Computational Prediction of Elapsed Narrative Time

Gregory Yauney
Cornell University
gyauney@cs.cornell.edu

Ted Underwood
University of Illinois,
Urbana-Champaign

David Mimno
Cornell University

Abstract

It is commonly asserted that the average amount of narrative time that elapses in a passage of fiction has decreased over the past several hundred years. But did this change in literary time really occur, and if so, how rapidly? Previous work on a small but carefully chosen set of hand-labeled passages finds that literary time becomes progressively shorter over the 19^{th} and early 20^{th} centuries. We explore computational methods for predicting literary time and apply these to a collection of more than 53,000 complete novels from the 1700s to the present day. We observe the same decrease in average narrative time up to the mid-20^{th} century. After establishing this result we compare metrics of estimated literary time to patterns of language use.

Defoe, *Robinson Crusoe* (long literary time)
But first I was to prepare more land, for I had now seed enough to sow above an acre of ground. Before I did this, I had a week's work at least to make me a spade, which, when it was done, was but a sorry one indeed, and very heavy, and required double labour to work with it.
Gaddis, *The Recognitions* (short literary time)
I'd rather not talk about it. But all right, I'm sorry, I didn't mean . . . He had started to move away from her but Esther was speaking to him, her voice going on as though she had not stopped, Because you've done the same thing, you've spent all your time too, you've put all your energy up against things that weren't there, but you put them there yourself just to have something to fight. . .

Table 1: Excerpts from passages labeled as having long and short elapsed narrative time.

1 Introduction

A goal of the study of the history of literature is to identify consistent changes in the writing of fiction. One of the more measurable characteristics is the narrative time elapsed in a passage. Table 1 shows example passages with long and short durations of elapsed narrative time. Narrative time by itself may not be particularly meaningful in isolation but if we can find consistent trends they may indicate larger changes in literary history.

We build on work that defines and measures elapsed narrative time in novels from the early 18^{th} through late 20^{th} centuries (Underwood, 2018). Underwood finds a consistent trend towards shorter elapsed time over the course of the 18^{th} and 19^{th} centuries from on average about one day down to about one hour, with an increase around the middle of the 20^{th} century. The previous work suggests that the trend in elapsed narrative time is not subject to the discontinuities previously hypothesized by some literary theorists (Genette, 1980), but the work has a small, carefully chosen dataset and large variation.

We present a new extension of the existing labeled dataset, which we use to validate previous findings. We then use a simple baseline method, linear regression from unigrams, to predict the log of elapsed time per passage length. In addition to the simple baseline method, we try a more sophisticated ranking-based approach that improves predictions at the loss of temporal resolution.

After validating our predictive models on labeled data we extend these results to an unlabeled dataset that is four orders of magnitude larger by token count. We replicate the result from (Underwood, 2018) that there is a strong gradual trend towards shorter narrative time over the course of the 19^{th} and 20^{th} centuries that flattens around 1950.

Measuring trends in the narrative presentation of elapsed time is significant in that it can lend additional credence to historical claims previously based on near-anecdotal evidence. Such similar gradual downward trends between the small labeled dataset and the large dataset serve to increase confidence in the gradual decrease in elapsed narrative time rather than the previously hypothesized discontinuity (Underwood, 2018). We interrogate the diachronic trends in narrative time using a learned topic model and preliminarily find that the trends are not due to a simple identification of century-specific style.

2 Related work

Much work has been done on computationally-assisted narrative interpretation, especially on such large-scale categorizations as genre, style, and nationality, as well as on quantifying literary-historical phenomena over time (Jockers, 2013; Piper, 2018). We focus on finer-grained textural information, as in work that quantifies intratextual qualities (Underwood, 2018; McGrath et al., 2018).

3 Datasets

We use three datasets of varying sizes: two small labeled datasets for learning to predict elapsed narrative time and to quantify the quality of that prediction. The third, much larger, dataset is for evaluating the extent to which the resulting literary historical claims about elapsed narrative time are justified on a larger corpus.

3.1 Existing labeled dataset

As a starting point, we begin with a dataset collected and analyzed by Underwood, which we refer to as the WLTMM dataset (Underwood, 2018). This dataset consists of 1,765 passages of about 250 words in length chosen from 108 canonical, best-selling, and randomly chosen novels.

3.2 Newly labeled dataset

In addition to the existing labeled passages, we created an additional set of labeled passages. These will be released publicly to the extent possible under copyright. In order to validate similarity within authors we label new passages from novels in the existing WLTMM dataset. To evaluate performance across authors, we also include a number of new, previously unlabeled novels.

We labeled elapsed narrative time for eight passages from each of seven novels for a total of 56 newly labeled passages. The novels were a combination of those already in the WLTMM dataset and new novels from across the time period covered by the WLTMM dataset: *Robinson Crusoe* (1719), *Pride and Prejudice* (1813), *Frankenstein* (1818), *Jane Eyre* (1847), *Middlemarch* (1871-2), *Mrs Dalloway* (1925), and *The Recognitions* (1955). Passages were 250-word segments chosen at random from the selected novels. No individual passages were repeated between this and the WLTMM dataset.

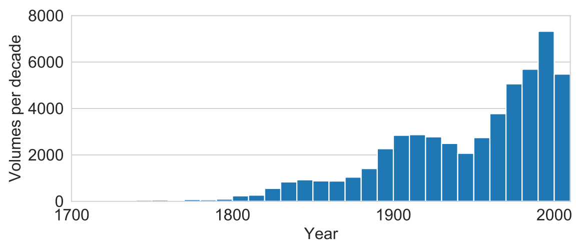


Figure 1: Estimated years of first publication in the sample of 53,539 HathiTrust volumes.

3.3 HathiTrust

In order to extend our analysis to a large-scale data set including contemporary works we need access to passage-level word count information from volumes that are in copyright. The HathiTrust Digital Library[1] provides, among other things, page-level word frequencies of millions of volumes of text. We sampled 116,434 volumes from HathiTrust that (a) had a genre metadata tag of fiction and (b) were published in English. We constructed bags of words for each volume. The vocabulary was limited for computational efficiency to the top 10,000 words in a sample of 1,000 HathiTrust volumes.

The year of publication field in the HathiTrust metadata is an insufficient proxy for year of first publication (Bamman et al., 2017). In order to estimate year of first publication, if a volume was published after the year of the author's death, we estimated the year of first publication as the year of the author's birth plus thirty years. We found this metadata-based strategy provided a reasonable estimation for our purposes, so we did not predict first publication from the text data as in (Bamman et al., 2017). Particular examples of failure were for novels first published posthumously or later editions published while an author was alive. Figure 1 shows the distribution of estimated years of first publication in our sample of the HathiTrust dataset.

[1] https://www.hathitrust.org/

Novel	Labeled	Predicted	Passage
The Recognitions Gaddis (1955)	-3.9120 (5 minutes)	-3.7238 (6 minutes)	But, and then, you're all right? I'd rather not talk about it. But all right, I'm sorry, I didn't mean . . . He had started to move away from her but Esther was speaking to him, her voice going on as though she had not stopped, Because you've done the same thing, you've spent all your time too, you've put all your energy up against things that weren't there, but you put them there yourself just to have something to fight . . . Esther ... So you wouldn't have to fight the real things. She spoke with great rapidity at him. And now you say you're tired? At your age, because you've been trying to make negative things do the work of positive ones . . . I wish I was an old man! he burst out at her, and then lowered his eyes again, his pale hand inside his coat holding the thick packet there. Because . . . damn it, this being young, it's like he said it was, it's like a tomb, this youth, youth, this thing in America, this accent on youth, on everything belongs to the young, and we, look at us, in this tomb, like he told me it could be, like he said it was . . . And Otto raised his eyes to see nothing moving in her face. Yes, you came here for him, didn't you, she said quietly. You only wanted to see him, didn't you. And
Frankenstein Shelley (1818)	1.7509 (1 day)	3.5959 (9 days)	narrower as I approached my native town. I discovered more distinctly the black sides of Jura, and the bright summit of Mont Blanc. I wept like a child. Dear mountains! my own beautiful lake! how do you welcome your wanderer? Your summits are clear; the sky and lake are blue and placid. Is this to prognosticate peace, or to mock at my unhappiness? I fear, my friend, that I shall render myself tedious by dwelling on these preliminary circumstances; but they were days of comparative happiness, and I think of them with pleasure. My country, my beloved country! who but a native can tell the delight I took in again beholding thy streams, thy mountains, and, more than all, thy lovely lake! Yet, as I drew nearer home, grief and fear again overcame me. Night also closed around; and when I could hardly see the dark mountains, I felt still more gloomily. The picture appeared a vast and dim scene of evil, and I foresaw obscurely that I was destined to become the most wretched of human beings. Alas! I prophesied truly, and failed only in one single circumstance, that in all the misery I imagined and dreaded, I did not conceive the hundredth part of the anguish I was destined to endure. It was completely dark when I arrived in the environs of Geneva; the gates of the town were already shut; and I was obliged to pass the night at Secheron, a village at the distance of half
Pride and Prejudice Austen (1813)	5.0831 (1 month)	-3.0814 (11 minutes)	farther northwards than Derbyshire. In that county there was enough to be seen to occupy the chief of their three weeks; and to Mrs. Gardiner it had a peculiarly strong attraction. The town where she had formerly passed some years of her life, and where they were now to spend a few days, was probably as great an object of her curiosity as all the celebrated beauties of Matlock, Chatsworth, Dovedale, or the Peak. Elizabeth was excessively disappointed; she had set her heart on seeing the Lakes, and still thought there might have been time enough. But it was her business to be satisfied–and certainly her temper to be happy; and all was soon right again. With the mention of Derbyshire there were many ideas connected. It was impossible for her to see the word without thinking of Pemberley and its owner. But surely," said she, "I may enter his county with impunity, and rob it of a few petrified spars without his perceiving me." The period of expectation was now doubled. Four weeks were to pass away before her uncle and aunt's arrival. But they did pass away, and Mr. and Mrs. Gardiner, with their four children, did at length appear at Longbourn. The children, two girls of six and eight years old, and two younger boys, were to be left under the particular care of their cousin Jane, who was the general favourite, and whose steady sense and sweetness of temper exactly adapted her for attending to them
Jane Eyre Brontë (1847)	-2.1203 (30 minutes)	1.2455 (14 hours)	dragged my exhausted limbs slowly towards it. It led me aslant over the hill, through a wide bog, which would have been impassable in winter, and was splashy and shaking even now, in the height of summer. Here I fell twice; but as often I rose and rallied my faculties. This light was my forlorn hope: I must gain it. Having crossed the marsh, I saw a trace of white over the moor. I approached it; it was a road or a track: it led straight up to the light, which now beamed from a sort of knoll, amidst a clump of trees–firs, apparently, from what I could distinguish of the character of their forms and foliage through the gloom. My star vanished as I drew near: some obstacle had intervened between me and it. I put out my hand to feel the dark mass before me: I discriminated the rough stones of a low wall–above it, something like palisades, and within, a high and prickly hedge. I groped on. Again a whitish object gleamed before me: it was a gate–a wicket; it moved on its hinges as I touched it. On each side stood a sable bush-holly or yew. Entering the gate and passing the shrubs, the silhouette of a house rose to view, black, low, and rather long; but the guiding light shone nowhere. All was obscurity. Were the inmates retired to rest? I feared it must be so. In seeking the door, I turned an angle: there

Table 2: Representative example passages with labeled elapsed time and predicted elapsed time. The top two passages show correct predictions for short duration and long duration. The bottom two passages show examples of extreme mismatches in prediction. Bright green: ten words with the greatest combined contribution in each passage indicative of long elapsed time. Dark green: next ten words in each passage indicative of long elapsed time. Bright purple: ten words with the greatest combined contribution in each passage indicative of short elapsed time. Dark purple: next ten words in each passage indicative of short elapsed time.

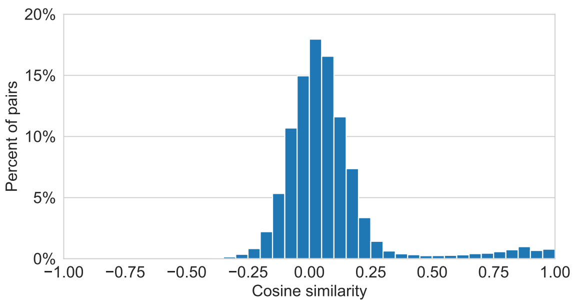


Figure 2: Distribution of cosine similarity after low-rank random projection between all pairs of 700 labeled HathiTrust volumes.

Duration	$\log(\frac{\text{minutes}}{\text{words}})$
a year	7.651
a month	5.152
a week	3.697
a day	1.751
an hour	-1.427
a minute	-5.521

Table 3: Time durations converted to log of minutes per words for 250-word passages.

3.4 Volume deduplication

Individual works appear in HathiTrust many times over. Reducing the number of duplicates is important to avoid the distorting effects that such duplicates can have, especially on semantic models (Schofield et al., 2017).

We used random low-rank projection of each volume from a vector with 10,000 entries weighted by inverse document frequency to a vector with 100 entries, and, after normalization, compared cosine similarities between all volumes. Only one volume was kept from any set of volumes with pairwise cosine similarities above a certain threshold.

The cutoff threshold was determined by running this process on a smaller labeled dataset of 700 volumes corresponding to 62 novels that were also in the WLTMM dataset. Figure 2 shows, via a histogram of pairwise cosine similarities, that the vast majority of such similarities were within the range of random while a small but noticeable fraction were very similar, as expected. We used the relatively conservative threshold of 0.75 that removed $87.5\% \pm 30.6\%$ of each novel's duplicates on average while only removing eight volumes that were not duplicates from two novels each.

4 Predicting elapsed narrative time

We learned to both directly predict elapsed narrative time using linear regression and to rank two passages' elapsed times via logistic regression. Furthermore, we used a latent Dirichlet allocation topic model to interpret the output of the linear regression model (Blei et al., 2003).

4.1 Linear regression

Following the work of Underwood, we trained and tested a linear regression model on the WLTMM dataset. The input to the model was the unigram

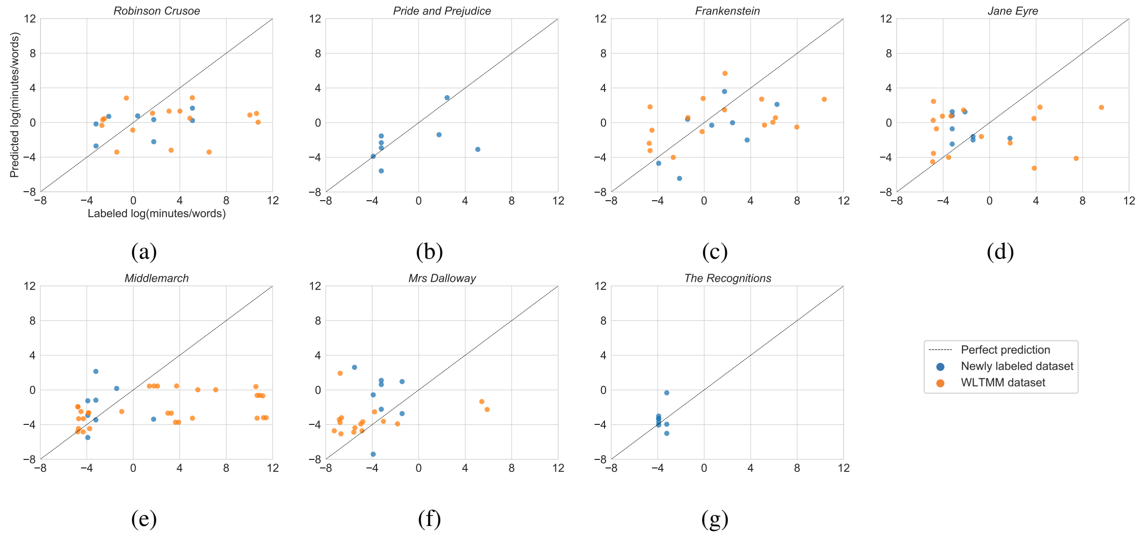


Figure 3: Correlation between labeled and linear-regression-predicted elapsed narrative time for the newly labeled passages (blue) with the results for passages in the WLTMM dataset (orange) included for comparison. The diagonal lines indicate what would be perfect prediction. Points above the line correspond to passages with predictions that over-predict elapsed time, i.e. those with a predicted elapsed time greater than that which was labeled, and points under the line similarly correspond to under-predictions.

Word	Coefficient
years	0.7331
life	0.6059
return	0.5924
family	0.5065
failure	0.5043
history	0.4867
wrote	0.4772
during	0.4694
nor	0.4337
aunt	0.4309

(a)

Word	Coefficient
woman	-0.3986
black	-0.4176
best	-0.4193
pocket	-0.4215
room	-0.4243
thou	-0.4315
struck	-0.4330
wish	-0.4536
moment	-0.4935
replied	-0.6549

(b)

Table 4: Coefficients from the linear regression for (a) words with coefficients most associated with longer elapsed time and (b) words with coefficients most associated with shorter elapsed time.

Word	Frequency	Coefficient	Total contribution
you	13	0.12	1.52
ve	4	0.31	1.25
to	9	0.08	0.68
and	6	0.08	0.48
eyes	2	0.23	0.46
		⋮	
her	4	-0.12	-0.50
have	2	-0.30	-0.60
him	4	-0.18	-0.71
but	4	-0.24	-0.95
thing	2	-0.49	-0.99

Table 5: Words that contribute most positively (top) and most negatively (bottom) to the prediction of elapsed time for the passage from *The Recognitions* shown at the top of Table 2.

frequencies from a passage, and output was the log of the number of minutes per words in the passage. Table 3 shows time durations with equivalent log values, for reference. The regression was trained on 1,487 passages from 90 novels, with a Pearson correlation coefficient of R=0.6095 between labeled and predicted narrative time on held-out passages.

We further validated the trained model on our newly labeled dataset. Table 2 shows representative examples of passages for which the classifier performed well and examples for which it performed poorly. Figure 3 shows regression predictions for each passage in our newly labeled dataset with comparisons to predictions on passages from the same novels in the WLTMM dataset, where available. The Pearson correlation coefficient between labeled and predicted elapsed time for all passages in our newly labeled dataset was R=0.3474 (p=0.0087). The regression performed best on passages from *Frankenstein*, *Pride and Prejudice*, and *The Recognitions*, despite no passages from the latter two appearing in the training set.

Table 4 shows the words with regression coefficients most indicative of long and short elapsed narrative time. Some words, examples of which include *you*, *to*, and *and* in Table 5, had coefficients

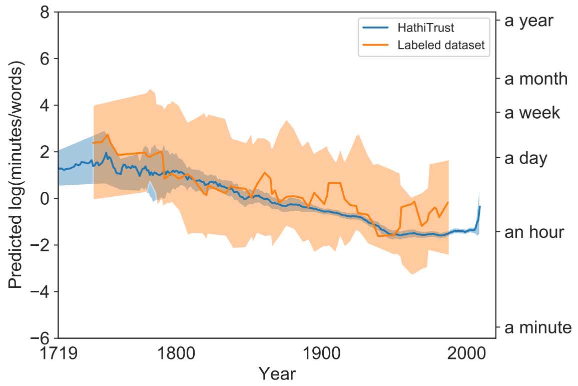
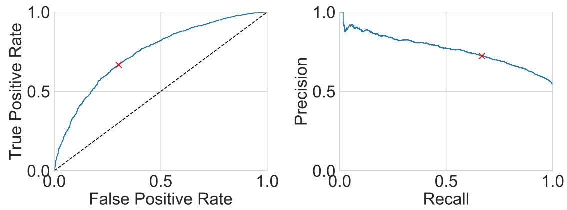


Figure 4: Moving means of average predicted elapsed narrative time per passage in the labeled training dataset (orange) and the larger HathiTrust dataset (blue). Shaded regions represent 95% confidence intervals, which do not account for prediction uncertainty.



(a) (b)

Figure 5: Receiver operating characteristic and precision-recall curves for the ranking classifier run on the test set of newly labeled passages. The training and test sets were identical to those used for linear regression. The 'x' identifies this classifier's performance.

much closer to zero but contributed greatly to a passage's predicted time due to their high frequency.

We ran the trained classifier on all pages in our sampled HathiTrust dataset. Figure 4 shows that predicted elapsed narrative time per page on the larger unlabeled HathiTrust dataset roughly parallels that of the predicted times for the WLTMM dataset when predictions are averaged by the novels' years of first publication. The average for passages in HathiTrust has a much narrower 95% confidence interval. This gives additional evidence across passages from full novels to the trend toward shorter elapsed narrative time throughout the 19^{th} and into the 20^{th} centuries identified in (Underwood, 2018).

Linear regression is limited in both its treatment of input features and output estimates. It is not able to account for feature conjunctions as every additional instance of a word has the same linear effect on the prediction regardless of context. Word sequence information and syntactic parses could provide more information, but with our current unigram-level access to material under copyright we would not be able to extend our analysis beyond the early 1920s. More sophisticated multi-layer architectures might be able to find useful conjunctions of features or avoid overweighting frequent terms. Another limitation of linear regression is that it is purely additive, so that there are no built-in constraints on the range of outputs, even to the point where predictions become physically implausible. As we are making predictions in the form of the log of elapsed time, the addition of a few words as input could result in predictions that might be narratively impossible, such as a duration of milliseconds or decades.

4.2 Ranking is an easier task than regression

In order to address some of the issues of the linear regression we trained a ranking classifier by performing logistic regression on the difference between two passages' bags of words. The goal was to see if determining which of the two input passages had a longer elapsed narrative time could be predicted with greater accuracy than directly predicting the log of the elapsed time. Specifically, we present the classifier with two passages and code a "correct" prediction as 0 if the first passage has a shorter duration and as 1 if the second passage has a shorter duration. The training data consisted 100,000 pairwise comparisons for random pairs of passages in the WLTMM dataset, and test data was pairwise comparisons between all pairs of passages in our newly labeled dataset.

Figure 5 shows receiver operating characteristic (ROC) and precision-recall curves for the test set; the classifier had an area under the ROC curve (AUC) of 0.7448, precision of 0.72, and recall of 0.67. Logistic regression outperforms the baseline of ranking passages by the linear regression's predicted elapsed time, which achieved an AUC of 0.6517, with precision and recall of 0.69 and 0.64, respectively. The words that are most associated with longer and shorter elapsed time remain mostly unchanged from those in linear regression. There do not seem to be specific novels on which the classifier performs better or more poorly.

4.3 ℓ_1 regularization for word importance

We find that although there is room for improvement, a number of different approaches for predicting elapsed time computationally have similar empirical performance. We next consider the limits of the set of input features. How few word types can we use without seriously impacting performance?

We use an ℓ_1-regularized ranking model with varying penalty parameter to progressively drive feature weights to zero. Surprisingly, a minimal set of eight words is sufficient to obtain 0.74 AUC on the training and test sets without any degradation in test performance. These words are *of, in, to, had* to indicate longer duration and *you, said, it, he* to indicate shorter duration.

While it is difficult to extrapolate from such words, all are short words that often appear in varieties of phrases. *Said* is additionally very indicative of dialogue, as—to a lesser extent—is *you* indicative of second-person dialogue as well.

5 Learned topics correlate with elapsed time

It is interesting to be able to predict human labels for elapsed time, but this by itself is only the beginning of the process of making empirical literary historical claims. We hope to use the larger corpus to reason about other trends that might be associated with such prediction. One type of trend is whether or not shorter and longer elapsed narrative time is grounded in specific language in the form of word co-occurrence patterns.

To this end, a topic model with 50 topics was trained on the HathiTrust volumes, with words that were found to be most predictive of elapsed narrative time excluded. Documents' topic proportions were used in a linear regression to predict a document's average elapsed narrative time.

To avoid simply learning from exactly the same data, the documents on which the topic model was trained had the eight words most indicative of elapsed time removed. The target value for the regression was the elapsed time predicted from the linear regression model on just the eight words excluded from the topic model. In the regression, 70% of the HathiTrust volumes were used for training and the remaining 30% were used for testing. Testing data had a Pearson correlation coefficient of R=0.6226 between average elapsed narrative time from the regression on unigram frequencies and predicted elapsed narrative time from this regression on topic proportions.

Figure 6 shows the topics most correlated with longer and shorter elapsed narrative time. Some topics seem to capture stylistic changes over time,

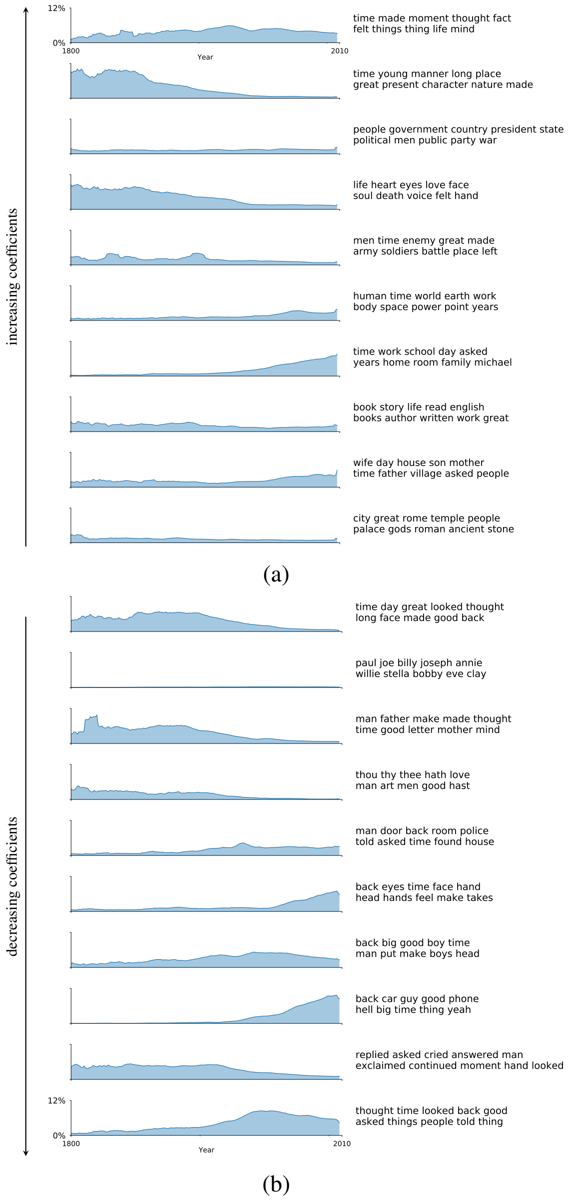


Figure 6: Prevalence of topics given publication date for (a) the ten topics that are most associated with longer elapsed narrative time and (b) the ten topics that are most associated with shorter elapsed narrative time ($K = 50$). Both sets mix older and more recent topics.

such as the topic with the third-strongest association with shorter elapsed time, which contains contemporary vernacular words. Notably, however, not all of the topics simply correlate with publication date. Some appear to be getting at other features of narrative, such as dialogue (*replied, asked, cried*), or content, such as government (*people, government*) and Rome (*city, great, rome*). Just as *said* is one of the words most associated with shorter elapsed time when ranking with ℓ_1-regularization, so, too, is a topic consisting of words that mark dialogue associated with shorter elapsed time here.

Though the distribution of topics does not preclude the possibility that the models are also learning other correlates of publication date, they do show that there are word co-occurrence patterns that distinctly correlate with longer and shorter elapsed time while not simply reproducing the gradual downward trend in narrative duration over time seen in Figure 4.

6 Discussion

Additional work can be done with our existing formulation of narrative time to detect narrative time with greater accuracy, to investigate whether our literary historical claims extend to even larger corpora, and to investigate whether other aspects of literary texture contribute to elapsed narrative time. The concept of literary time could be formulated otherwise; for instance, other formulations might detect phrases that signal the rapid passage of time, such as "The next morning . . . ". We have presented models that learn to predict elapsed narrative time and have furthermore used them to evaluate empirical historical claims about the shortening of literary time, finding that such claims remain valid even with the increased evidence afforded by narrative time prediction.

References

David Bamman, Michelle Carney, Jon Gillick, Cody Hennesy, and Vijitha Sridhar. 2017. Estimating the date of first publication in a large-scale digital library. In *Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries*, pages 149–158. IEEE Press.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Gérard Genette. 1980. *Narrative Discourse: An Essay on Method*. Cornell University Press, Ithaca.

Matthew L Jockers. 2013. *Macroanalysis: Digital methods and literary history*. University of Illinois Press.

Laura McGrath, Devin Higgins Higgins, and Arend Hintze. 2018. Measuring modernist novelty. *Journal of Cultural Analytics*.

Andrew Piper. 2018. *Enumerations: Data and Literary Study*. The University of Chicago Press.

Alexandra Schofield, Laure Thompson, and David Mimno. 2017. Quantifying the effects of text duplication on semantic models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2737–2747.

Ted Underwood. 2018. Why literary time is measured in minutes. *ELH*, 85(2):341–365.