narratology, history, systems theory, discourse analysis, computational narrative studies, digital humanities, natural language processing

1. Introduction

In his 1976 essay “Boundaries of Narrative,” Gérard Genette invited readers to “consider the principal plays of oppositions through which narrative defines and constitutes itself in the face of various nonnarrative forms” (p. 1) [8]. Over the ensuing decades, researchers have elaborated a variety of schemas to characterize narrative communication, creating a well-established corpus of theoretical work [9, 1, 7, 11]. Underpinning much of this work is the belief that there are intrinsic linguistic properties that predictably, if not universally, adhere within narrative forms of communication [30]. Narrative, according to these theoretical frameworks, is a detectable linguistic phenomenon.

One of the principal shifts to occur in the field of narratology over the past several decades has been an emerging understanding of narrative as a matter of degree rather than of kind [11, 10, 22]. “Narrativity” according to these theories is a quality that can best be understood not as a global binary class (a document either is or is not narrative), but as a local, multi-dimensional scalar property. As Ochs and Capps [19] write, “We believe that narrative as genre and activity can be fruitfully examined in terms of a set of dimensions that a narrative displays to different degrees and in different ways” (p. 19). In this sense, a narrative document, such as a novel, may exhibit greater or lesser degrees of narrativity at different moments

CHR 2021: Computational Humanities Research Conference, November 17–19, 2021, Amsterdam, The Netherlands

✉ andrew.piper@mcgill.ca (A. Piper); sunyam.bagga@mail.mcgill.ca (S. Bagga);
laura.monteiro@mail.mcgill.ca (L. Monteiro); andrew.yang3@mail.mcgill.ca (A. Yang);
marie.labrosse@mail.mcgill.ca (M. Labrosse); yu.l.liu@mail.mcgill.ca (Y.L. Liu)

📄 0000-0001-9663-5999 (A. Piper)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

in the text, just as ostensibly non-narrative documents, such as scientific reports, may also exhibit degrees of narrativity. Herman [11] has taken this understanding one step further to suggest that narrativity is not simply a matter of the local interplay of formal and textual features, but emerges through the interaction between readers and texts. Narrativity can thus be understood as a potentially rising or falling quality within documents (or other forms of communication) that depends on the interaction of different linguistic or semiotic features combined with readers' responses.

While a wealth of recent work in the field of natural language processing has engaged with the detection of different dimensions of narrativity (such as causal and temporal relations [18], turning points [21, 2], reportable events [20], frames [6], etc.), no work to our knowledge has undertaken the more elementary task of narrativity detection itself. Can we reliably predict whether a span of text is engaging in narrative communication and if so, to what degree of intensity? Such work has the potential not only to contribute to our theoretical understanding of narrativity as a form of communication. It can also provide empirical insights into the distribution of narrativity across different discursive domains and time periods giving us insights into the social functions of narrative communication over time. The latter will be our concern here.

In this paper, we develop computational models to detect “narrativity” as a local, multi-dimensional textual quality across four different discursive domains over a two- to three-hundred year time-period. Our aim in doing so is to test the relationship between narrative communication and the process of functional differentiation among social systems as theorized by the sociologist Niklas Luhmann [16]. According to Luhmann, social systems are governed by communicative practices (“codes”) that maintain a system’s internal coherence in distinction from other systems. As societies modernize, differentiation strengthens over time as each system evolves to maintain its internal coherence in distinction from its environment (i.e. other systems).

The question we wish to test here is whether the practice of narrative participates in this process of functional differentiation between the social systems Luhmann labels “art” and “science.” According to Luhmann, art’s function lies in its ability to communicate the sensory process of observation, to “allow a world to appear within the world” (p. 241), whereas the function of science is to “structure the field of possible statements [about the world] with the help of the code true/untrue” (p. 227) [15]. As a form of communication strongly associated with the idea of “world-building” [11], we would thus expect narration to be highly associated with artistic forms of expression but not necessarily negatively associated with scientific discourse. There is nothing intrinsic to narrative communication that makes it an inappropriate vehicle for fact-based discourse. After all, one can tell true and untrue stories.

And yet according to the data and models used here, we can observe a very clear historical trajectory of the *de-narrativization of truth-based discourse*. Our findings bring to light longstanding and growing tensions between what Hayden White first introduced as the relationship between *knowing* and *telling* [29]. For White, the function of narration should be understood as “a solution to a problem of general human concern, namely, the problem of how to translate knowing into telling, the problem of fashioning human experience into a form assimilable to structures of meaning that are generally human rather than culture-specific” (p. 5) [29]. Narration is a key communicative mode for White that makes knowledge “assimilable” to individual human beings and collective societies. As a growing body of research has indicated, narrative is indeed an effective means of addressing the problem of knowledge sharing across a variety of social domains, from economics [25] to climate change [5] to political polarization

[13].

Our findings, preliminary as they are, suggest the need for further research into this disassociation of science and narration and its potential social effects. Is the growing public distrust in science related to the denarrativization of scientific communication? Do efforts of “public science” or science journalism have a positive effect on reversing public distrust and are such effects related to their degree of narrativity? If narration is increasingly seen as belonging to the domain of art, has this differentiation from science unintentionally contributed to the devaluation of the arts (or their study)? How might the arts instead participate more explicitly in the process of knowledge transfer, i.e. help “recouple” in Luhmann’s terms the practices of knowing and telling?

In order to detect narrativity in our historical collections, we undertake the following steps, which we describe in greater detail in the following sections. First, we construct a data set of 335,245 documents to represent our two primary social systems of art and science, which we subdivide into four domains of fiction, poetry, science, and non-fiction. We then develop a working theory of “narrativity” drawn from existing theoretical literature that informs our manual annotation of the data. Building a team of three trained student annotators, we hand-annotate 401 passages according to a scalar understanding of narrativity derived through numerous meetings and discussion. This data is then used to train and test our machine learning models, which we describe in Section 3. We present our results in Section 4 and include a discussion of their potential implications as well as limitations (Section 5). Finally, we conclude with a brief discussion of where future work in computational narrative studies might lead.

2. Data

2.1. Annotated Data

In order to annotate training data for the presence of narrativity, we rely on the following theoretical schema developed by Herman [11]. According to this schema, narrative communication consists of the following four elements:

1. Situatedness: narrativity depends on the social context in which it occurs
2. Event sequencing: narrativity depends on temporally ordered events
3. World making: narrativity depends on the fact of disequilibrium such that we can observe a change in the world
4. Feltness: narrativity captures the *experience* of events, i.e. “what it is like”

Herman’s categories can be seen as syntheses of previous narratological frameworks, capturing a good degree of consensus in the field. The emphasis on feltness, for example, is strongly indebted to the argument by Fludernik [7] that “Experientiality reflects a cognitive schema of embodiedness that relates to human existence and human concerns” (p. 9). Similarly, event-sequencing is strongly indebted to the work of theorists like Genette [9], Sternberg [26, 27], and Ricoeur [24] and their emphasis on temporality as a central component of narrative communication, while world making derives from the work of Labov and Waletzky [14] and Bruner [3].

In general then, Herman’s model is guided by the notion that, “*Narrative roots itself in the lived, felt experience of human or human-like agents interacting in an ongoing way with their*

cohorts and surrounding environment” (our emphasis). Thus for Herman what matters most about narrativity is: a) the centralization of one or more agents; b) the sequencing of events and thus time; and finally, c) the idea of “lived experience in an environment”, i.e. a sense of world building.

Based on this theoretical framework, we hand-annotate 401 passages drawn from the experimental data using the following steps:

First, we assembled a team of three annotators who all have majors in the humanities. These are readers who have high levels of education and exposure to training in textual analysis. Second, over the course of several weeks we engaged in discussions and experiments regarding the concept of “narrativity” with respect to the theoretical framework discussed above as well as different kinds of text passages. These discussions culminated in a codebook, which is included in the supplementary material.¹ Annotators were then asked to code a given passage across three dimensions of narrativity, which were defined for the annotators as “agency,” “event sequencing,” and “world making.” Note how we translated Herman’s “feltness” into “agency” to better account for the idea of experientiality at the heart of most major narrative theories.

For each passage, readers were asked to respond to the following statements using a five-point Likert scale:

- “This passage foregrounds the lived experience of particular agents.” (Agency)
- “This passage is organized around sequences of events that occur over time.” (Event sequences)
- “This passage creates a world that I can see and feel.” (World making)

1. Strongly disagree
2. Somewhat disagree
3. Unsure
4. Somewhat agree
5. Strongly agree

Notice how we do not expressly ask if readers felt that the passage was “narrative” or not. Rather, we ask them to consider their feelings with respect to these three primary narrative dimensions, which we then average into a single “narrativity” score. We found that this increased reader agreement and allowed for more nuanced understandings of narrative communication. For example, it was not uncommon for some types of discourse to emphasize sequential events but lack an emphasis on agency or building a world.

We provide a few sample passages that received low and high average narrativity scores by readers. Note that passages have been truncated from their actual length.

Non-Fiction - Average Reader Score 1.2 *The employment of the uninterpretable symbol in the intermediate processes of trigonometry furnishes an illustration of what has been said. Lapprehend that there is no mode of explaining that application which does not covertly assume the very principle in question. But that principle, though not, as I conceive, warranted by formal reasoning based upon other grounds, seems to deserve a place among those axiomatic*

¹Note that we provide the reader-annotated data, annotator’s codebook, metadata, code for all models and concrete implementation details of custom features from Table 1 in the Supplementary Material. It is available at <https://doi.org/10.7910/DVN/DAWVME>

truths which constitute in some sense the foundation of general knowledge, and which may properly be regarded as expressions of the mind's own laws and constitution.

Fiction - Average Reader Score 1.44 *It is too weak for a shield, too transparent for a screen, too thin for a shelter, too light for gravity, and too threadbare for a jest. The wearer would be naught indeed who should misbeseem such a wedding garment. But wherefore does the sheep wear wool? That he in season sheared may be, And the shepherd be warm though his flock be cool.*

Science - Average Reader Score 4.55 *I assisted at the opening of her Body, and having found in the matrix a little round mass of the bigness of a great black Cherry, I took the husband aside, and asked him, Num a tempore fluxus menstruorum uxorem cognevisset? And having received for answer, that he had, I prayed him to let me carry home with me this little ball, which I had found in her womb. I was no sooner come home but I opened it, and found, that nature had wrought with so much activity in so small a time...*

Fiction - Average Reader Score 4.55 *Whereupon a sudden outcry arose within the house, and a head popped angrily out of the aperture so suddenly created. But as instantly it returned within. For Jorian tossed the lattice to the ground by the door and thrust his spear-head into the cravat of red which the man had about his throat, shouting to him all the while in the name of the Prince, of the Duke, of the Emperor, of the Archbishop, of all potentates, lay and secular, to come down and open the gates.*

Because our annotations use a multi-point scale, we assess inter-rater reliability (IRR) using the average deviation index as discussed in Burke, Finkelstein, and Dusig [4]. We report an average deviation of 0.48 (± 0.27). This indicates that on average our annotators' judgments per passage fall within just under 0.5 points of each other on our 5-point Likert scale, suggesting reasonable levels of agreement. A one-way ANOVA was conducted to compare the effect of genre on average deviation among annotators, with a significant effect observed [$F(3, 397) = 7.56, p = 6.26e - 05$], with poetry generating significantly more deviation among annotators than the other genres (mean AD of 0.58). We also note that as seen in Figure 1, annotator scores were not normally distributed around 3.0, but rather exhibit a skewed central tendency between 2.0 and 2.5. 65% of the annotations were below 3, suggesting there were only a minority of passages where our annotators were confident of the passage's narrativity.

2.2. Experimental Data

Our experimental data consists of five separate collections that are designed to capture the two social systems of "art" and "science," which we represent as the four discursive domains of fiction, non-fiction, poetry, and science. Doing so allows us to see aggregate behavior across the two systems as well as potential internal differences based on discourse type. Our data consists of:

- **Fiction & Non-Fiction.** This data is derived from the Hathi Trust Digital Library and is drawn from Piper and Bagga [23]. It encompasses 85,130 passages of fiction and 99,968 passages of non-fiction spanning the years 1800-1999 written in English. The labels are generated using modified predictive models based on prior work [28]. The distribution

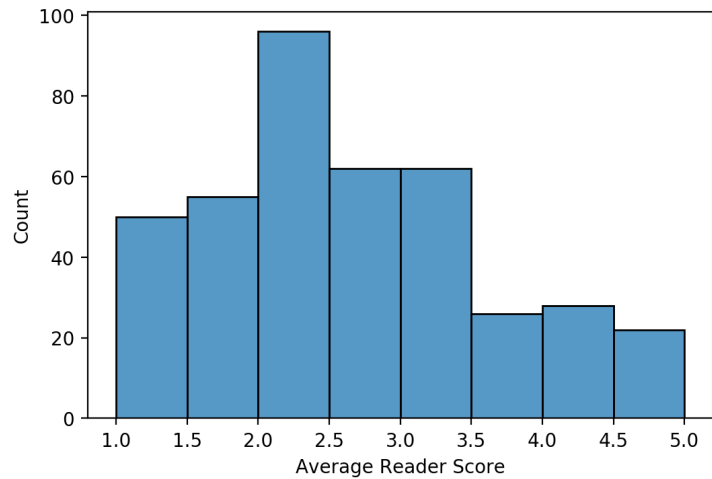


Figure 1: Distribution of the averaged annotator scores for “narrativity.”

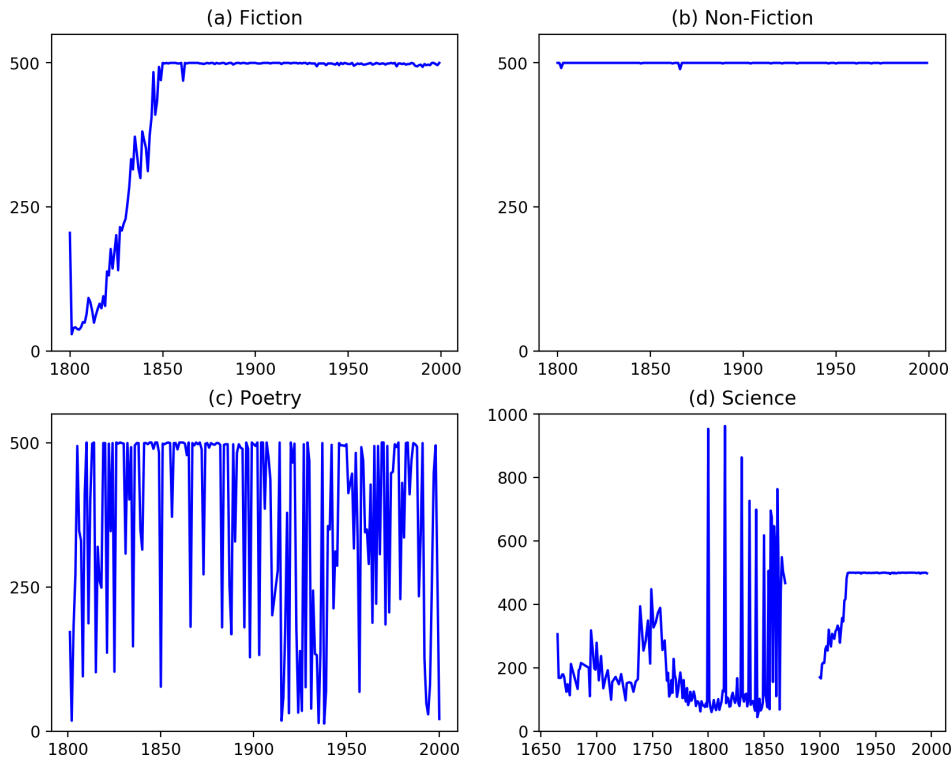


Figure 2: Distribution of the number of passages per year for each of the four domains of (a) Fiction, (b) Non-Fiction, (c) Poetry, and (d) Science.

of the number of passages per year is indicated in Figure 2. Years represent year of publication, not year of composition or first printing. Our data reflects reading material available in a given year as archived by academic libraries.

- **Poetry.** This data is drawn from the Literature Online Poetry database. It consists

of 73,077 poems by 857 authors who wrote in English and who were alive during the nineteenth and twentieth centuries. To estimate year of publication, we use the author’s birth-date plus 35 years to capture an estimated career midpoint. Because of the relatively small number of poets in our dataset, we are not able to capture a consistent number of poems per year.

- **Science.** To represent the domain of scientific writing, we use two different data sets. The first is drawn from the Royal Society Corpus (RSC 4.0) based on the first two centuries of the Philosophical Transactions of the Royal Society of London from its beginning in 1665 to 1869 [12]. Due to copyright restrictions, no data is publicly available after 1869. This dataset consists of 31,698 documents. To augment this data, we use a collection of 45,439 randomly selected articles drawn from top 100 most common articles in the JSTOR Data for Research platform organized under the heading “physical sciences” published between the years 1900 and 2015. The distribution of articles over time is captured in Figure 2.

Because our interest is in local narrativity, i.e. the extent to which a span of tokens expresses narrative communication, we represent our documents as randomly selected sequences of 5 sentences in length. This number has been indicated in prior work as a reasonable frame in which completed narratives can transpire [17]. We can thus assume that “narrativity” can be present in spans of this length. Future work will want to explore this parameter further.

3. Models

For the purposes of our project, we use the predicted probability of a passage’s “narrativity” as an indicator of the degree of narrative communication present in that passage. In order to build a model to predict a passage’s narrativity, we train and validate our models using our reader annotated data. We experiment with three widely-used algorithms (Logistic Regression, Support Vector Machines, and Random Forests) and multiple combinations of different features to identify the best-performing model. We present our feature components in Table 1² and present the performance of each model using different feature combinations in Figure 3. As can be seen in Figure 3, Random Forest performs the best out of the three learning algorithms. Table 2 presents a brief overview of the top-5 performing models using Random Forest.

We assess model performance according to Pearson’s correlation coefficient rather than the more traditional F1 score (although we also report traditional classification metrics in Table 2). Because our metric of narrativity is predicted probability and not a binary classification, the question we want to address is how well our models correlate with the *scalar* nature of reader judgments.

To construct our experimental feature spaces, we aggregate our features into three general categories: lexical features (ngrams), syntactical features (part-of-speech and dependency relationships), and higher-level custom features designed to capture specific narratological theories, including time, concreteness, animate entities and perceptuality. For a full discussion of the custom features, see the supplementary material.

As we can see in Figure 3, all three classifiers behave similarly and achieve their maximum performance on a variety of feature combinations. Interestingly, unigrams tend to perform

²Note that the results shown in Figure 3 correspond to a maximum of 100 features per category. This is why # Features for word-bigrams, for example, is 100 although the complete feature space involved 25,434 word-bigrams. Experiments with other values of max-features yielded similar results.

Table 1

Description of the individual feature categories used by our learning models.

Feature-Category	Description	# Features
pos1	part-of-speech unigrams	38
pos2	part-of-speech bigrams	100
pos3	part-of-speech trigrams	100
pos23	part-of-speech bigrams & trigrams	100
dep1	dependency tag unigrams	45
dep2	dependency tag bigrams	100
dep3	dependency tag trigrams	100
dep23	dependency tag bigrams & trigrams	100
word1	word unigrams	100
word2	word bigrams	100
word3	word trigrams	100
word23	word bigrams & trigrams	100
Tense	tokens annotated with time-related tags by NER	1
Mood	measures of setting, concreteness, eventfulness and saying	4
Voice	animate entities and perceptual vocabulary	4
% Quoted	ratio of words in dialogue	1

Table 2

Random Forest’s performance with different feature-combinations on the reader-annotated passages using 5-fold cross validation. The top-5 performing Random Forest models are shown here (see Figure 3 for all feature spaces.) Note that TMV is used as shorthand for Tense + Mood + Voice features.

Feature-Set	Pearson-r	F1-Score	Precision	Recall
pos1 + TMV + Pct-Quoted	0.742	0.787	0.796	0.780
pos1 + TMV	0.740	0.788	0.801	0.780
All Categories	0.735	0.790	0.798	0.785
pos1 + Mood	0.732	0.783	0.803	0.770
pos1 + dep1 + TMV	0.717	0.782	0.799	0.770

better than the limited sets of bi- or trigrams for lexemes, pos, and dependency tags. The best performing model consists of part-of-speech unigrams, % dialog and custom-built features that aim to capture “event sequences”, “world building”, and “agency” for which we use the categories *tense*, *mood*, and *voice*. There appears to be a strong grammatical signature to narrativity that marginally grows in strength when we add in features that capture the notion of “environment” emphasized in Herman’s theory above [11]. We leave a deeper exploration of these issues to future work.

As we can see in Figure 4, the correlation between reader judgments and predicted probability is approximately linear and indicates a reasonable level of agreement ($r = 0.742$). We observe

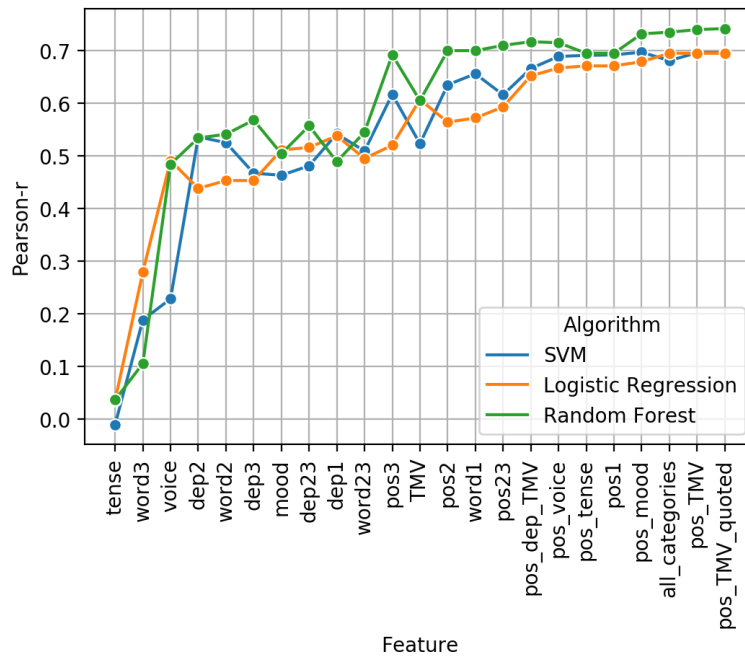


Figure 3: A comparison of different learning algorithms using different combinations of features. Pearson-r is computed on 5-fold cross-validated predictions.

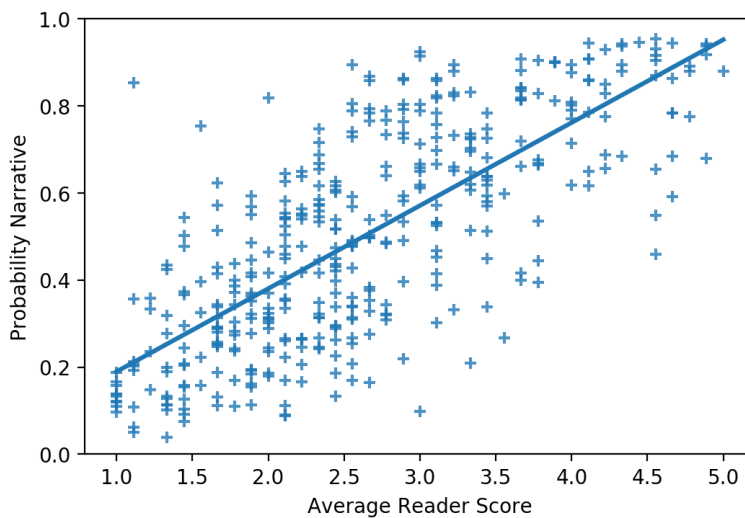


Figure 4: Visualization of the correlation between the averaged reader-annotations and our best model’s predicted probability of a passage’s “narrativity” ($r = 0.742$).

higher levels of variability in the middle range of annotations between 2.0 and 3.0, which is to be expected. As readers’ judgments become less certain, so too do we observe more variability in our models’ predictions. Future work will want to explore the extent to which more annotations lead to higher levels of correlation or whether we achieve some kind of maximum level of correlation between computational models and human judgments in this area.

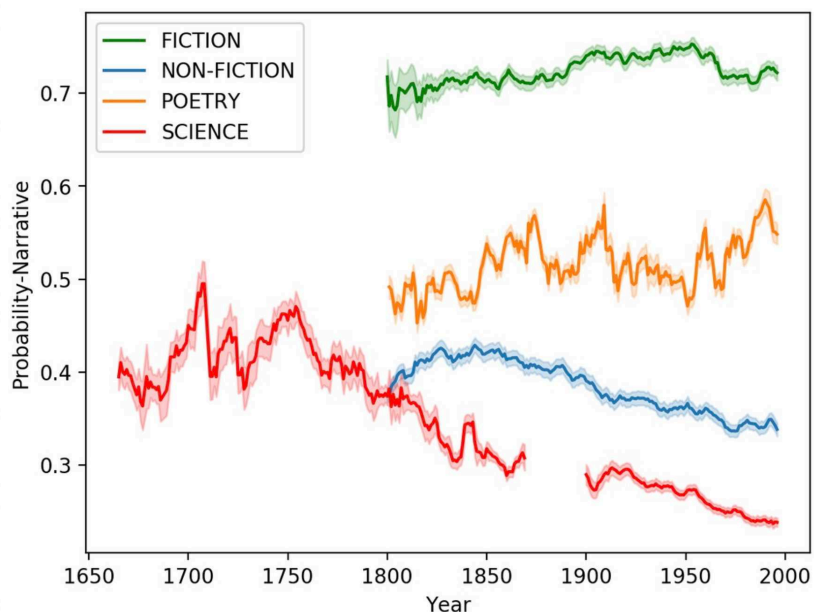


Figure 5: Five-year rolling averaged yearly probability of the text being narrative across all four domains by the best-performing model: pos1 + TMV + Pct-Quoted.

4. Results

Applying our best performing model (Random Forest with POS-unigrams + Pct-Quoted + *tense*, *mood* and *voice* features) on the experimental data described above, we generate the average yearly predicted probability of narrativity across all four domains as shown in Figure 5. According to our models the four domains behave in distinctive fashion with respect to narrativity, providing support for the idea that narrativity may be another facet underlying Luhmann’s thesis about functional differentiation [16]. Second, with respect to the fact-based domains of nonfiction and science, we also observe meaningful decays in the estimated intensity of narrativity over our time period. For science, we see a drop from an average five-year high of 0.495 estimated narrativity in 1707 to a five-year low of 0.236 in 1994, while for non-fiction we see a less dramatic decline from 0.428 (in 1844) to 0.338 (in 1996). Because our non-fiction class can potentially contain scientific reports archived in Hathi, we cannot definitively tell if this decline of narrativity in non-fiction is due to the growth of science writing in Hathi or the decline of narrativity in non-scientific forms. While we provide some validation of this in the discussion, future work in this direction will depend on more fine-grained genre classification with respect to non-fictionality.

In terms of our two “literary” domains, we see little change over time, suggesting relative stability of these domains’ relationship to narrativity. While this does not run counter to expectations with respect to fiction, theorists of poetry might be surprised to see such continuity given the popularity of long narrative poems in the nineteenth century (for example in the work of Walter Scott or Longfellow to name two prominent examples). Future work will want to explore in greater depth whether there is a meaningful break with respect to poetic narrativity for authors born after the late nineteenth-century that then potentially reverses course for younger poets born closer to the end of the twentieth-century as indicated in Figure 5. More

domain-specific training data would be needed along with more careful sampling techniques to gain confidence about any such shifts given how slight they are with respect to our models.

Taken altogether, our models suggest that narrativity is strongly socially differentiated across different discursive domains and that at least with respect to fact-based discourses this differentiation is increasing strongly over time as both non-fiction writing and specifically scientific writing exhibit declines in their reliance on narrative communication. We take up the implications of these findings in our discussion.

5. Discussion

Our study raises a number of questions for future research. Representing the social systems of “art” and “science” is a challenging task. In our work, we have tried to capture at least two larger subdomains of writing within these systems to better understand the kind of internal differentiation may be at work. While future work will want to experiment with different samples dependent on different archival resources, we do observe interesting differences with respect to the narrative behavior of our subdomains. For example, we see how the narrativity of poetry is not only considerably lower than prose fiction, it consistently hovers around the fifty-percent mark, suggesting that one of the potential social functions of poetry as a genre could be its ability to communicate narrative ambiguity. Such ambiguity is corroborated by the higher average deviation among our annotators with respect to the poetry training data. This suggests an interesting potential theory one could pursue for the future study of poetry in a larger social context along with the potential increase in narrativity that we observe for poets born after the 1960s.

Because our models have been trained to understand cross-domain behavior of narrativity, our work cannot however speak to within-domain distinctions with respect to narrativity. For example, an interesting question to be pursued in the future is the intensity and extent of narrativity at the document level within our different discourses. When do we see novels for instance engage in more explicitly “narrative” communication, are there reliable patterns of the rise and fall of narrativity, or finally what kinds of novels (genres) indicate greater degrees of narrativity overall? Similarly, for science documents while we observe an overall process of denarrativization of scientific documents, are there still portions of articles that engage in more narrative-like behavior or portions of the field (i.e. disciplines) that engage in more narrative communication than others? These questions would help provide insights into the relationship between categories like genre, discipline and narration.

Further reflection could also be given to our framework of “truth-based discourse,” which is not exactly synonymous with “science,” which is one of the reasons we also model “non-fictional” writing as well. Scientific writing is one kind of communication that makes truth claims, but there are numerous others that belong to different institutional frameworks. We note that in a random sample of two-hundred passages drawn from our non-fiction experimental data that the number of “scientific” texts moves from 6 in the nineteenth century to 10 in the twentieth. While this represents a large increase, it is still a very small fraction of all writing in our non-fiction sample suggesting that the decline in narrativity in non-fiction is not strongly related to a rise of scientific writing in Hathi Trust. In other words, writing classified as non-fiction is exhibiting similar trends to science but is being produced in different institutional contexts. Future work could explore more deeply why we see this denarrativization of non-fiction along with science writing.

At the level of data annotation, while we demonstrate solid agreement between readers and reasonable model correlation with reader judgments, we are not able to annotate large amounts of data to better calibrate our models. Hand-annotation is a slow and expensive process and future work will want to explore mechanisms that allow for scaling annotation while maintaining quality. We assume model accuracy will increase with increased amounts of annotated data, which may or may not have a bearing on the historical trends we observe here. The observed declines in our science and non-fiction corpora are so steep and consistent that we would be surprised if future work indicated significant changes in this regard.

In terms of our theoretical framework, it is important to underscore that our approximation of narrativity is just that. While we do not observe significant shifts in the distribution of narrativity according to feature-space selection, our models are still guided by a particular theoretical framework with respect to narrativity. Future work will want to explore alternative theories and feature representations of narrativity to see if the historical trends we are observing continue to emerge.

Finally, future work will want to explore the extent to which our findings are or are not culturally specific, i.e. the extent to which they hold in other language communities and the extent to which those correlations are driven by social factors such as national wealth or education levels. Just how universal is this process of functional differentiation and denarrativization with respect to truth-based discourses that we have observed here? Is this indeed a marker of “modernization”?

6. Conclusion

Our work has demonstrated that narrative communication is a detectable linguistic quality of texts from the perspective of human readers and machine learning. Even with a small set of training data we can achieve reasonable levels of predictive accuracy and correlation with trained reader judgments across very different kinds of texts over relatively long historical time spans. We also show that with sufficient training readers can agree quite well regarding the intensity of a passage’s narrativity.

Being able to identify the degree of narrativity in large-scale historical document collections allows us to gain a better understanding of the distribution of narrative communication across documents that serve different social functions. Modeling narrativity at the computational level suggests that narrative is a form of communication that participates in Luhmann’s theory of functional differentiation, at least with respect to the social systems of art and science. While culturally and historically universal – narration is present in all linguistic communities and recorded time periods – narration is far from being socially universal. Indeed, it appears that in modern, highly differentiated societies narration is increasingly aligned with the particular social system of the arts as truth-based discourse becomes less and less narrativized over time. How this may impact urgent large-scale questions such as trust in science or particular collective responses to social problems such as climate change or health pandemics remains an open, yet important question for future research.

References

- [1] M. Bal. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, 2009.

- [2] R. L. Boyd, K. G. Blackburn, and J. W. Pennebaker. “The Narrative Arc: Revealing Core Narrative Structures through Text Analysis”. In: *Science Advances* 6.32 (2020), eaba2196.
- [3] J. Bruner. “The Narrative Construction of Reality”. In: *Critical Inquiry* 18.1 (1991), pp. 1–21.
- [4] M. J. Burke, L. M. Finkelstein, and M. S. Dusig. “On Average Deviation Indices for Estimating Interrater Agreement”. In: *Organizational Research Methods* 2.1 (1999), pp. 49–68.
- [5] S. Bushell, G. S. Buisson, M. Workman, and T. Colley. “Strategic Narratives in Climate Change: Towards a unifying narrative to address the action gap on climate change”. In: *Energy Research & Social Science* 28 (2017), pp. 39–49.
- [6] N. Chambers and D. Jurafsky. “Unsupervised Learning of Narrative Event Chains”. In: *Proceedings of ACL-08: HLT*. Columbus, Ohio: Association for Computational Linguistics, 2008, pp. 789–797.
- [7] M. Fludernik. *Towards a ‘Natural’ Narratology*. Routledge, 2002.
- [8] G. Genette. “Boundaries of Narrative”. In: *New Literary History* 8.1 (1976), pp. 1–13.
- [9] G. Genette. *Narrative Discourse: An Essay in Method*. Vol. 3. Cornell University Press, 1983.
- [10] R. Giora and Y. Shen. “Degrees of Narrativity and Strategies of Semantic Reduction”. In: *Poetics* 22.6 (1994), pp. 447–458.
- [11] D. Herman. *Basic Elements of Narrative*. John Wiley & Sons, 2009.
- [12] H. Kermes, S. Degaetano-Ortlieb, A. Khamis, J. Knappen, and E. Teich. “The Royal Society Corpus: From Uncharted Data to Corpus”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*. 2016, pp. 1928–1931.
- [13] E. Kubin, C. Puryear, C. Schein, and K. Gray. “Personal Experiences Bridge Moral and Political Divides Better than Facts”. In: *Proceedings of the National Academy of Sciences* 118.6 (2021).
- [14] W. Labov and J. Waletzky. “Narrative Analysis: Oral Versions of Personal Experience.” In: (1967).
- [15] N. Luhmann. *Die Kunst der Gesellschaft*. Suhrkamp, 1995.
- [16] N. Luhmann. *Social Systems*. Stanford University Press, 1995.
- [17] N. Mostafazadeh, N. Chambers, X. He, D. Parikh, D. Batra, L. Vanderwende, P. Kohli, and J. Allen. “A Corpus and Cloze Evaluation for Deeper Understanding of Commonsense Stories”. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016, pp. 839–849.
- [18] N. Mostafazadeh, A. Grealish, N. Chambers, J. Allen, and L. Vanderwende. “CaTeRS: Causal and Temporal Relation Scheme for Semantic Annotation of Event Structures”. In: *Proceedings of the Fourth Workshop on Events*. San Diego, California: Association for Computational Linguistics, 2016, pp. 51–61. DOI: 10.18653/v1/W16-1007. URL: <https://www.aclweb.org/anthology/W16-1007>.

- [19] E. Ochs and L. Capps. *Living Narrative: Creating Lives in Everyday Storytelling*. Harvard University Press, 2009.
- [20] J. Ouyang and K. McKeown. “Modeling Reportable Events as Turning Points in Narrative”. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, 2015, pp. 2149–2158. DOI: 10.18653/v1/D15-1257. URL: <https://www.aclweb.org/anthology/D15-1257>.
- [21] P. Papalampidi, F. Keller, and M. Lapata. “Movie Plot Analysis via Turning Point Identification”. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, 2019, pp. 1707–1717. DOI: 10.18653/v1/D19-1180. URL: <https://www.aclweb.org/anthology/D19-1180>.
- [22] F. Pianzola. “Looking at Narrative as a Complex System: The Proteus Principle”. In: *Narrating Complexity*. Springer, 2018, pp. 101–122.
- [23] A. Piper and S. Bagga. *HATHI 1M: Million Page Historical Prose Data in English from the Hathi Trust*. 2021.
- [24] P. Ricoeur. *Time and Narrative, Volume 1*. University of Chicago press, 2012.
- [25] R. J. Shiller. *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press, 2020.
- [26] M. Sternberg. “Telling in Time (I): Chronology and Narrative Theory”. In: *Poetics Today* 11.4 (1990), pp. 901–948.
- [27] M. Sternberg. “Telling in Time (II): Chronology, Teleology, Narrativity”. In: *Poetics Today* 13.3 (1992), pp. 463–541.
- [28] T. Underwood, P. Kimutis, and J. Witte. “NovelTM Datasets for English-Language Fiction, 1700-2009”. In: *Journal of Cultural Analytics* 5.2 (May 28, 2020). DOI: 10.22148/001c.13147.
- [29] H. White. “The Value of Narrativity in the Representation of Reality”. In: *Critical Inquiry* 7.1 (1980), pp. 5–27.
- [30] S. Zeman. “Grammatik der Narration”. In: *Zeitschrift für germanistische Linguistik* 48.3 (2020), pp. 457–494.