Modeling Narrative Revelation

Andrew Piper, Hao Xu and Eric D. Kolaczyk

McGill University, Montreal, QC H3A 2M7, CANADA

Abstract

A core aspect of human storytelling is the element of narrative time. In this paper, we propose a model of narrative revelation using the information-theoretic concept of relative entropy, which has been used in a variety of settings to understand textual similarity, along with methods in time-series analysis to model the properties of revelation over narrative time. Given a beginning state of no knowledge about a story (beyond paratextual clues) and an end state of full knowledge about a story's contents, what are the rhythms of dissemination through which we arrive at this final state? Using a dataset of over 2,700 books of contemporary English prose, we test for various time-dependent characteristics of narrative revelation against four stylistic categories of interest: audience age level, prestige, point-of-view, and fictionality.

Keywords

narratology, information theory, contemporary literature, discourse structure, narrative revelation

1. Introduction

Italo Calvino was fond of quoting a Sicilian expression that "time takes no time in a story" [8]. A narrator can tell a story that traverses centuries in a few sentences or can slow time down to the point where a few seconds takes minutes to describe. Such manipulations of time – one of the great loves of narrative theory [28, 31, 32] – hide a more elementary fact about stories: no matter how much they may compress or dilate time, they still take time to tell. All stories, even the shortest, happen in time and cannot be told all at once.

The fact that stories take time means that the dissemination of information – the ordering and divulging of facts about the storyworld – plays an important role in the meaning of the story. Independent of *what* is told, *how* it is told is a key aspect of a story's meaning. Narrative theorists refer to this discrepancy as "discourse structure" [13, 1, 6] and it has largely been framed as an ordering problem, i.e. the discrepancy between how narrative information is revealed and the underlying logic of events within the story. A sizable body of empirical studies has shown, for example, the way modulations in narrative order – such as withholding salient information or reordering events in non-linear fashion – can influence the emotional or affective response of audiences [6, 3].

Less attention has been paid to the more elementary question of the amount of novel information imparted at any given moment in a story. Given a beginning state of no knowledge about a story (beyond paratextual clues [14]) and an end state of full knowledge about

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France
andrew.piper@mcgill.ca (A. Piper); hao.xu6@mail.mcgill.ca (H. Xu); eric.kolaczyk@mcgill.ca (E. D. Kolaczyk)
0000-0001-9663-5999 (A. Piper)

a story's contents, what are the rhythms of dissemination through which we arrive at this final state? Does a narrator introduce more information early and then spend time going over more familiar terrain or, conversely, withhold key pieces of information that we only learn about towards the close of the story? Are there periods of local exploration, where narrators spend more time introducing novel information, and periods of exploitation (to borrow a classic framework from computer science), where narrators immerse audiences in already established characters, themes, and situations? Do these practices exhibit predictable, periodic behavior or are they more akin to random walks? Finally, how are such practices impacted by the social situatedness of narrative [15]? When a narrator is crafting a story for younger audiences or telling a true versus fictional story or appealing to literary elites on prize committees, do we see modulations in the way narrative information is revealed?

In this paper, we draw on the fields of information theory and statistics (including time series analysis, in the latter case) to develop a model of narrative revelation to capture the relative amount of new information communicated by authors over narrative time. We use the information-theoretic concept of relative entropy or Kullback-Leibler divergence to quantify how much new information is introduced in a given book during a window of text at a given time T, relative to the prior window at time $T - 1$. We then employ various techniques from statistics and time series analysis to characterize the temporal dynamics of the resulting traces, at both the aggregate level (across our corpus) and at the level of individual books. Relative entropy has been applied to the study of textual difference in numerous settings [9], including parliamentary discourse [2] and the evolution of scientific English [10, 4], as well as been shown to be a good predictor of human visual attention [16], linguistic processing [18, 19], and has more recently been proposed as a model of implicit cultural learning [34]. Statistics – and in particular, time series analysis – provides us with a well-developed set of tools for detecting and describing aspects of the temporal behavior in the relative entropies for our corpus, such as trend, periodicity, and statistical dependency of the present on the past (e.g., [7]).

We apply our measure of narrative revelation to the CONLIT dataset [24], which includes approximately 2700 books from 12 genres drawn from contemporary English prose published since 2001. We use available partitions in the data to test the relationship between patterns of narrative revelation and different social categories. In particular, we concentrate on the following categories in our analysis, including the relevant classes from the CONLIT data: fictionality (fiction / non-fiction), prestige (prizewinning novels / bestsellers), age level (YA + Middle School / Adult Fiction), and point-of-view (first person / third person). Note that all but the first condition on fictional narratives.

Understanding the dynamics of narrative revelation can provide an important window into the nature of human storytelling using computational methods. First, it can provide an objective measure of informational novelty within texts, which can then be associated with reader judgments. While beyond the scope of the present work, future work will want to explore this relationship between the rate of novel information and readers' affective states. Such a measure can also provide insights into the effects that social settings have on the revelation of narrative information, such as audience type or the narrator's goals regarding the instrumentality of information being communicated (facticity/fictionality), as well as potentially reveal audience preferences for story structure when it comes to the distribution of new information. In particular, it can give us the means to model what is known as the explore/exploit trade-off when it comes to narrative communication [33]. When telling a story we assume that narrators will oscillate between periods of exploration (introducing and developing novel ideas and characters) and periods of exploitation (deepening our understanding/attachment to the agents and experiences already introduced). And yet we currently have little knowledge about how these relationships evolve over narrative time as it relates to long narrative forms and whether social factors impact this behavior. Our work thus attempts to provide a novel method for modeling the dissemination of information over narrative time further contributing to more general inquiry into the temporal properties of human storytelling.

2. Related Work

A number of approaches to the computational modeling of discourse structure have been proposed. Schmidt [30] used topic modeling to identify thematic arcs in television screenplays, while Thompson, Wojtowicz, and DeDeo [33] used topic models to study thematic progression in philosophical texts and social media. Reagan, Mitchell, Kiley, Danforth, and Dodds [27] used sentiment analysis to model the concept of narrative fortune [12], for which Elkins [11] provides a more in-depth study of the validity of sentiment arcs as models of narrative structure. Boyd, Blackburn, and Pennebaker [5] used particular word types to capture three primary narrative stages, and Sap, Jafarpour, Choi, Smith, Pennebaker, and Horvitz [29] used the predictability of next sentences to capture the concept of narrative "flow," though this is not applied to questions of narrative time. Piper and Toubia [26] used word embeddings to model narrative non-linearity using the traveling salesman problem. Ouyang and McKeown [22] and Piper [23] devised methods for predicting narrative "turning points" as larger structural qualites, drawing on Aristotelian and Augustinian theories of narrative respectively. Finally, McGrath, Higgins, and Hintze [21] and Liddle [20] have used information theoretic frameworks to model stylistic novelty over narrative time with respect to small collections of literary documents.

Our work builds on this prior work in at least two important ways. First, we utilize a large and diverse collection of publicly successful long narrative forms [24]. This overcomes limitations surrounding prior work's use of artificially constructed corpora [29] or small historical literary collections [21, 20]. Second, in using an information-theoretic model of narrative revelation, quantifying surprise through similarity of word-count distributions in adjacent windows of text, our models are agnostic with respect to linguistic or thematic content. In contrast, prior work conditioned on topical distributions [30, 33], particular word types [5], or limited semantic frameworks such as sentiment [27, 11]. In this sense, our models approach the question of discourse structure from a more general perspective.

Our reliance on Kullback-Leibler divergence as our principal measure of "information revelation" also brings analytical affordances. Prior work has shown its relevance for understanding a variety of cultural domains (see Chang and DeDeo [9] for an overview), including the study of the novelty of parliamentary discourse [2], the evolution of scientific English [10, 4], human visual attention [16], linguistic processing [18, 19], and implicit cultural learning [34]. Other modeling options such as word embeddings, PCA, or topic modeling require knowledge of the entire text and thus would pollute our measurement of local information novelty relative to a prior window, where the subsequent direction of the text is assumed to be unknown. While transformer models or LLMs could potentially be useful for this task, they run the risk of introducing cultural bias into our models due to the opacity of training data. KLD only measures the particular linguistic shifts within a text bringing in no external information. We take up limitations surrounding the use of KLD to capture the concept of narrative revelation in our discussion section.

Our work is perhaps closest in spirit to that of Thompson, Wojtowicz, and DeDeo [33] and their conceptualization of the explore/exploit paradigm in a narrative setting, although it differs in three key ways. First, their data derive primarily from time-ordered acts of speech (including from parliamentary and social media sources), rather than long narrative forms. Second, they use distributions derived from topic modeling within adjacent windows, while we use word-count distributions when computing Kullback-Leibler divergence for adjacent time windows. Third, whereas they use random-walks based on Levy Flights to model their resulting time-indexed sequences of Kullback-Leibler divergences, with an eye specifically towards capturing narrative (dis)continuity, our focus is on more fundamental properties of time-indexed data like average, trend, and, in particular, dependency structure, for which we use statistical regression and time series analysis.

3. Methods

We define narrative revelation as the practice of disseminating novel information over narrative time with respect to a local prior window of text. Given what has come immediately before, how surprising is any new passage? To capture this concept of surprise, we use Kullback-Leibler divergence (KLD), which calculates the relative entropy (or divergence) between two probability distributions:

$$D_{KL}(p,q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \quad , \tag{1}$$

where X is a (discrete) state space and p and q are probability mass functions defined on X, such that $q(x) = 0$ implies $p(x) = 0$. Note that the quantity in (1) is always greater than or equal to zero, with equality holding if and only if p and q are equal.

For our purposes, the state space is time-varying, with X_T defined as the union of all words in the T-th and $(T-1)$-st adjacent (non-overlapping) windows of 1000 words each. The functions p and q are estimated from the word frequencies in these two respective windows, using Laplace smoothing to avoid 0 values. While conditioning on word frequencies limits the amount of semantic context that can be inferred from a given window of text, it has the advantage of observing the literal distribution of information over narrative time. The end result of this approach is a time series of KLD values, say $\{KLD_T\}_{T=1}^{T_{max}}$, capturing the extent to which information disseminated at time T within a given narrative is novel compared to that disseminated just previously at time $T-1$, with larger KLD corresponding to greater novelty. As such, KLD_T is intended to capture the amount of new information disseminated over narrative time. We refer to these representations as the non-normalized time series. Figure 1 provides examples of this approach and the resulting values and behavior.

In order to test for an association between revelation and narrative time in aggregate, we also create normalized representations of KLD_T for each book to control for differing book lengths. To do so, we first subset all books into 50 equal parts, then subsample 1000 words for each part, and then compute KLD_T for the $T_{max} = 49$ resulting pairs of adjacent windows. We refer to these representations as the normalized time series. As we discuss in Section 4.2 these serve as the basis of our regression analysis to better understand the linear trends of KLD_T at the aggregate level.

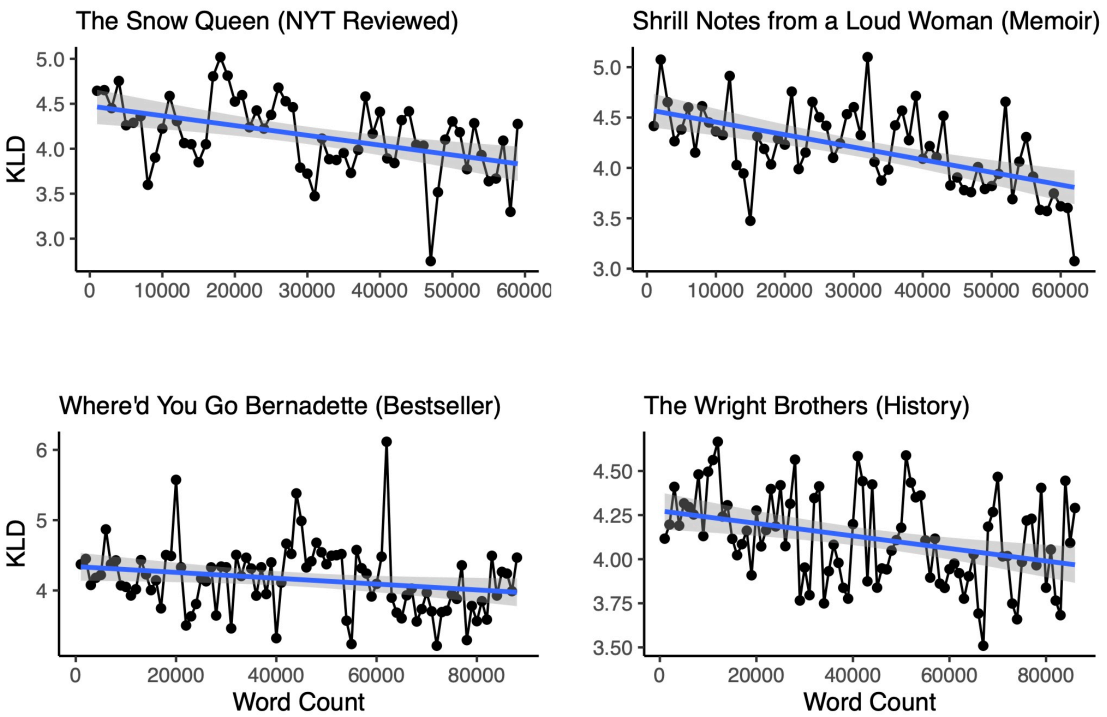

Figure 1: Examples of KLD values over narrative time for four sample books from our corpus from different genres.

Using these approaches, we formulate the following hypotheses:

H1. Average Rate of Revelation. We expect average revelation at the book level to vary by all of our measured social categories. Specifically, prior work [29] has indicated that fictional narratives are more predictable at the sentence level and thus we expect to see lower levels of average revelation with respect to fictionality at the document level. We also expect average revelation to be negatively associated with reading level and positively associated with prestige (more information being more "difficult" for readers to process and thus potentially more valued by elite readerships).

H2. The Slope of Revelation. We expect there to be an association between revelation and narrative time. Prior theoretical work has suggested that narratives exhibit structural patterns [17], which has been confirmed in different ways through empirical work [5, 27, 11]. A general linear increase in surprise would support a theory of narrative investment in the value of plot twists (or “surprise endings”), while a general linear decrease would support the theory of narrative immersion, i.e. once novel information is introduced a narrative spends less time introducing more information (exploration) and more time exploiting known information. While we expect there to be an association between revelation and time, prior work does not give clear indications of which directionality to expect.

H3. Dependency Patterns of Revelation. Given assumptions about the value of narrative structure to narrative meaning, we expect there to be discernible dependency patterns to the rise and fall of revelation, with the present extent of revelation driven by that of the past in non-trivial ways (e.g., lagged dependency). While no prior work has suggested that narrative revelation should follow predictable dependency patterns, it could be the case that this is a latent structural feature to narrative plotting and potentially drives reader enjoyment.

4. Results

4.1. Average Revelation (H1)

We quantify the average revelation by calculating, for each book, the average of the values KLD_T over times T for our non-normalized time series, standardized to account for the considerable differences in book length in the CONLIT data. To evaluate the support in our data for the specific hypotheses with respect to our various two-level factors of social categories, we use two-sample t-tests and report the results in the form of Cohen's d as a measure of effect size (see Table 1). We find that average revelation is associated with all social variables in our data set with the exception of point-of-view. The largest effect size is reserved for the factor of instrumentality: non-fiction books engage in higher rates of average information revelation over narrative time (see also Figure 2). Surprisingly, prestige as captured by prize-winning novels exhibit effects almost as large as instrumentality and greater than those associated with reading level. This supports prior work that has shown significant stylistic differences between prizewinning and bestselling novels [25] and adds a further dimension to understand the ways in which prestige-driven selection effects prioritize distinctive stylistic traits.

Table 1
Measuring the effect size of different factors on average narrative surprise. Cohen's d is used here as a measure of effect size.

Category	d
Fictionality	-0.60 (medium)
Prestige	0.52 (medium)
Age Level	0.25 (small)
POV (1P)	0.14 (negligible)

4.2. The Slope of Revelation (H2)

A regression analysis was conducted to test the association of our narrative revelation variable KLD_T with our narrative time variable T and fictionality, as well as their interaction, here us-

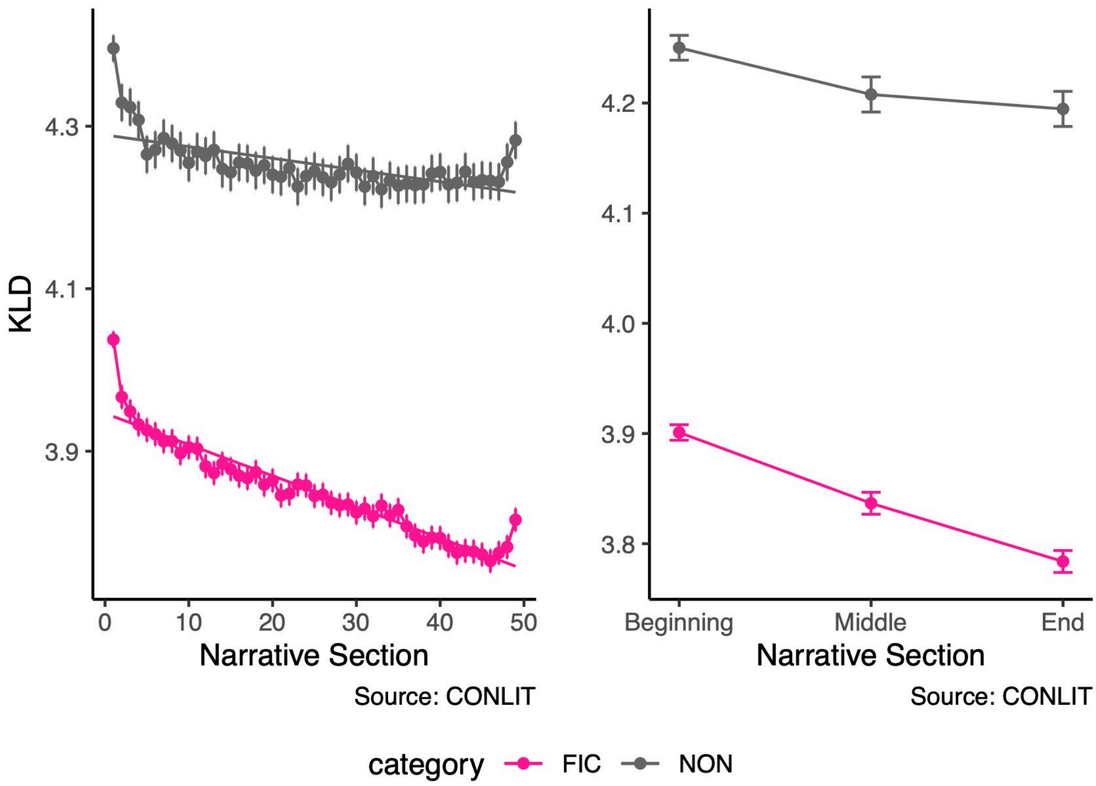

Figure 2: Estimates of average KLD at different points in narrative time using models that standardize book lengths into 50 and 3 parts respectively.

ing the normalized time series. All effects were found to be statistically significant (regression coefficient p-values < 10^{-16}). The association between the amount of narrative revelation and narrative time was small but negative (slope coefficient -0.0038, in comparison to an intercept of 3.947), suggesting that as narratives progress, narrative revelation decreases slightly on average in our entire corpus.

Figure 2 provides the average KLD value for each narrative section by category along with the standard error. For the purposes of visualization, we show our 50-part model as described above as well as a 3-part model, where we divide each book's KLD values for all windows into three equal-sized sections and take the average. As we can see for our 50-part model, fiction books had on average a lower intercept (0.343 lower) and a steeper slope (0.002 steeper) than non-fiction, indicating that fictional books have lower overall levels of narrative revelation (as shown above in 4.1) and also a more pronounced decay in narrative revelation. We also note that for both categories we observe increases in average KLD in the final 1-2 sections of the 50-part model, suggesting that a common approach to narrative closure involves introducing increased levels of novel information toward the end (something we miss in the more generalized 3-part model). Such distinctive structure towards the close of narratives is considerably less pronounced however than the severity of decline of information revelation in the opening sections of a book. Finally, we found that youth fiction was similarly associated with greater decreases of revelation over narrative time, but that prestige and point-of-view were not.

4.3. Dependency Patterns of Revelation (H3)

The results in the previous two sections pertain to the behavior of the average and slope of narrative revelation in aggregate across books in the CONLIT data. Understanding the behavior of revleation at the level of individual books is also of substantial interest but requires a more nuanced analysis. The sequences $\{KLD_T\}_{T=2}^{T_{max}}$ are time series, not only of varying lengths but also, as it turns out, of varying complexity.

Exploratory analysis of the non-normalized KLD time series reveals that, while they in general oscillate, they nevertheless do not typically have a dominant frequency (as determined using the `findfrequency` function of the R `Forecast` package), suggesting the absence of strictly periodic (and hence easily predicted) behavior of narrative revelation over narrative time. Further exploration of the autocorrelation behavior of the KLD time series suggests the use of ARIMA models. Such models are the workhorse of modern time series analysis and consist of three components: autoregressive (AR), integrated (I), and moving average (MA). The autoregressive component refers to behavior where the value KLD_T at time T can be predicted by earlier values $KLD_{T-1}, \dots, KLD_{T-p}$, for some p lags, suggesting a regression-based relationship with itself. The moving average component allows for this regression-based relationship to have dependent errors, say over time scales of length q. And the integrated component allows for such combined AR-MA behavior to ride on top of a polynomial trend of order d, akin to the way a line with slope underlies a cloud of points within classical linear regression.

We used the `auto.arima` function in the R `Forecast` package to fit a separate ARIMA model to each KLD time series in the CONLIT data. This function includes data-driven selection of the triple (p, d, q), which we take as the unit of primary interest in our analysis. Of the 2754 books analysed, 59% exhibited a trend ($d > 0$). Of those, 80% exhibited downward trends (i.e. negative slopes). Non-fiction books were 2x more likely to be in the positive slope class. For our second variable, 39% of all books exhibited autoregressive behavior ($p > 0$), meaning that in a strong minority of books the successive values of narrative revelation are correlated. Within this group we see that 75% have first order dependencies (p=1) and another 20% have second-order (p=2), accounting for almost all books with auto-regressive behavior. Where there is a correlation between successive windows, it tends to reach only 1-2 windows back. Finally, $q = 0$ was selected for all books in the data set, indicating that these characteristics of autoregression and/or trend can be viewed as occurring with a backdrop of white noise.

As in Section 4.1, we test the distribution of our two variables of interest, trend (d) and dependence (p) across our four social categories (Table 2). We report the percentage of books associated with each kind of time-dependent behavior for each category. As we can see from the breakdowns, there are only two scenarios where we observe meaningful differences between categories (> 5% difference among books). The first is at the level of dependence for Youth books, where we see 7% fewer books exhibiting auto-regressive behavior. This suggests that books targeting younger audiences skew in favor of less patterning and more consistency when it comes to narrative revelation. This is underscored by the fact that this effect is even stronger for Middle School books compared to Young Adult books. The second notable difference is similar to what we observed in Figure 2. Fiction books are more likely to exhibit a detectable trend in the levels of revelation over narrative time, a trend which is overwhelmingly negative (downward).

Table 2
Comparing dependence (p) and trend (d) across our four social categories. We report the fraction of books exhibiting a positive value for each class.

Category	Classes	Dependence (p)	Trend (d)
Fictionality	Fiction / Nonfiction	0.38 / 0.41	0.61 / 0.56
Prestige	Prizewinners / Bestsellers	0.44 / 0.47	0.59 / 0.61
Age Level	Youth / Adult	0.34 / 0.41	0.60 / 0.62
POV	1P / 3P	0.38 / 0.39	0.63 / 0.59

5. Discussion

Our work has aimed to continue prior efforts in modeling the temporal dimensions of narrative communication. Narratives have a fundamental temporal dimension that impacts their meaning. Accordingly, we have provided a novel method for capturing the dissemination of new information over narrative time as well as highlighted the utility of well-established statistical methods for capturing temporal relationships in time-series data. Our hope is that these frameworks can be applied towards the further study of computational narrative understanding to deepen our knowledge about the typicalities and particularities of human storytelling.

Our models support prior work [29, 26] in showing how fictional narratives exhibit significantly more investment in patterns of narrative exploitation than narrative exploration. Fiction tends to engage in lower levels of narrative revelation overall and those levels decline more precipitously over narrative time. Fictional narratives invest more heavily in immersing readers in well-known information rather than continuously introducing novel information, an effect that grows stronger over the course of a book's narrative. Both fiction and non-fiction exhibit a tendency to increase narrative revelation in the final closing sections of a book, suggesting a more universal narrative tendency with regards to narrative structure [5].

When it comes to books targeting different reading audiences, we see that the intended age level of audiences and the selection preferences of elite audiences do appear to effect levels of narrative revelation (and for younger audiences lower levels of temporal dependence). Authors engage in lower overall levels of revelation when writing for younger audiences, and higher levels when attempting to appeal to elite audiences.

One major open question for this line of research is the degree to which KLD covers the diverse ways that "narrative revelation" may instantiate itself. Changes in vocabulary distribution over narrative time that our models capture is one way of thinking about novel information in a narrative. But we can also imagine how new or surprising information could be encoded in very similar language but provides a key as yet unknown insight. The revelation of a murderer in a mystery is the most obvious example where a single name would provide very high levels of “revelation,” but low levels of KLD. The fact that our measures are inversely associated with audience reading levels suggest that narrative revelation as we are modeling it may be capturing the information *load* and/or narrative complexity as much as the potential cognitive disposition of “surprise” on the part of readers. On the other hand, the trend towards increases in late-section rises of KLD that we are seeing suggests that our models may be capturing this idea of narrative revelation as a function of novel information as it relates to key plot points.

Similarly, because our models condition on local revelation – where the amount of novel information is measured with respect to an immediate prior window – we cannot know if such late-stage increases in revelation are absolutely novel or a return to information that references earlier parts of the narrative (performing a sense of narrative “closure”). Future models could explore the extent of revelation with respect to larger windows of text or even the entire text, in essence capturing the absolute novelty of any new passage with respect to what a book has divulged up to that point. Such work would also open the door to questions of non-linearity, as when a passage refers back to a distant prior passage and continues the narrative after some interlude [26]. We thus see a key avenue for future work to focus on validating and making precise what aspects of narrative revelation KLD captures and what other kinds and structures of revelation over narrative time are possible.

Acknowledgments

This research was generously supported by the Social Sciences and Humanities Research Council of Canada (895-2013-1011).

References

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

[2] A. T. Barron, J. Huang, R. L. Spang, and S. DeDeo. “Individuals, Institutions, and Innovation in the Debates of the French Revolution”. In: *Proceedings of the National Academy of Sciences* 115.18 (2018), pp. 4607–4612.

[3] J. Bermejo-Berros, J. Lopez-Diez, and M. A. G. Martínez. “Inducing Narrative Tension in the Viewer through Suspense, Surprise, and Curiosity”. In: *Poetics* 93 (2022), p. 101664.

[4] Y. Bizzoni, S. Degaetano-Ortlieb, P. Fankhauser, and E. Teich. “Linguistic Variation and Change in 250 years of English Scientific Writing: A Data-Driven Approach”. In: *Frontiers in Artificial Intelligence* 3 (2020), p. 73.

[5] 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.

[6] W. F. Brewer and E. H. Lichtenstein. “Stories are to Entertain: A Structural-Affect Theory of Stories”. In: *Journal of Pragmatics* 6.5-6 (1982), pp. 473–486.

[7] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer science & business media, 2009.

[8] I. Calvino. *Six Memos for the Next Millennium.* Harvard University Press, 1988.

[9] K. K. Chang and S. DeDeo. "Divergence and the Complexity of Difference in Text and Culture". In: *Journal of Cultural Analytics* 5.2 (2020).

[10] S. Degaetano-Ortlieb, H. Kermes, A. Khamis, and E. Teich. "An Information-Theoretic Approach to Modeling Diachronic Change in Scientific English". In: *From Data to Evidence in English Language Research.* Brill, 2018, pp. 258–281.

[11] K. Elkins. *The Shapes of Stories: Sentiment Analysis for Narrative.* Cambridge University Press, 2022.

[12] G. Freytag. *Technique of the Drama: An Exposition of Dramatic Composition and Art.* S. Griggs, 1895.

[13] G. Genette. *Narrative Discourse: An Essay in Method.* Vol. 3. Cornell University Press, 1983.

[14] G. Genette. *Paratexts: Thresholds of Interpretation.* 20. Cambridge University Press, 1997.

[15] D. Herman. *Basic Elements of Narrative.* John Wiley & Sons, 2009.

[16] L. Itti and P. Baldi. "Bayesian Surprise Attracts Human Attention". In: *Vision Research* 49.10 (2009), pp. 1295–1306.

[17] W. Labov and J. Waletzky. "Narrative Analysis: Oral Versions of Personal Experience." In: (1967).

[18] R. Levy. "Expectation-Based Syntactic Comprehension". In: *Cognition* 106.3 (2008), pp. 1126–1177.

[19] R. Levy. "Memory and Surprisal in Human Sentence Comprehension". In: *Sentence Processing* 78 (2013), pp. 142–195.

[20] D. Liddle. "Could Fiction Have an Information History? Statistical Probability and the Rise of the Novel". In: *Journal of Cultural Analytics* 4.2 (2019).

[21] L. McGrath, D. Higgins, and A. Hintze. "Measuring Modernist Novelty". In: *Journal of Cultural Analytics* 3.1 (2018).

[22] 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.

[23] A. Piper. "Novel Devotions: Conversional Reading, Computational Modeling, and the Modern Novel". In: *New Literary History* 46.1 (2015), pp. 63–98.

[24] A. Piper. "The CONLIT Dataset of Contemporary Literature". In: *Journal of Open Humanities Data* 8 (2022).

[25] A. Piper and E. Portelance. "How Cultural Capital Works: Prizewinning Novels, Bestsellers, and the Time of Reading". In: *Post45* 10 (2016).

[26] A. Piper and O. Toubia. "A Quantitative Study of Non-linearity in Storytelling". In: *Poetics* 98 (2023), p. 101793.

[27] A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, and P. S. Dodds. "The Emotional Arcs of Stories are Dominated by Six Basic Shapes". In: *EPJ Data Science* 5.1 (2016), pp. 1–12.

[28] P. Ricoeur. *Time and Narrative, Volume 1*. University of Chicago Press, 2012.

[29] M. Sap, A. Jafarpour, Y. Choi, N. A. Smith, J. W. Pennebaker, and E. Horvitz. "Quantifying the Narrative Flow of Imagined versus Autobiographical Stories". In: *Proceedings of the National Academy of Sciences* 119.45 (2022), e2211715119.

[30] B. M. Schmidt. "Plot Arceology: A Vector-Space Model of Narrative Structure". In: *2015 IEEE International Conference on Big Data (Big Data)*. Ieee. 2015, pp. 1667–1672.

[31] M. Sternberg. "Telling in Time (I): Chronology and Narrative Theory". In: *Poetics Today* 11.4 (1990), pp. 901–948.

[32] M. Sternberg. "Telling in Time (II): Chronology, Teleology, Narrativity". In: *Poetics Today* 13.3 (1992), pp. 463–541.

[33] W. H. Thompson, Z. Wojtowicz, and S. DeDeo. "Lévy Flights of the Collective Imagination". In: *arXiv preprint arXiv:1812.04013* (2018).

[34] S. P. Veissière, A. Constant, M. J. Ramstead, K. J. Friston, and L. J. Kirmayer. "Thinking Through Other Minds: A Variational Approach to Cognition and Culture". In: *Behavioral and Brain Sciences* 43 (2020), e90.