

## University of Groningen

### Stylometric Literariness Classification

van Cranenburgh, Andreas; Ketzan, Erik

*Published in:*

Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2021

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

van Cranenburgh, A., & Ketzan, E. (2021). Stylometric Literariness Classification: the Case of Stephen King. In S. Degaetano-Ortlieb, A. Kazantseva, N. Reiter, & S. Szpakowicz (Eds.), *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (pp. 189-197). Association for Computational Linguistics (ACL).

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Stylometric Literariness Classification: the Case of Stephen King

Andreas van Cranenburgh

University of Groningen  
Department of Information Science  
a.w.van.cranenburgh@rug.nl

Erik Ketzan

Trinity College Dublin  
Centre for Digital Humanities  
ketzane@tcd.ie

## Abstract

This paper applies stylometry to quantify the literariness of 73 novels and novellas by American author Stephen King, chosen as an extraordinary case of a writer who has been dubbed both “high” and “low” in literariness in critical reception. We operationalize literariness using a measure of stylistic distance (Cosine Delta) based on the 1000 most frequent words in two bespoke comparison corpora used as proxies for literariness: one of popular genre fiction, another of National Book Award-winning authors. We report that a supervised model is highly effective in distinguishing the two categories, with 94.6% macro average in a binary classification.

We define two subsets of texts by King—“high” and “low” literariness works as suggested by critics and ourselves—and find that a predictive model does identify King’s *Dark Tower* series and novels such as *Dolores Claiborne* as among his most “literary” texts, consistent with critical reception, which has also ascribed postmodern qualities to the *Dark Tower* novels. Our results demonstrate the efficacy of Cosine Delta-based stylometry in quantifying the literariness of texts, while also highlighting the methodological challenges of literariness, especially in the case of Stephen King. The code and data to reproduce our results are available at <https://github.com/andreascv/kinglit>

## 1 Introduction

Stephen King, as both a highly successful and critically acclaimed author, has been referred to as “a contradiction in the literary world, a genre novelist who has achieved both popular and critical success,” (Anderson, 2020, p. 17). In this paper we investigate the literariness of King’s corpus using data-driven stylometric methods.

Stylometry is often operationalized using distances in vector space based on the most frequent

words (MFW) in texts. Stylometry was canonically developed for questions of authorship attribution, but it has increasingly been applied to other research questions, including author profiling (e.g. age and gender), and stylochronometry, or the order in which an author wrote their works (Neal et al., 2017). In this paper we propose to further the quantification of literariness as an additional application of stylometry.

The term *literariness* was introduced by the Russian formalist Jakobson (1921, p. 11), although we follow a more general definition: “The sum of special linguistic and formal properties that distinguish literary texts from non-literary texts” (Baldick, 2008). Recent work has approached literariness with data-driven methods. The Riddle of Literary Quality project (2012-2019) focuses on predicting reader judgments of literary quality (Koolen et al., 2020) using textual features. Van Cranenburgh and Koolen (2015) report that machine learning solely using bigrams “can successfully separate novels that are seen as highly literary from less literary ones.” Moreover, content and style bigrams achieve comparable accuracy, indicating that literariness is reflected in both thematics and stylistics of texts. Van Cranenburgh and Bod (2017) report that an ensemble of lexical, syntactic, and text complexity features “contribute to explaining judgments of literature,” while *individual* syntactic constructions and *individual* bigrams are weak predictors of literariness, as “the top 100 features contribute only 34% of the total [literariness] prediction.” Van Cranenburgh et al. (2019) report that even with short fragments of several pages, literariness can be predicted effectively; a keyword analysis reveals that literariness is associated with particular stylistic differences, such as strong and weak pronouns. Piper and Portelance (2016) present another study of contemporary fiction and focus on the distinctions between prize-winning, bestselling, and genre fiction in American novels, employing LIWC fea-

**K03:** *The Shining* (Smith, 2002; Hendrix, 2013)  
**K06:** *The Dead Zone* (Smith, 2002; Hendrix, 2013)  
**K11:** *Shawshank* (Romano, 2018)  
**K16:** *Pet Sematary* (Smith, 2002)  
**K19:** *It* (King, 2006; Gates, 1986)  
**K20, 29, 38, 45-48, 60:** *Dark Tower* series (Semenza, 2006, p. 73)  
**K21:** *Misery* (King, 2006; Katzenbach, 1987; Banks, 1987; Smith, 2002)  
**K32:** *Dolores Claiborne* (King, 2006)  
**K39:** *Bag of Bones* (Cruz, 2014; USA Today, GQ: Powells, 1999; Hendrix, 2015a)  
**K51:** *Lisey's Story* (Maslin, 2006)

Table 1: King works categorized as high literary.

tures, multidimensional scaling, and supervised models. Underwood and Sellers (2016) present a study with a similar diachronic aim, demonstrating that a logistic regression based on word frequencies is effective in distinguishing between amateur poetry and critically reviewed poetry.

The studies above demonstrate that literariness is readily reflected in textual features, but these studies worked with sizeable datasets and reliable indicators of literariness. But what if literariness itself is in question in the texts? What if there is only limited data available, as in the case of a single author? Moreover, in the absence of consensus on literariness based on reader surveys or literary award juries, we must contend with a “ground truth” that is potentially more subjective and unreliable. We propose to build upon these studies by exploring stylometric methods for quantifying the literariness of works by a single author, and, again, not just any author, but an author known for occupying an exceptional position across high and low literariness in fiction. Since our data and task are challenging, we opt for a proven stylometric method: Cosine Delta (Smith and Aldridge, 2011; Evert et al., 2017), a variant of Burrow’s Delta (Burrows, 2002). While this method has been especially successful for authorship attribution, we expect it to be effective for stylistic aspects in general, of which literariness is an instance.

## 2 Materials and Methods

**King novels and novellas** (73 texts). The main corpus contains all fiction solely authored by Stephen King published through 2020. Of these texts, 17 are categorized as high literary (Table 1), 13 as low literary (Table 2), and the rest are not categorized. The motivation for this categorization is presented in the following section.

**K04:** *Rage*\*  
**K07:** *Firestarter*\*  
**K09:** *Cujo*\*  
**K10:** *The Running Man*\*  
**K15:** *Christine* (Allen, 2012; Hendrix, 2013)  
**K17:** *The Eyes of the Dragon*\*  
**K18:** *Thinner*\*  
**K22:** *The Tommyknockers*\*  
**K37:** *The Regulators* (Hendrix, 2015b)  
**K40:** *The Girl Who Loved Tom Gordon* (Allen, 2012)  
**K43:** *Dreamcatcher*\*  
**K50:** *Cell* (King, 2006; Noonan, 2006)  
**K52:** *Blaze*\*

Table 2: King works categorized as low literary. Texts with \* are suggested by the authors.

We contrast the King corpus with two comparison corpora:

**Literary fiction** (38 novels). Inclusion criteria include: National Book Award winning authors published from 1974–2020, and, to attempt to control for diachronic change, American authors roughly the same age (within 10 years) as Stephen King (born 1947). For each author, we select the novel which won the Award, and an arbitrarily chosen second novel by the same author, to increase the corpus size. No effort was made to control for genre in this corpus, but this could be considered in future work.<sup>1</sup>

**Popular fiction** (36 novels). Inclusion criteria include: *New York Times* Bestsellers (top 10 only) published from 1974–2020, American authors near the same age (within 10 years) as Stephen King, in genres King himself has written in: thriller and/or horror. We select two novels for each author.<sup>2</sup>

<sup>1</sup>Literary fiction corpus: Thomas Pynchon, *Gravity’s Rainbow* (1974). Robert Stone, *Dog Soldiers* (1975). Tim O’Brien, *Going After Cacciato* (1979). John Irving, *The World According to Garp* (1980). Alice Walker, *The Color Purple* (1983). Larry Heinemann, *Paco’s Story* (1987). Pete Dexter, *Paris Trout* (1988). John Casey, *Spartina* (1989). Charles R. Johnson, *Middle Passage* (1990). Charles Frazier, *Cold Mountain* (1997). Alice McDermott, *Charming Billy* (1998). Julia Glass, *Three Junes* (2002). Lily Tuck, *The News from Paraguay* (2004). Richard Powers, *The Echo Maker* (2006). Denis Johnson, *Tree of Smoke* (2007). Jaimy Gordon, *Lord of Misrule* (2010). Louise Erdrich, *The Round House* (2012). James McBride, *The Good Lord Bird* (2013). Sigrid Nunez, *The Friend* (2018).

<sup>2</sup>Popular fiction corpus: Allan Folsom, *The Day After Tomorrow* (1994); *Day of Confession* (1998). Anne Rice, *The Witching Hour* (1990); *Taltos* (1994). Catherine Coulter, *Riptide* (2000); *Knock Out* (2009). Charlaine Harris, *Definitely dead* (2006); *Dead Reckoning* (2011). Dean Koontz, *Lightning* (1998); *By the Light of the Moon* (2002). Iris Johansen, *Final Target* (2001); *Countdown* (2005). James Patterson, *Cat and Mouse* (1997); *Cross the Line* (2016). Janet Evanovich, *Hot Six* (2000); *Plum Spooky* (2009). John Grisham, *The Firm* (1991); *The Racketeer* (2012). Michael Connelly, *A Darkness*

The fundamental issue with using this corpus as a proxy for low literariness is whether the identification of “popular” “genre” fiction with “low literariness” is valid, and whether the distinctions between high and low fiction as described by e.g., (Huysen, 1986) as “the Great Divide” are as great in recent fictional texts. We acknowledge this issue but, in the absence of e.g. large datasets of reader surveys on the literariness of English-language novels, choose this as the best option.

## 2.1 “High” and “low” literariness King novels

Despite having sold an estimated 350 million books (Heller, 2016), King’s texts have been “the object of very little critical study” (Anderson, 2017, p. 5). While this overstates the case, questions of literariness certainly suffuse early King criticism and much academic discourse on King to this day, with early critics providing value judgments on whether King is or is not “literature,” whether he is a “mere” horror or “genre” writer or somehow more “literary,” most vociferously Harold Bloom, who wrote that “King has replaced reading” (Bloom, 2006, p. 2) and “King’s books [...] are not literary at all, in my critical judgment” (Bloom, 2006, p. 207). Despite such criticism, King has gradually attained “high” literary approval, through, *inter alia*, the National Book Foundation’s 2003 Medal for Distinguished Contribution to American Letters and publications in such highbrow venues as *The New Yorker* and *The Atlantic*. “King’s avowed mission [has been] to dismantle the either/or logic of the binary opposition between popular and serious” fiction, per Birke (2014), through novels which, to this day, fall variously along spectrums of “genre” and “high” subject matter and stylistics.

King himself identified some of his books which may be more or less literary in an interview with *The Paris Review*, suggesting that some of his novels are “entertainments,” including *Cell* (a horror thriller in which mobile phones transform people into mindless killers) while others “work on more than one level,” including *Misery*, *Dolores Claiborne*, and *It*, the latter of which King assessed as “the most Dick-

ensian of my books” (King, 2006). Scholars are, of course, under no obligation to follow King’s classification of his books into “entertainments” and literary novels which “work on more than one level.” Although Koolen et al. (2020) have performed literariness experiments benefiting from reader surveys for literariness ratings, no such resource exists for King, which leads us to compare the results of our stylistometric experiments with some novels which King and critics have claimed are more or less literary.

Identifying King’s high and low literariness texts is certainly a subjective exercise which only foregrounds problematic issues of literariness as a concept. Genre and literariness are related but separate concepts—*Pet Sematary*, for instance, is unambiguously a horror novel, yet has been considered so fine in its execution that it becomes “literary,” and media critics who assign value judgments of “good” or “bad” can do so quite separately from issues of literariness; a King novel can be “low literary” but “good”—an unputdownable thriller—or “high literary” but “bad”—ambitious attempts which fail. We have thus begun with two buckets of novels in which critics specifically raise the issue of literariness. While competing critics have sometimes proclaimed the same King novel as high or low literariness (e.g. *11/22/63*), other King works have attained a critical consensus, most notably, the *Dark Tower* series, “what many critics believe to be Stephen King’s most literary of works,” per (Semenza, 2006, p. 73), with some academic commentators discussing the series in terms of post-modernism, a high literary designation (Anderson, 2020; Buday, 2015). It is easier to find critics proclaiming a King novel to be high in literary qualities than damning a King novel as unliterary trash, however, so we supplement our “low literariness King” bucket with nine more texts based upon our critical judgment, especially examples of relatively straightforward genre exercises which have not been acclaimed by literary-minded critics. This provides us with data points to test stylistometric classification of King’s literariness.

## 2.2 Stylometry and Machine Learning

The texts are lowercased and tokenized with a regular expression that includes apostrophes in words (no other punctuation is considered inside words). We apply the Cosine Delta method (Smith and Aldridge, 2011; Evert et al., 2017), a variant of Bur-

---

*More Than Night* (2001); *The Wrong Side of Goodbye* (2016). Michael Crichton, *Congo* (1980); *Timeline* (1999). Nelson DeMille, *The Charm School* (1988); *Night Fall* (2004). Patricia Cornwell, *All That Remains* (1992); *The Scarpetta Factor* (2009). Peter Straub, *Ghost Story* (1979); *Koko* (1988). Robert Crais, *Taken* (2012); *Suspect* (2013). Sue Grafton, *H Is for Homicide* (1991); *T is for Trespass* (2007). Thomas Harris, *Red Dragon* (1981); *Hannibal* (1999). Tom Clancy, *The Hunt for Red October* (1984); *Rainbow six* (1998).

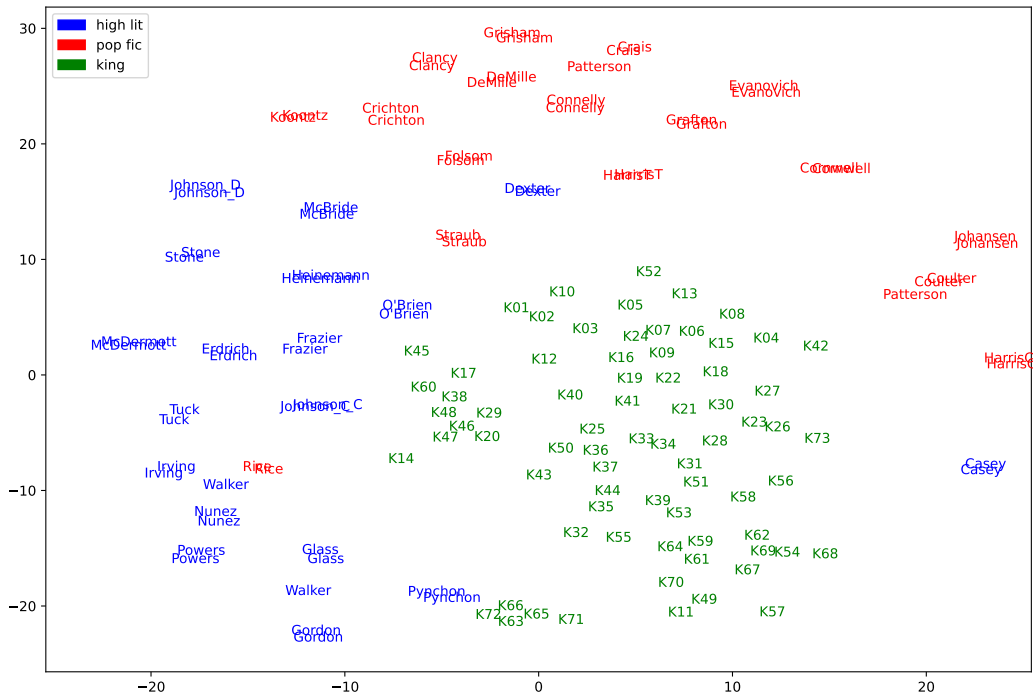


Figure 1: A t-SNE scatter plot of z-scored frequencies of 1000 MFW.

row’s Delta (Burrows, 2002), which has proven to be highly effective for authorship attribution. Based on the results in Evert et al. (2017), we work with the 1000 most frequent words; a higher number works as well, but we wished to limit the influence of non-stylistic features such as character names and topical words.

The Cosine Delta method begins by extracting relative frequencies, which are standardized into z-scores. Relative frequencies ensure that document length is ruled out as a factor. Standardization involves subtracting the mean such that each feature has a mean of 0, as well as normalizing to unit variance (standard deviation of 1). This ensures that the differences between document vectors are not dominated by high frequency words, since each feature contributes equally.

We select the 1000 most frequent words and fit the standardized frequencies based on the two comparison corpora (high literary and popular fiction), and apply the same selection and standardization on the King texts. This ensures that the feature selection and standardization is not influenced by the King texts.

The Cosine Delta provides distance measure document vectors which reflect the texts’ stylistic similarity, which can be explored in unsupervised fashion with dimensionality reduction and clustering methods. However, as we are particularly inter-

ested in estimating the literariness of texts from their word frequencies, we operationalize this by using the popular fiction and high literary novels as proxies for low and high literariness, respectively. Cosine Delta provides a distance between any two texts, and we could classify the literariness of a text by selecting its nearest neighbor, but this assumes that each word frequency is equally important for literariness, and does not tell us where to draw the boundary between popular and literary.

We therefore train a crossvalidated, regularized logistic regression model on predicting whether a text is from the high literary or popular fiction subset (the King texts are not included in this experiment). During crossvalidation, we ensure that both novels by an author are either in the training or in the test fold, not both; this avoids the situation where the model is able to learn a shortcut by exploiting the author signal. Regularization ensures that the model is able to work with a large number of features effectively, even though the number of features is much larger than the number of data points ( $1000 > 148$ ). Logistic regression, despite the name, is a linear classification model, which has an advantage over models such as Support Vector Machines (SVM) in that it produces calibrated probabilities which indicate the confidence of a classification on a scale of 0 to 1. For a given document vector, we use this probability as an indication of

how likely it is that the document is a high literary novel, which we then apply to the King texts.

### 3 Experiments and Analysis

#### 3.1 Unsupervised results

We first present some unsupervised explorations in which we analyze the observed z-scores directly, after which we turn to supervised models that are optimized to predict the literariness labels. By inspecting the mean z-scores in each corpus, we can determine which word frequencies diverge the most (on average) from the comparison corpora (high literary and popular novels) as a whole. The top 15 words with the lowest and highest z-scores are:

**Typical for King:** ones, kid, thought, one, sounded, least, began, idea, although, mouth, it, sound, mind, gone, almost

**Atypical for King:** meet, moved, family, figure, attention, we, near, whether, minute, learned, met, worked, within, though, return

**Typical for high literature:** way, whose, girls, dream, funny, among, fall, each, whole, every, one, sometimes, own, themselves, day

**Typical for popular fiction:** door, hell, okay, we're, second, opened, check, probably, realized, checked, phone, killed, information, area, minutes

These bags of words resemble the output of topic modeling, an unsupervised method to discover latent semantic structures in texts (Blei, 2012). Even though these lists were not produced by topic modeling, the cautions against naively interpreting bags of words produced by topic modeling are relevant here; Schmidt (2012) suggests that topic models are known for creating “unexpected juxtapositions,” while Goldstone and Underwood (2012) argue that “interpreters really need to survey a topic model as a whole, instead of considering single topics in isolation.” With these cautions in mind, we observe that the typical words for popular fiction seem action- and procedure oriented, and include markers of informality (‘okay,’ contractions). These typical words for high literature tend to be conceptual, while these typical words for King display a number of words relating to inner state, such as *thought*, *idea*, *mind*, which may relate to claims by e.g. Magistrale (2013, p. 355), who praises “King’s ability to place the reader in the consciousness of a character.” The keyness of *Kid*, meanwhile, probably relates to the fact that “King regularly makes use of the child protagonist in his novels” (Olson, 2020, p. 2). On the other hand, among the atypical words for King are words related to family and time, which have been reported as distinctive keywords of literary works (Piper and Portelance, 2016; van

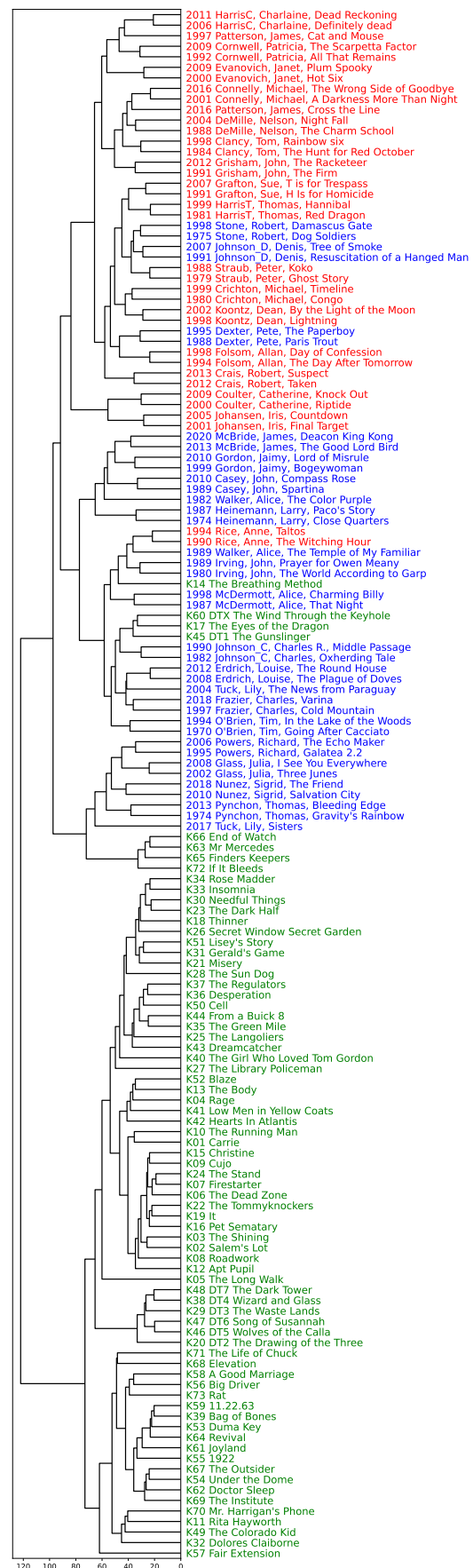


Figure 2: A hierarchical clustering of the whole corpus based on Cosine Delta of 1000 MFW.

	prec.	recall	F1	support
high lit	94.7	94.7	94.7	38
popular	94.4	94.4	94.4	36
macro avg	94.6	94.6	94.6	74

Table 3: Crossvalidated classification scores on predicting high vs popular in the comparison corpora.

Cranenburgh and Koolen, 2015). However, it is expected that individual features provide limited information. To get a better understanding of the stylistic differences captured by the z-scores, we should consider all of the 1000 MFW.

Figure 1 shows a dimensionality reduction scatter plot using t-SNE with a perplexity (i.e., number of nearest neighbors) of 15. In this plot, the subcorpora are shown in different colors for visualization purposes. Note that these labels were not given to t-SNE; this scatter plot is therefore still an unsupervised analysis. We can observe that high literary fiction and popular fiction tend to cluster together, with the strongest outlier being vampire fiction author Anne Rice. The texts in the left of King’s cluster are those most firmly gravitating toward the high literary cluster, and interestingly, at least eight *Dark Tower* books are placed there by the model, consistent with critical assessments of the *Dark Tower* series as among King’s most experimental and even “postmodern.” We had suggested *The Eyes of the Dragon* (K17) as a low literariness text, as it is King’s only attempt at a Tolkien-esque fantasy (and thus within the purview of the popular fantasy genre), but the model places it firmly within high literary. Another cluster in the bottom left (and thus closest to high literary) features a number of recent King texts which are thematically quite different, but perhaps united in stylistics (K63, 65-66, 71-72). As an aside, one of the King texts suggested as most literary by this model, “The Breathing Method” (K14), is a largely forgotten early novella, but we suggest that the minor power of models such as these to simply highlight texts for literary scholars to revisit (with fresh eyes and a newfound attention to literariness) is a small but worthwhile humanistic contribution.

Another way of exploring the intertextual distances is with a hierarchical clustering, which provides a more ordered way of assessing the stylistic similarities in the corpus (Figure 2). While both the scatter plot and hierarchical clustering highlight which texts are most similar to each other and

	prec.	recall	F1	support
high lit	70.6	70.6	70.6	17
popular	61.5	61.5	61.5	13
macro avg	66.1	67.1	66.1	30

Table 4: Classification scores on the King novels with literariness categorizations, trained on the high lit and popular corpora. The support column shows the number of datapoints in each subset.

	prec.	recall	F1	support
high lit	71.4	58.8	64.5	17
popular	56.2	69.2	62.1	13
macro avg	63.8	64.0	63.3	30

Table 5: Crossvalidated classification scores on the King novels with literariness categorizations, trained on the King corpus.

form clusters, the scatter plot has the limitation of having to summarize the similarities across 1000 dimensions in two dimensions, and since t-SNE is a randomized algorithm with several important parameters that need to be tuned, results can vary and sometimes highlight different aspects in the data. The hierarchical clustering, on the other hand, does not reduce the dimensionality and shows a different kind of structure in the data, in a deterministic fashion. The tree is built bottom up, starting with each text in its own cluster; at each step, the two most similar clusters are then merged based on cosine distance. The horizontal length of the branches are proportional to the similarity of the clusters (Evert et al., 2017). Since Cosine Distance is so adept at detecting authorship, most authors with two novels form a cluster.<sup>3</sup> In other respects, this method confirms clusterings that are also apparent in the scatter plot, which is an indication of the robustness of those results.

### 3.2 Supervised results

In the unsupervised results, many types of variation in the z-scores may play a role, and we do not obtain a direct measure of how well literariness is reflected in the z-scores. We therefore turn to supervised models in this section.

Table 3 presents classification results for the high literature and popular fiction comparison corpora. The evaluation is done using 3-fold crossvalidation. The classification is highly accurate, indicating that

<sup>3</sup>There are two exceptions: Alice Walker and Lily Tuck.

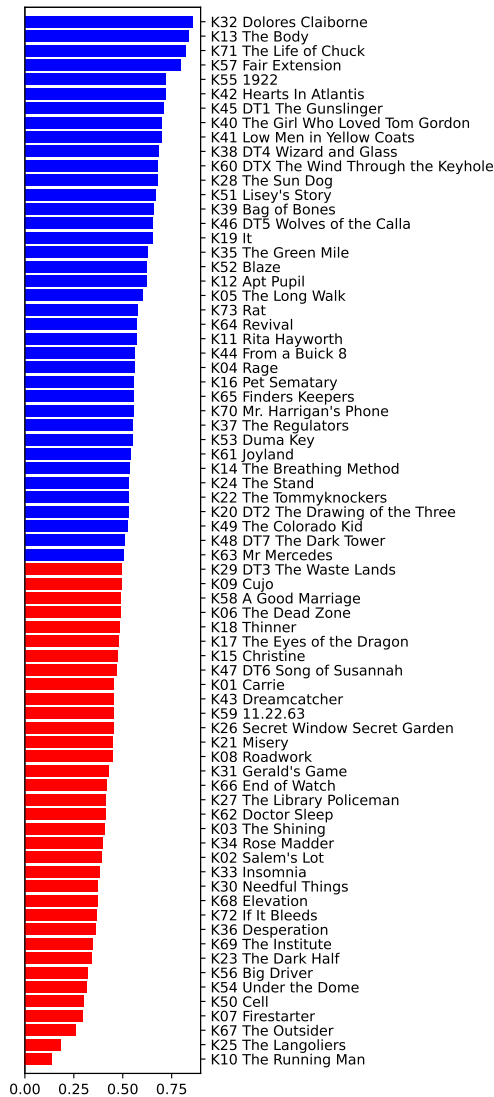


Figure 3: Probability that each King text is high literary according to logistic regression train on popular fiction and high literary novels.

there is a robust difference in word usage across these two corpora, and confirming, as in previous work, that literariness is predictable from textual features. This result also supports the suitability of Cosine Delta for modeling literariness.

Table 4 reports the classification results for the King novels we categorized as high and low literary. In early experiments, the model predicted almost all King novels as literary. There are two possible explanations for this: either all the King novels are indeed literary, or this points to a mismatch of the literariness differences in the King novels compared to the training data which we use as proxy for literariness. Since we know that there are degrees of literariness in the King corpus, we consider the second explanation more likely. In this case we

can conclude that the King novels are closer to the literary novels in our reference corpus than to the popular novels. We therefore need to correct for this imbalance in the training data. We achieve this by specifying during training that misclassifications on popular novels are twice as important as misclassifications on literary novels (i.e., a skewed class prior). The result is that about half of King novels are predicted to be literary. With a macro average of 66.1%, the results outperform a random baseline, which produces macro average F1 scores with a mean of 50%.

We can inspect the behavior of the model by looking at predictions for individual novels. Figure 3 shows the probability of being a high literary novel for each text in the King corpus, ranked from most to least literary. Compared to the scatter plot, this provides a more direct way of assessing to what extent literariness is reflected in the textual features, since the predictive model focuses on finding the differences between literary and popular novels while ignoring other variation in the z-scores. A number of the *Dark Tower* novels again score among the highest in the literariness score, while other King novels from the high literary bucket in Table 1 also score high here, including *Lisey's Story* and *It*. The highest scoring novel, *Dolores Claiborne*, is consistent with critical assessment as being high literary for King, with critics such as Senf (1998) praising the realist psychological portraits of its female characters. The literariness of the 2nd, 5th, and 6th top results can somewhat be implied by the fact that these have been adapted as critically acclaimed films, *The Body* as the 1986 film *Stand by Me*, with *1922* and *Hearts in Atlantis* as films of the same titles. Not all of our predicted low literariness texts score particularly low by this measure, but *Cell*, which King singled out as one of his “entertainments,” is the fifth lowest-scoring text.

To investigate the possible difference in standards between the categorizations of high literary and popular novels on the one hand, and the labels for low and high literary King novels on the other, we evaluate a different classifier on the same set of King novels, but trained only on the King novels using 5-fold crossvalidation; see the scores in Table 5. For this model we did not use a skewed class prior. The results are slightly worse, but given the much smaller training set, the results are still quite good.



## 4 Conclusion and Future Work

Our results confirm that popular fiction and high literary novels can be distinguished remarkably effectively, as shown by the classification F-scores. Applying the same model to the King corpus gives promising results, but the classification F-scores are not as impressive. On the one hand, this is entirely expected: the model is evaluated on a different domain than what it was trained on, and some of the patterns it exploits in the training data will not apply. On the other hand, there may be a deeper problem; as many of King's works display aspects of high and low literariness, in both thematics and perhaps style, King may be an exceptionally problematic case in any quantitative measurement of literariness.

Moreover, the lower performance on the set of labeled King texts may indicate that more training data and more sophisticated models are needed, but it may also be a sign that the categorizations of the King novels are not consistent with the textual features or the categorizations of the popular and high literary novels in the comparison corpora. What is taken as a misclassification by the model in our evaluation could also point to a misclassification by critics; moreover, since our labels reflect the perspective of different individuals, they are not necessarily consistent with each other; these critics may have applied different standards. These are methodological challenges that should be addressed in future work.

Robust evidence emerging from the experiments on King's work is the consistent ranking by the models of novels in King's *Dark Tower* series and other texts as being among his more literary works, consistent with critical assessments. This shows promise for Cosine Delta and other stylometric methods in literariness quantification. While the question of King's literariness is not settled, our results confirm the notion that King's work spans the spectrum of low and high literariness.

We also leave for future work the case of literariness and Richard Bachman, a pen name used by King in early novels and two later works. Despite the nom de plume's practical origins (King and his publishers were wary of publishing two novels in one year by the same author), King later described Bachman as a "persona" and "alter ego," stating that the early novels had been written "in a Bachman state of mind: low rage, sexual frustration, crazy good humor, and simmering despair" (King, 1996). We suggest that an exploration of King/Bachman

would merit a dedicated, mixed-method study, particularly as there is far from critical consensus on the literariness of the various Bachman books, individually or collectively, let alone whether the Bachman novels (some separated by decades) can convincingly be argued to share distinctive features. *The Running Man*, a Bachman book, was scored by our model as the lowest King novel in literariness, and while a hypothetical critic might dismiss that novel as a low-literary dystopian thriller, the book has many literary aspects, as discussed by [Texter \(2007\)](#). Meanwhile, *The Long Walk*, the second Bachman book, scored relatively highly in literariness in our model, consistent with our own critical assessment of the novel as a fine dystopian allegory of the Vietnam War. Is Bachman a signal, or merely noise?

## Acknowledgements

We are grateful to Corina Koolen, Karina van Dalen-Oskam, and three anonymous reviewers for their feedback.

## References

- Dwight Allen. 2012. [My Stephen King problem: A snob's notes](#). Los Angeles Review of Books.
- James Arthur Anderson. 2017. *The Linguistics of Stephen King: Layered Language and Meaning in the Fiction*. McFarland.
- James Arthur Anderson. 2020. *Excavating Stephen King: A Darwinist Hermeneutic Study of the Fiction*. Lexington Books.
- Chris Baldick. 2008. Literariness. In *The Oxford Dictionary of Literary Terms*. Oxford University Press, USA.
- Carolyn Banks. 1987. Summer thrillers from three masters. *The Washington Post*.
- Dorothee Birke. 2014. [Challenging the divide? Stephen King and the problem of 'popular culture'](#). *The Journal of Popular Culture*, 47(3):520–536.
- David M. Blei. 2012. [Probabilistic topic models](#). *Communications of the ACM*, 55(4):77–84.
- Harold Bloom. 2006. *Stephen King*. Infobase Publishing.
- Maroš Buday. 2015. [Exploring „the dark tower“: Stephen King's postmodern epic](#). *Studia Anglica Resoviensia*, 12:135–145.
- John Burrows. 2002. 'Delta': a measure of stylistic difference and a guide to likely authorship. *Literary and linguistic computing*, 17(3):267–287.

- Gilbert Cruz. 2014. [All 64 Stephen King books, ranked](#). New York Vulture.
- Stefan Evert, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. 2017. [Understanding and explaining Delta measures for authorship attribution](#). *Digital Scholarship in the Humanities*, 32(Supplement 2):ii4–ii16.
- David Gates. 1986. The creature that refused to die. Newsweek.
- Andrew Goldstone and Ted Underwood. 2012. [What can topic models of PMLA teach us about the history of literary scholarship?](#) *Journal of Digital Humanities*, 2(1).
- Karen Heller. 2016. Meet the writers who still sell millions of books. Actually, hundreds of millions. *The Washington Post*, 20.
- Grady Hendrix. 2013. [The great Stephen King reread: Christine](#). Tor.com.
- Grady Hendrix. 2015a. [The great Stephen King reread: Bag of Bones](#). Tor.com.
- Grady Hendrix. 2015b. [The great Stephen King reread: The Regulators](#). Tor.com.
- Andreas Huyssen. 1986. *After the great divide: Modernism, mass culture, postmodernism*, volume 399. Indiana University Press.
- Roman Jakobson. 1921. *Novejšaja russkaja poëzija*. Politika, Prague.
- John Katzenbach. 1987. Sheldon gets the ax: Misery by Stephen King. *New York Times Book Review*, 31.
- Stephen King. 1996. The importance of being Bachman. In *The Bachman Books*, pages v–ix. Plume.
- Stephen King. 2006. [Stephen King, the art of fiction no. 189. Interviewed by Nathaniel Rich & Christopher Lehmann-Haupt](#). The Paris Review.
- Corina Koolen, Karina van Dalen-Oskam, Andreas van Cranenburgh, and Erica Nagelhout. 2020. [Literary quality in the eye of the Dutch reader: The national reader survey](#). *Poetics*.
- Tony Magistrale. 2013. Why Stephen King still matters. In Charles L. Crow, editor, *A Companion to American Gothic*, pages 353–65. John Wiley & Sons.
- Janet Maslin. 2006. [Her Story of Him, Both Tender and Terrible](#). *The New York Times*.
- Tempestt Neal, Kalaivani Sundararajan, Aneez Fatima, Yiming Yan, Yingfei Xiang, and Damon Woodard. 2017. [Surveying stylometry techniques and applications](#). *ACM Comput. Surv.*, 50(6).
- Erica Noonan. 2006. [In King’s uneven ‘Cell,’ evil rings up Boston](#). The Boston Globe.
- Debbie Olson. 2020. Introduction: Stephen King’s fictional children. In Debbie Olson, editor, *Children and Childhood in the Works of Stephen King*, pages 1–12. Lexington Books.
- Andrew Piper and Eva Portelance. 2016. [How cultural capital works: Prizewinning novels, bestsellers, and the time of reading](#). *Post45*, 10.
- Powells. 1999. [Synopses & reviews of Bag of Bones by Stephen King](#). Publisher’s website.
- Aja Romano. 2018. [Stephen King has spent half a century scaring us, but his legacy is so much more than horror](#).
- Benjamin M. Schmidt. 2012. [Words alone: Dismantling topic models in the humanities](#). *Journal of Digital Humanities*, 2(1).
- Gregory M. Colón Semenza. 2006. [Adapting Milton for children: Margaret Hodges’s Comus \(1996\) and the tale of Childe Rowland](#). In *Milton in Popular Culture*, pages 71–82. Springer.
- Carol Senf. 1998. Gerald’s Game and Dolores Claiborne: Stephen King and the evolution of an authentic female narrative voice. In Kathleen Margaret Lant and Theresa Thompson, editors, *Imagining the Worst: Stephen King and the Representation of Women*, pages 91–110. Greenwood Press, London.
- Greg Smith. 2002. [The literary equivalent of a Big Mac and fries?: academics, moralists, and the Stephen King phenomenon](#). *The Midwest Quarterly*, 43(4):329.
- Peter W. H. Smith and W. Aldridge. 2011. [Improving authorship attribution: Optimizing Burrows’ delta method](#). *Journal of Quantitative Linguistics*, 18(1):63–88.
- Douglas W. Texter. 2007. [A funny thing happened on the way to the dystopia: The culture industry’s neutralization of Stephen King’s The Running Man](#). *Utopian Studies*, 18(1):43–72.
- Ted Underwood and Jordan Sellers. 2016. [The longue durée of literary prestige](#). *Modern Language Quarterly*, 77(3):321–344.
- Andreas van Cranenburgh and Rens Bod. 2017. [A data-oriented model of literary language](#). In *Proceedings of EACL*, pages 1228–1238.
- Andreas van Cranenburgh and Corina Koolen. 2015. [Identifying literary texts with bigrams](#). In *Proceedings of CLFL*, pages 58–67.
- Andreas van Cranenburgh, Karina van Dalen-Oskam, and Joris van Zundert. 2019. [Vector space explorations of literary language](#). *Language Resources and Evaluation*, 53(4):625–650.