Correlations between GoodReads Appreciation and the Sentiment Arc Fractality of the Grimm brothers' Fairy Tales

Yuri Bizzoni[1], Mads Rosendahl Thomsen[2], Ida Marie S. Lassen[1] and Kristoffer Nielbo[3]

[1]*Center for Humanities Computing Aarhus, Aarhus University, Jens Chr. Skous Vej 4, Building 1483,DK-8000 Aarhus C*

[2]*School of Communication and Culture - Comparative Literature, Aarhus University, Langelandsgade 139, Building 1580, DK-8000 Aarhus C*

[3]*Center for Humanities Computing Aarhus & Interacting Minds Centre, Aarhus University, Jens Chr. Skous Vej 4, Building 1483, 3rd floor, DK-8000 Aarhus C*

Abstract

Despite their widespread popularity, fairy tales are often overlooked when studying literary quality with quantitative approaches. We present a study on the relation between sentiment fractality and literary appreciation by testing the hypothesis that fairy tales with a good balance between unpredictability and excessive self-similarity in their sentiment narrative arcs tend to be more popular and more appreciated by audiences of readers. In short, we perform a correlation study of the degree of fractality of the fairy tales of the Grimm brothers and their current appreciation as measured by their Goodreads scores. Moreover, we look at the popularity of these fairy tales through time, determining which ones have come to form a strong "internal canon" in the corpus of the authors and which one have fallen into relative obscurity.

Keywords

computational narratology, sentiment analysis, fractal analysis, literary quality assessment

1. Introduction

In recent years the increase in size of available corpora and the possibility of performing complex operations on textual data have given rise not only to an explosion of tools for the exploratory analysis of literary collections [9], but also to new hypotheses in literary and aesthetic studies, all the while making existing complex hypotheses relatively easier to test [34, 39]. This change in the ways we can study texts has on one hand brought about the possibility of testing for complex patterns in linguistic data, such as the presence of long-ranging regularities and multifactorial structures, defining new research questions for the field. On the other hand, it has become possible to study traditional questions with brand new methodologies.

CHR 2022: Computational Humanities Research Conference, December 12 – 14, 2022, Antwerp, Belgium
yuri.bizzoni@cc.au.dk (Y. Bizzoni); madsrt@cc.au.dk (M. R. Thomsen); idamarie@cas.au.dk (I. M. S. Lassen); kln@cas.au.dk (K. Nielbo)
0000-0002-6981-7903 (Y. Bizzoni); 0000-0002-4975-6752 (M. R. Thomsen); 0000-0001-6905-5665 (I. M. S. Lassen); 0000-0002-5116-5070 (K. Nielbo)

One growingly popular object of research falling in the second category is the concept of literary quality [43]. Making use of large-scale quantitative means, several works have tried to define and study literary quality, focusing on different aspects of literary texts and their contexts [14]. The riddle of quality is as intriguing as elusive: while large numbers of readers can often converge on their overall judgment of a piece of literature, they can often be a loss to justify why a different text, on similar topics or with a similar perspective, fails to produce for them the same experience. At the same time, no single text in the history of literature has attracted unanimous approval, and the reasons that different readers bring for their judgment of a piece can be both varied and hard to define rigorously.

The first major difficulty in this line of research is how to find proxies to represent quality itself: should one do it through prestigious prices, book sales, or crowd voting? While this question has been problematic in any study on quality, it becomes perhaps even more stringent when dealing with quantitative analyses, since it is necessary to find a way to represent it into a number or a boolean value.

It has become relatively common among quantitative studies on literary quality to use large online platforms for readers such as *GoodReads* [26] as a way to approximate the 'mean perceived quality" of a text. Platforms like *GoodReads* represent a particularly appealing source of information since they gather data from unconstrained readers, over relatively long spans of time, and can average on large numbers of individual scores [42]. Usually these studies do not take into account the average number of raters in order to avoid completely conflating perceived quality (how much readers liked a book) with popularity (how many readers rated the book). Even so, using scores from a general purpose platform like GoodReads deals with quality as a statistical mean of all raters' appreciation, which is a democratic but not unproblematic take on concept. Since it allows to take into account the perspective of very different readers, giving to each the same value, it could be argued that literary quality is a multi-dimensional problem and by using GoodReads scores we are measuring one dimension of it, while other means could be used to compound it with its other aspects.

Beyond the question of how to represent quality, the second main problem in this line of research is to decide which features we should select to explore it.

Some studies have looked into classic stylometric features, while others have tried more sophisticated properties, exploring the semantic and psycholinguistic dimensions of texts.

A set of works has looked specifically into the linearity of textual narratives, representing them as series extended through the virtual time of reading (or listening). In such cases, the focus has been on how to best represent such narratives lines or arcs and to define their most typical or relevant shapes [36].

A relatively prominent line of research has looked into the sentimental and emotional aspect of literary texts [24, 25], as well as renown movies scripts or song lyrics [8], following the idea that the emotions expressed in texts, as drawn through sentiment analysis resources [3, 1, 23, 30] and emotion recognition tools [2, 10] working at the word [31, 32, 22] or paragraph level [27, 44], might play a fundamental role in readers' engagement and response [15].

This line of research seems to have returned some promising results [40], and often these studies focus on the cultural context of narratives in general, linking their sentimental or emotional values to social or historical changes in the broadest sense [41, 17, 4].

Finally, some studies have recently linked the linear and temporal dimension of texts with that of fractal analysis, the ensemble of techniques to assess the fractal characteristics of an object and assign a fractal dimension to a dataset [13, 29, 6, 16]. Such studies hypothesize that the quality of literary narratives might be partly due to fractal properties embedded in texts [12, 19]. For example, Mohseni, Gast, and Redies [33] looked into the fractality of shallow stylometric features such as TTR and POS-R, while Hu, Liu, Thomsen, Gao, and Nielbo [21] explored the self-predictability of a novel's narrative arc. Following the latter, Bizzoni, Peura, Nielbo, and Thomsen [7] brought this analysis to the fairy tales of H.C. Andersen, finding a positive correlation between the level of fractality in the fine-grained sentiment arc of a narrative and its appreciation: on average, fairy tales with more "fractal" sentiment arcs would receive higher scores on GoodReads.

In this work we attempt to take this line of research slightly further, hypothesizing that while the level of fractality might correlate with reader appreciation or perceived literary quality, there most likely is a breaking point or "sweet spot" after which a text's sentimental fractality stops being a good thing.

To check this hypothesis we reproduce the experimental setting of Bizzoni, Peura, Nielbo, and Thomsen [7] on a selection of fairy tales from the Grimm brothers, and we test whether (a) a correlation between sentiment fractality and GoodReads scores for the Grimm brothers exists at all; and (b) whether this correlation is linear or follows a rise-and-fall pattern.

2. Data

We followed Taylor and Edwardes's 2001 English translation of *The Grimms' Fairy Tales* [20]. We chose this edition simply because it is the most popular on the Gutenberg Project, and that would help us select the fairy tales from the Grimm brothers that are most likely to have received ratings from more than one or two reviewers on GoodReads; this selection is composed of 62 tales. We drew the text from Project Gutenberg's website [38].

We worked on English both to replicate as close as possible the work of Bizzoni, Peura, Nielbo, and Thomsen [7] and to have access to some of the most tested resources for Sentiment Analysis.

This selection includes most of the best-known stories from the Grimm brothers such as *Rapunzel*, *Hansel and Gretel*, and *Red Riding Hood*. It also contains several titles that are relatively less widespread, such as *Clever Hans*, *The Salad* or *The Turnip*. Most titles of these selections are nonetheless read enough to have earned more than ten different ratings on GoodReads. As we show in Figure 1, there seems to be a weak correlation between the number of ratings and the average rating of a fairy tale, indicating that the most liked stories are also the ones attracting most votes (link between popularity and appreciation).

3. Method

We performed the sentiment analysis of the fairy tales using the VAD lexicon [31], one of the most popular word-based sentiment resources in computational linguistics [32]. Since we are working on unusually short pieces of narratives, we drew our sentiment arcs from each word score, without averaging them on sentences or paragraphs. This also allows us to operate at

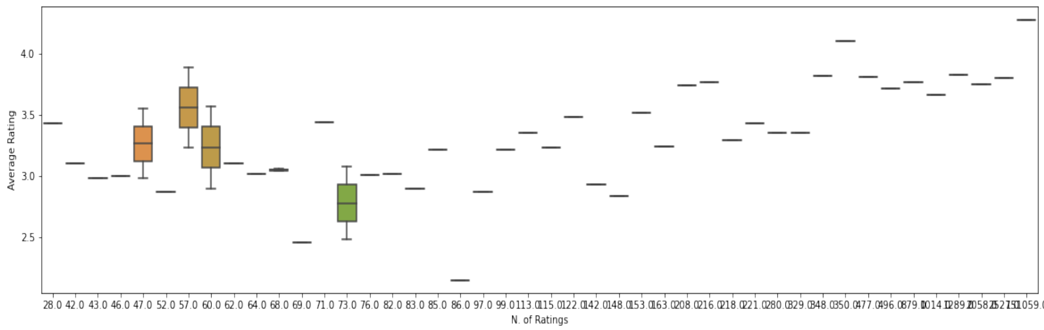

Figure 1: Number of GoodReads ratings (on the x axis) plotted against the average rating score (on the y axis) for our selection of Grimm fairy tales. A growing number of ratings seems to overall correspond to a higher average rating: more appreciated stories tend to become also more popular.

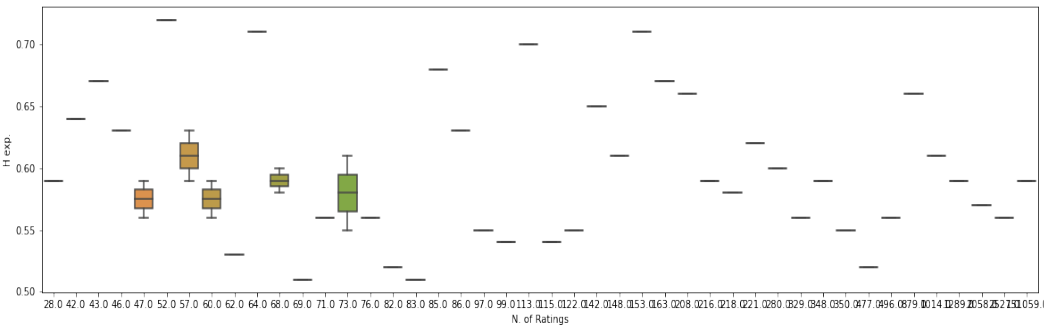

Figure 2: Number of GoodReads ratings (on the x axis) plotted against the Hurst exponent (on the y axis) for our selection of Grimm fairy tales.

the finest grain of sentiment analysis, keeping us at the interface between narrative and style, which is particularly important when working with literary quality.

As suggested by [21], we use the Hurst exponent to approximate the story arc's inner coherence. The Hurst exponent, H, is a measure of self-similar behavior. In the context of story arcs, self-similarity means that the arc's fluctuation patterns at faster time-scales resemble fluctuation patterns at slower time scales [37]. We use Adaptive Fractal Analysis (AFA) to estimate the Hurst exponent [18]. AFA is based on a nonlinear adaptive multi-scale decomposition algorithm [18]. The first step of the algorithm involves partitioning an arbitrary time series under study into overlapping segments of length $w = 2n + 1$, where neighboring segments overlap by $n + 1$ points. In each segment, the time series is fitted with the best polynomial of order M, obtained by using the standard least-squares regression; the fitted polynomials in overlapped regions are then combined to yield a single global smooth trend. Denoting the fitted polynomials for the $i - th$ and $(i + 1) - th$ segments by $y^i(l_1)$ and $y^{(i+1)}(l_2)$, respectively, where

$l_1, l_2 = 1, \cdots, 2n+1$, we define the fitting for the overlapped region as

$$y^{(c)}(l) = w_1 y^{(i)}(l+n) + w_2 y^{(i+1)}(l), \quad l = 1, 2, \cdots, n+1$$

where $w_1 = \left(1 - \frac{l-1}{n}\right)$ and $w_2 = \frac{l-1}{n}$ can be written as $(1 - d_j/n)$ for $j = 1, 2$, and where d_j denotes the distances between the point and the centers of $y^{(i)}$ and $y^{(i+1)}$, respectively. Note that the weights decrease linearly with the distance between the point and the center of the segment. Such a weighting is used to ensure symmetry and effectively eliminate any jumps or discontinuities around the boundaries of neighboring segments. As a result, the global trend is smooth at the non-boundary points, and has the right and left derivatives at the boundary [37]. The global trend thus determined can be used to maximally suppress the effect of complex nonlinear trends on the scaling analysis. The parameters of each local fit is determined by maximizing the goodness of fit in each segment. The different polynomials in overlapped part of each segment are combined using Equation 5 so that the global fit will be the best (smoothest) fit of the overall time series. Note that, even if $M = 1$ is selected, i.e., the local fits are linear, the global trend signal will still be nonlinear. With the above procedure, AFA can be readily described. For an arbitrary window size w, we determine, for the random walk process $u(i)$, a global trend $v(i), i = 1, 2, \cdots, N$, where N is the length of the walk. The residual of the fit, $u(i) - v(i)$, characterizes fluctuations around the global trend, and its variance yields the Hurst parameter H according to the following scaling equation:

$$F(w) = \left[\frac{1}{N}\sum_{i=1}^{N}(u(i) - v(i))^2\right]^{1/2} \sim w^H$$

By computing the global fits, the residual, and the variance between original random walk process and the fitted trend for each window size w, we can plot $\log_2 F(w)$ as a function of $\log_2 w$. The presence of fractal scaling amounts to a linear relation in the plot, with the slope of the relation providing an estimate of H[1]. Accordingly, a H higher than .5 indicates a degree of linear coherence (e.g., positive sentiments are followed by positive sentiments), while H lower than .5 indicates a series that tends to revert to the mean (e.g. a positive emotion always follows a negative emotion).

After determining the Hurst exponent for each fairy tale, we correlated it with the average scores on Goodreads. To compute the strength of the correlations, we used the most popular metrics: Pearson [5] and Spearman [35] correlations. After computing them on the whole dataset, we calculated each of these measures on what we considered the possible "upward" and "downward" trends of the data. To choose the breaking point to divide the data between a potentially positive and a potentially negative correlation trend, we computed a correlation matrix between all of the Hurst scores and all of the GoodReads values in our dataset (see Figure 3).

We also computed the probability distribution of seeing a highly rated fairy tale versus the probability of seeing a non-highly rated fairy tale at different Hurst values (see Figure 4). To do this, we first conventionally set a high rate threshold at 3.5 GoodReads stars, following the practice of Maharjan, Arevalo, Montes, González, and Solorio [28], and then for each Hurst

[1]Code for computing DFA and AFA is available at https://github.com/knielbo/saffine

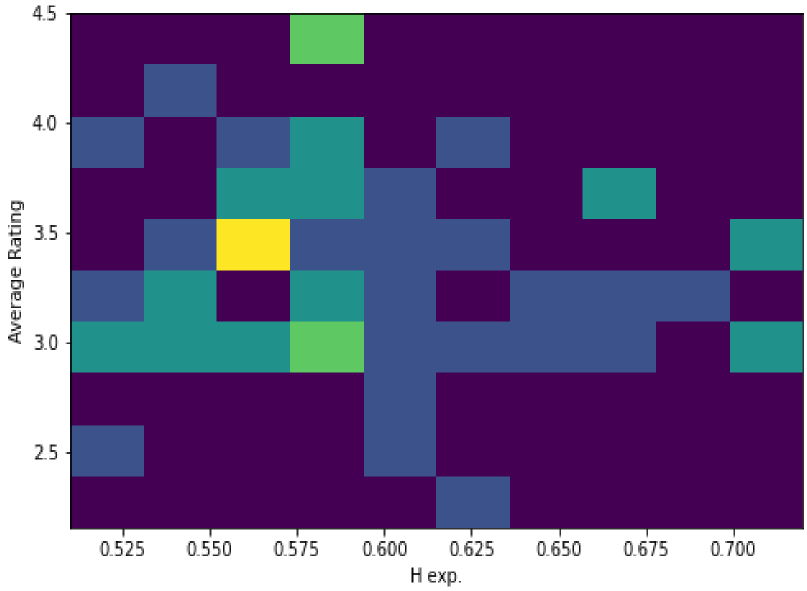

Figure 3: Correlation matrix between Hurst exponents and average GoodReads scores. The higher correlations appear to be between Hurst exponents between 0.55 and 0.6, and avg. ratings between 3.0 and 4.5. The highest correlations seem set between H=0.56 and H=0.58.

value (H=0.5, H=0.51, and so forth) we computed the probability of seeing a highly rated tale (higher than the threshold) versus the probability of seeing a non-highly rated tale.

Both these tests point to a peaking positive correlation of arc fractality with GoodReads scores between a Hurst exponent of 0.56 and a Hurst exponent of 0.6 circa, which appears in line with the findings of Bizzoni, Peura, Nielbo, and Thomsen [7].

4. Results

Our analysis shows that the correlation between the number of ratings a fairy tale receives and its Hurst exponent, on the other hand, does not appear so obvious (Figure 2). a positive correlation between a tale's Hurst exponent and its average rating up until a given Hurst value, and a negative correlation afterward. As we said in Section 3, we defined this breaking point both by checking the probabilities of a "high scoring tale" (Figure 3) and by computing the correlation matrix between Hurst exponent and GoodReads' scores (Figure 4). We found that the point that most clearly divides the positive from the negative correlation in our data appears to be Hurst=0.57, although naturally it is the whole area between H=0.56 and H=0.59 that seems to indicate an inversion in the correlation between these two variables. In this way, we can easily identify two opposing trends in our data, that appear by computing both Pearson's and Spearman's correlations. The Pearson correlation is of 0.5 (*p-value*: 0.01) for the Hurst interval 0.48-0.58, and of -0.3 (*p-value*: 0.08) for the interval 0.57-0.75. The Spearman correlation is of

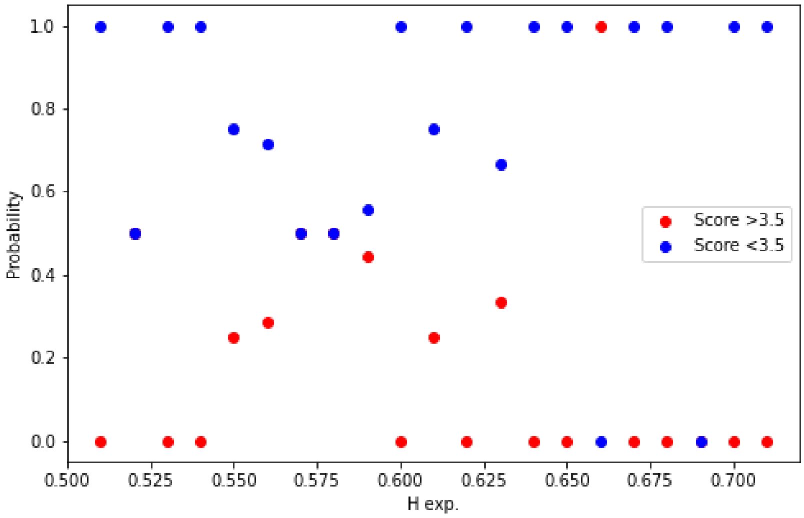

Figure 4: Probability of seeing a high scoring vs a non-high scoring fairy tale for different Hurst values. The probability of a high-scoring tale seems to peak between Hurst=0.56 and Hurst=0.6 circa.

0.45 (*p-value*: 0.02) for the Hurst interval 0.48-0.58, and of -0.37 (*p-value*: 0.02) for the interval 0.57-0.75. Following standard boundaries for the strength of linear correlations' ranges [11], we can define all these correlations as moderate or robust. In average, all but the second Pearson correlation have a p-value under the significance threshold of 0.05 [2]. In all cases, the negative correlations appear weaker (and less significant) than the positive ones. This seems to confirm what can also be seen in Figure 5: after the " sweet spot" of arc fractality, the steepness of the trendline is smoother and the link between a fairy tale's average score and its Hurst exponent decreases. Looking into the individual titles of the fairy tales is also interesting: the stories falling into the "internal canon" of the Grimm brothers seems to form a loose cluster. For example, fairy tales that have the highest number of readers on GoodReads, that have elicited reproductions in movies or cartoons, or that tend to appear most often in choice anthologies of the authors appear to cluster at the center of the graph, namely at the high point of GoodReads' appreciation and fractal equilibrium. As can be seen for example in Figure 6, *Little Red Riding Hood/Little Red-Cap* (1900 ratings), *Hansel and Gretel* (2711 ratings), *Rapunzel* (2303 ratings) fall on the upper right quadrant, while in Figure 7 *Ashputtel*, the Grimms' version of *Cinderella* (1706 ratings), *King Grisly-Beard* (230 ratings), *The Travelling Musicians [of Bremen]* (479 ratings) tend to cluster on the upper left quadrant of the graph. At the margins of the plots we can find less widespread pieces like *The Fox and the Horse* (55 ratings), *Sweetheart Roland* (88 ratings) or *Clever Hans* (75 ratings). But independently from the number of ratings, the reader

[2]The second Pearson's p-value is just slightly above the formal threshold of significance. What this indicates, rather than a Yes/No significance output, is that the correlation between Hurst values and readers' scores becomes less obvious after the peak area around H=.57. Both the value and the significance of the negative correlations are weaker than the positive correlations' ones.

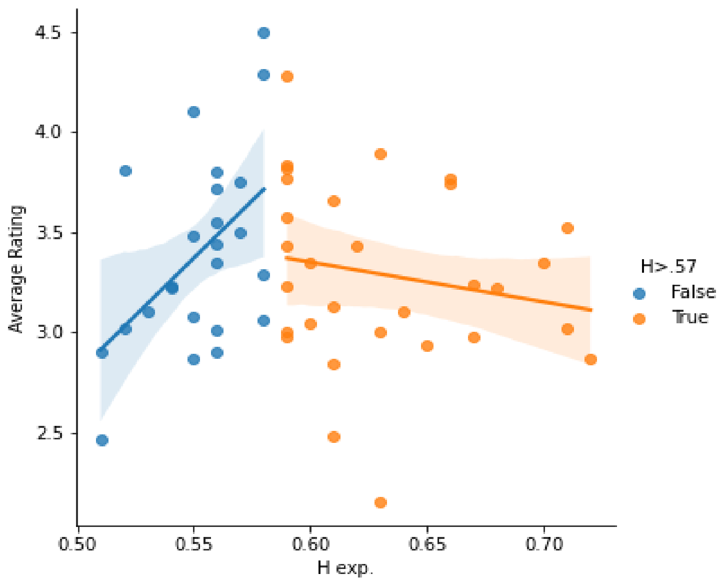

Figure 5: Correlations between Hurst exponent and average GoodReads rating, breaking point set at H=0.57.

will recognize the stories in the center as being among the best known from the authors.

5. Discussion

We have computed a correlation between sentiment arcs' Hurst exponents and average GoodReads' scores on a selection of Grimms' fairy tales, both confirming the findings of previous literature on H.C. Andersen [7] and validating our hypothesis of the presence of an ideal balance or "sweet spot" between deficient and excessive fractality. Based on our results, it seems that the correlation between these two variables is positive and increasingly robust up to a certain point, and becomes negative and weaker from that point on wards. The fact that our correlations are never "very strong" (e.g. they are never higher than Pearson=0.5) appears to us as reassuring rather than discouraging. The fractality of a fairy tale's sentiment arc is not, naturally, the sole cause for its literary quality or popular appreciation, and a large number of factors are likely to influence the appreciation of a text - we would not expect very strong correlations from this one feature.

Naturally, our study's take on literary quality has a number of limitations. The average ratings of GoodReads have the substantial advantage of representing a large number of readers, who annotate a text in completely natural circumstances (e.g. they are not being paid or brought in a lab to do the annotation) over a relatively long span of time. At the same time, like all works using GoodReads scores or similar averages, we are consciously conflating quality with popularity, a perspective that might contrast with a prestige-driven view of quality, in which a smaller number of experts defines the quality of a text. Also, using the simple average of GoodReads reduces judgments that might be related to different aspects of the text to one

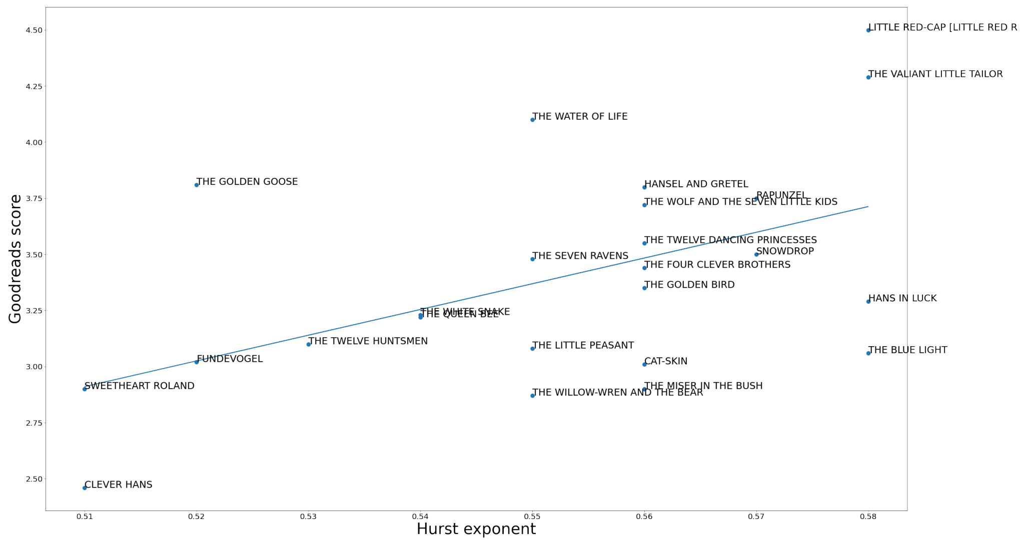

Figure 6: Closer look into the titles in the upward trend part of the plot. Pearson correlation=0.44, p-value=0.03.

single dimension. Although this problem is less severe when we consider a dataset of a single genre, from a small group of authors, it would become increasingly relevant if we were to apply this way of measuring quality to heterogeneous collections of texts.

Overall, our study confirms the perhaps surprising finding that quality (in terms of GoodReads' scores) and sentiment fractality hold a correlation for literary fairy tales, but adds the insight that such correlation might not be always linear: there could be a "sweet spot" for fractality's effects on readers. If we compare the study from Bizzoni, Peura, Nielbo, and Thomsen [7] with our own findings, we see that the most iconic titles from both collections fall in a similar interval for Hurst scores: *The Little Mermaid* or *The Ugly Duckling* for H.C. Andersen, *Red Riding Hood* or *Hansel and Gretel* for the Grimm brothers are between H=0.55 and H=0.60 approximately.

In future, we will try to test other quality "proxies" (prestigious awards, established canons) and to expand our feature set beyond sentiment arcs. It is reasonable to assume that different stylistic characteristics complement each other. It would also be interesting to attempt to predict the popularity of a text through automatic classification. Overall, sentiment fractality appears to be a promising feature to further our understanding of literary quality.

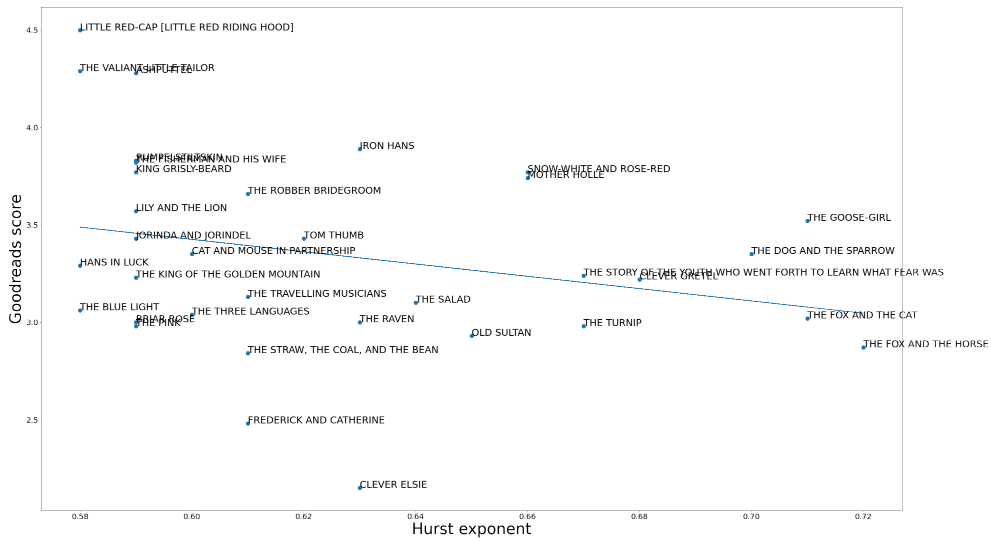

Figure 7: Closer look into the titles in the downwards trend part of the plot. Pearson correlation=-0.43, p-value=0.03.

References

[1] C. O. Alm. "Characteristics of high agreement affect annotation in text". In: Proceedings of the fourth linguistic annotation workshop. 2010, pp. 118–122.

[2] C. O. Alm, D. Roth, and R. Sproat. "Emotions from text: machine learning for text-based emotion prediction". In: *Proceedings of human language technology conference and conference on empirical methods in natural language processing*. 2005, pp. 579–586.

[3] E. C. O. Alm. *Affect in*text and speech*. University of Illinois at Urbana-Champaign, 2008.

[4] K. Bebbington, C. MacLeod, T. M. Ellison, and N. Fay. "The sky is falling: evidence of a negativity bias in the social transmission of information". In: *Evolution and Human Behavior* 38.1 (2017), pp. 92–101.

[5] J. Benesty, J. Chen, Y. Huang, and I. Cohen. "Pearson correlation coefficient". In: *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.

[6] J. Beran. *Statistics for Long-Memory Processes*. 1st ed. New York: Chapman and Hall/CRC, 1994.

[7] Y. Bizzoni, T. Peura, K. Nielbo, and M. Thomsen. "Fractal Sentiments and Fairy Tales-Fractal scaling of narrative arcs as predictor of the perceived quality of Andersen's fairy tales". In: *Journal of Data Mining & Digital Humanities* (2022).

[8] C. O. Brand, A. Acerbi, and A. Mesoudi. "Cultural evolution of emotional expression in 50 years of song lyrics". In: *Evolutionary Human Sciences* 1 (2019).

[9] J. Brooke, A. Hammond, and G. Hirst. "GutenTag: an NLP-driven tool for digital humanities research in the Project Gutenberg corpus". In: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. 2015, pp. 42–47.

[10] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco. "Affective computing and sentiment analysis". In: *A practical guide to sentiment analysis*. Springer, 2017, pp. 1–10.

[11] J. Cohen. "Quantitative methods in psychology: A power primer". In: *Psychological Bulletin* (1992), pp. 153–159.

[12] J. Cordeiro, P. R. M. Inácio, and D. A. B. Fernandes. "Fractal Beauty in Text". In: *Progress in Artificial Intelligence*. Ed. by F. Pereira, P. Machado, E. Costa, and A. Cardoso. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 796–802. DOI: 10.1007/978-3-319-23485-4_80.

[13] T. M. Cox. "Long range dependence: A review". In: *Iowa State University*. Press, 1984.

[14] A. van Cranenburgh and C. Koolen. "Identifying literary texts with bigrams". In: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. 2015, pp. 58–67.

[15] I.-A. Drobot. "Affective Narratology. The Emotional Structure of Stories". In: *Philologica Jassyensia* 9.2 (2013), p. 338.

[16] A. Eke, P. Herman, L. Kocsis, and L. R. Kozak. "Fractal characterization of complexity in temporal physiological signals". In: *Physiological Measurement* 23.1 (2002), R1. URL: http://stacks.iop.org/0967-3334/23/i=1/a=201.

[17] K. Eriksson and J. C. Coultas. "Corpses, maggots, poodles and rats: Emotional selection operating in three phases of cultural transmission of urban legends". In: *Journal of Cognition and Culture* 14.1-2 (2014), pp. 1–26.

[18] J. Gao, J. Hu, and W.-w. Tung. "Facilitating Joint Chaos and Fractal Analysis of Biosignals through Nonlinear Adaptive Filtering". In: *PLoS ONE* 6.9 (2011). Ed. by M. Perc, e24331. DOI: 10.1371/journal.pone.0024331. URL: http://dx.plos.org/10.1371/journal.pone.0024331.

[19] J. Gao, M. L. Jockers, J. Laudun, and T. Tangherlini. "A multiscale theory for the dynamical evolution of sentiment in novels". In: *2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC)*. Ieee. 2016, pp. 1–4.

[20] J. Grimm and W. Grimm. *Grimms' Fairy Tales*. Trans. by E. Taylor and M. Edwardes. Gutenberg Project, 2001.

[21] Q. Hu, B. Liu, M. R. Thomsen, J. Gao, and K. L. Nielbo. "Dynamic evolution of sentiments in Never Let Me Go: Insights from multifractal theory and its implications for literary analysis". In: *Digital Scholarship in the Humanities* 36.2 (2021), pp. 322–332.

[22] S. M. Islam, X. L. Dong, and G. de Melo. "Domain-specific sentiment lexicons induced from labeled documents". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2020, pp. 6576–6587.

[23] S. Jain, S. Malviya, R. Mishra, and U. S. Tiwary. "Sentiment Analysis: An Empirical Comparative Study of Various Machine Learning Approaches". In: *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*. Kolkata, India: NLP Association of India, 2017, pp. 112–121. URL: https://aclanthology.org/W17-7515.

[24] M. Jockers. *Syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text (Version 1.0. 1)*. 2017.

[25] E. Kim and R. Klinger. "A survey on sentiment and emotion analysis for computational literary studies". In: *arXiv preprint arXiv:1808.03137* (2018).

[26] K. Kousha, M. Thelwall, and M. Abdoli. "Goodreads reviews to assess the wider impacts of books". In: *Journal of the Association for Information Science and Technology* 68.8 (2017), pp. 2004–2016. DOI: 10.1002/asi.23805. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23805.

[27] X. Li, L. Bing, W. Zhang, and W. Lam. "Exploiting BERT for end-to-end aspect-based sentiment analysis". In: *arXiv preprint arXiv:1910.00883* (2019).

[28] S. Maharjan, J. Arevalo, M. Montes, F. A. González, and T. Solorio. "A multi-task approach to predict likability of books". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 2017, pp. 1217–1227.

[29] B. Mandelbrot. *The Fractal Geometry of Nature*. Updated ed. edition. San Francisco: Times Books, 1982.

[30] M. V. Mäntylä, D. Graziotin, and M. Kuutila. "The evolution of sentiment analysis–A review of research topics, venues, and top cited papers". In: *Computer Science Review* 27 (Feb. 1, 2018), pp. 16–32. DOI: 10.1016/j.cosrev.2017.10.002. URL: https://www.sciencedirect.com/science/article/pii/S1574013717300606.

[31] S. M. Mohammad and P. D. Turney. "Nrc emotion lexicon". In: *National Research Council, Canada* 2 (2013).

[32] S. M. Mohammad. "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words". In: *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*. Melbourne, Australia, 2018.

[33] M. Mohseni, V. Gast, and C. Redies. "Fractality and Variability in Canonical and Non-Canonical English Fiction and in Non-Fictional Texts". In: *Frontiers in Psychology* 12 (2021). URL: https://www.frontiersin.org/article/10.3389/fpsyg.2021.599063.

[34] F. Moretti. *Distant reading*. Verso Books, 2013.

[35] L. Myers and M. J. Sirois. "Spearman correlation coefficients, differences between". In: *Encyclopedia of statistical sciences* 12 (2004).

[36] 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.

[37] M. A. Riley, S. Bonnette, N. Kuznetsov, S. Wallot, and J. Gao. "A tutorial introduction to adaptive fractal analysis". In: *Frontiers in Physiology* 3 (2012). DOI: 10.3389/fphys.2012.00371. URL: http://journal.frontiersin.org/article/10.3389/fphys.2012.00371/abstract.

[38] B. Stroube. “Literary freedom: Project gutenberg”. In: *XRDS: Crossroads, The ACM Magazine for Students* 10.1 (2003), pp. 3–3.

[39] T. Underwood. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, Feb. 14, 2019. DOI: 10.7208/9780226612973. URL: https://www.degruyter.com/document/doi/10.7208/9780226612973/html.

[40] S. Veleskia. “Weak Negative Correlation between the Present Day Popularity and the Mean Emotional Valence of Late Victorian Novels”. In: *Proceedings http://ceur-ws. org ISSN* 1613 (2020), p. 0073.

[41] C. J. Walker and B. Blaine. “The virulence of dread rumors: A field experiment.” In: *Language & Communication* (1991).

[42] M. Walsh and M. Antoniak. “The Goodreads ‘Classics’: A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism”. In: *Journal of Cultural Analytics* 4 (2021), pp. 243–287.

[43] M. Wilkens. “Canons, close reading, and the evolution of method”. In: *Debates in the digital humanities* (2012), pp. 249–58.

[44] H. Xu, B. Liu, L. Shu, and P. S. Yu. “DomBERT: Domain-oriented language model for aspect-based sentiment analysis”. In: *arXiv preprint arXiv:2004.13816* (2020).