# Not just Plot(ting): A Comparison of Two Approaches for Understanding Narrative Text Dynamics*

Pascale Feldkamp Moreira[1], Yuri Bizzoni[1], Emily Öhman[2] and Kristoffer Nielbo[1]

[1]Center for Humanities Computing, Aarhus University, Aarhus, Denmark
[2]School of International Liberal Studies, Waseda University, Tokyo, Japan

### Abstract
This paper presents the outcomes of a study that leverages emotion annotation to investigate the narrative dynamics in novels. We use two lexicon-based models, VADER sentiment annotation and a novel annotation of 8 primary NRC emotions, comparing them in terms of overlaps and assessing the dynamics of the sentiment and emotional arcs resulting from these two approaches. Our results indicate that whereas the simple valence annotation does not capture the intricate nature of narrative emotions, the two types of narrative profiling exhibit evident correlations. Additionally, we manually annotate selected emotion arcs to comprehensively analyse the resource.

### Keywords
sentiment analysis, narrative emotions, fractal analysis, computational literary studies

## 1. Introduction

The complexity involved in modeling the sentiments and emotions "expressed" in a text poses significant challenges, and extensive efforts have been dedicated to developing linguistic frameworks and computational tools for detecting emotional and sentiment profiles of texts. These challenges are amplified in analysing *literary* texts, where the connotative and evocative dimensions of words at various narrative levels (plot, narrator, characters, style, etc.) play a role for the reading experience. Therefore, going beyond simple binary valences, as well as assessing the reliability of sentiment annotation against the narrative dynamics, is a particular challenge in the study of literary texts, where the interaction and strengths of two main approaches – sentiment and emotion annotation – are not sufficiently studied.

In this study, we examine a new resource of emotion arcs based on emotion intensity lexicons [36] that are enhanced by word embeddings and manual fine-tuning of word associations in a large corpus of novels, aiming to understand its overall reliability. We use an independent system, the VADER lexicon, to compare results and observe whether these methods converge in their overall profiling of the novels. Given that the methods represent different

*CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France*
*Corresponding author.
†These authors contributed equally.
pascale.moreira@cc.au.dk (P. F. Moreira); yuri.bizzoni@cc.au.dk (Y. Bizzoni); ohman@waseda.jp (E. Öhman); kln@cas.au.dk (K. Nielbo)
0000-0002-2434-4268 (P. F. Moreira); 0000-0002-6981-7903 (Y. Bizzoni); 0000-0003-1363-7361 (E. Öhman); 0000-0002-5116-5070 (K. Nielbo)

approaches – the first one-dimensional, rule- and sentence-based, and the second multidimensional, embedding- and paragraph-based – a convergence of their results on a large-scale literary corpus would help in gaining an understanding of both methods' reliability as more or less diverging methods for studying literary texts. A radical divergence, instead, could indicate that the complexity of literary language is too high for relatively simplistic analyses and that more sophisticated methods are needed.

## 2. Previous Works

To capture meaningful aspects of the reading experience, previous work has tested the potential of **sentiment analysis** [1, 19] at the word [34], sentence [32], or paragraph level [28], to model, i.a., sentiment arcs of novels [23, 42, 21]. Sentiment arcs have been used to evaluate literary texts in terms of shape or plot [42] and progression [16], as well as mood [39]. Moreover, certain arc dynamics have been connected to reader appreciation, considering both simple and complex narratives [3, 2], and Bizzoni, Moreira, Thomsen, and Nielbo [2] have shown that sentiment features have an effect even when compared to the stylistic features usually employed for this type of task [27, 30]. Studies usually draw positive or negative sentiment scores or valences of words or sentences via lexica [18] or machine learning approaches based on human annotations [35]. Several studies seek to develope suitable and specific methods to annotate sentiments and emotions for different domains [9, 37, 49].

Approaches to **emotion annotation** are predominantly based on the theory of universal emotions by Ekman [11], including Plutchik's wheel of emotions [41], and SenticNet [5], although recent studies have shown promise in expanding these models [8]. Studies of emotion in literary texts face challenges that inhere to emotion annotation, including the volatility and overlap of emotions as it is a task where there are large disagreements even between human annotators [38], with a lack of ground truth due to the subjective nature of emotions. Despite these inherent challenges, both Koljonen, Öhman, Ahonen, and Mattila [26] and Schmidt and Burghardt [46] show that emotion intensity (or polarity) is more congruent with human interpretations of affective content compared to a simple binary lexicon-based approach; and emotion annotation has been used to model literary genre [24], as well as reader appreciation, where Maharjan, Kar, Montes, González, and Solorio [31] have shown that the emotion flow in literary texts is connected to reader appreciation, indicating the potential of going beyond simple valences.

## 3. Data

### 3.1. The Chicago Corpus

We use the "Chicago Corpus", which spans 9,089 novels published in the US (1880-2000), and is a unique corpus both in terms of size,[1] and diversity. It was compiled based on the number of libraries holding each novel with preference for very circulated works. It is not homogeneous

[1]Studies on literary quality often rely on corpora of $< 1{,}000$ books [14, 27].

in terms of genre, as library holdings reflect a diverse demand across genre, and features both prestigious and popular works from Mystery to Science Fiction [29].[2]

### 3.2. Emotion Annotation

For annotation of emotion intensities, we use the NRC Affect Intensity Lexicon [36] of emotion labels, because it is an extensive emotion intensity lexicon (compared to other similar lexicons) that has been used and validated in various emotion detection tasks. The NRC lexicon was created based on human annotations and contains 9,829 lexemes with at least one emotion label, connected to a value between 0 and 1 to represent the intensity of the labeled emotion(s) calculated utilizing best-worst-scaling in the annotation process [25]. The emotions are *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, and *trust*. As this lexicon is not specific to the domain of literary texts, we used the novels in our dataset to create a semantic vector space model with Word2Vec [33]. We then expanded the lexicon by extracting emotions for lexemes that were not in the lexicon but had high cosine similarity values with lexemes in the lexicon, as well as various iterations of manual evaluation checking for unsubstantiated emotion associations of words [39, 44, 15]. This enabled us to create a fine-tuned domain-specific lexicon.

### 3.3. Sentiment Analysis

To annotate for valence, we chose a simple lexicon-based approach, using **VADER** at the sentence level [17], where each sentence is assigned compound score, ranging from negative (-1) to positive (1). We applied VADER because of its transparency, being based on a lexicon and a small set of rules. It is widely employed and shows good performance and consistency across domains [45, 43], which is beneficial when dealing with a corpus diverse in genre. Moreover, the origin of VADER in social media analysis does not appear to hinder annotation of literary texts [4]. Elkins and Chun [13] observe that the arc appears comparable to the **Syuzhet-package**, specifically developed for narrative [21]. However, using VADER side-steps some of the problems inherent to this package and to word-based annotation [47]. Dictionary-based methods seem to perform well even on so-called "nonlinear" narratives [**richardson_linearity_2000**, 13], and do not appear to perform worse than state-of-the-art Transformer-based approaches [12, 7].

## 4. Methods

We compare the emotion-based annotation of the novels with their simple annotation of VADER valence, using two central representations of the texts: their overall average emotional and sentimental intensity and the dynamics of emotion and sentiment arcs as estimated by the Hurst exponent (indicating arc persistence), as going beyond the sentiment or emotion intensities of novels to look at arc dynamics allows us to compare the similarities and strengths of each approach.

[2] Other studies have used the corpus [51, 6], see https://textual-optics-lab.uchicago.edu/us_novel_corpus.

## 4.1. Average Emotion

The initial inquiry focuses on the distribution of emotion intensities within the corpus and its relationship with the overall VADER valence distribution. By examining the average emotional intensity of the novels, we examine predominant tendencies of the corpus. Moreover, by comparing the distributions of the emotions as well as the correlation between mean emotion intensities and mean valence of novels, we see whether valence subsumes emotions. If there was no correlation between novels' mean values in emotions and mean valence, it may mean that emotion-based annotation and sentiment annotation capture different text facets. Similarly, a strong correlation between all emotion values and the valence of texts may indicate that emotion-based annotation does not contribute much beyond that which a valence-based approach captures.

## 4.2. Arc Dynamics

In addition to assessing the average emotion, we assess the dynamics of emotion and sentiment arcs in the narratives, which have recently appeared promising in modeling reader appreciation [2]. The **Hurst** exponent is a statistical measure that estimates the self-similarity of a time series, which has been proposed as an indicator of arc coherence [16]. In this particular study, we apply adaptive fractal analysis [**Gao2011**, 50] instead of the more commonly used detrended fluctuation analysis [40], due to the inherent noisiness and non-linearity of arcs. While the estimation of the Hurst exponent is beyond the scope of this paper, we use the following heuristic for arc coherence [16]. The range of the Hurst exponent $H$ for well-behaved time series is $0 \leq H \leq 1$. For $H \geq 0.5$, arcs are persistent such that increases are followed by increases and decreases by further decreases. For $H = 0.5$, arcs appear as white noise and are only characterized by short-range correlations; and for $H < 0.5$, arcs are anti-persistent and display mean-reverting behavior, that is, increases are followed by decreases and decreases by increases. In terms of arc coherence, persistent story arcs appear as more coherent narratives, where emotional intensity develops at longer time scales. Story arcs that only display short-range correlations lack coherence, while anti-persistent story arcs will oscillate around an average and undifferentiated emotional state [16].

We observe the distribution of the level of persistence in emotion and sentiment arcs using both emotion-based and sentiment-based annotations. If the two resources returned radically different Hurst exponents, it would mean that the patterns elicited by a simple analysis for valence are very different from those elicited by emotion analysis. In other words, the "composition" of all the emotions in one single dimension gives way to dynamics that are different from the patterns of any individual emotion. In contrast, overlap between the Hurst distribution of the sentiment arcs and that of the emotion arcs would indicate comparability of the two annotations. It could give insight into which emotions are drawing the overall sentiment Hurst of the corpus towards higher versus lower exponents.

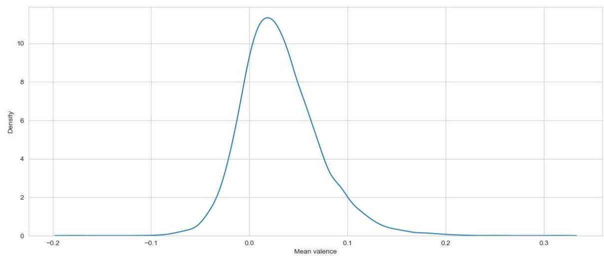


**Figure 1:** Distribution of VADER valences in the Chicago corpus, which range from -1 to 1.

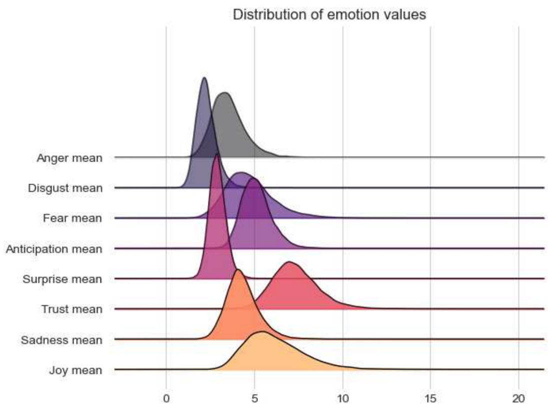


**Figure 2:** Distribution of emotion intensities in the Chicago corpus, ranging from 0 to 20.

**Table 1**
Mean and standard deviation of emotion values in our corpus, their correlation with valence, and mean and standard deviation of Hurst per emotion arc. While mean valence ranges from -1 to 1, emotion intensities are attributed per word from 0-1, then summed into 300-word bins. The averages of each title (average of all bins), lie in the interval 0–20 in our corpus. In the table, the highest value in each column is in bold.

| Emotion | Mean | Std. deviation | Correlation w. valence | Hurst mean | Hurst std. deviation |
|---|---|---|---|---|---|
| Valence | 0.031 | 0.039 | x | 0.608 | 0.037 |
| Joy | 6.126 | **1.598** | **0.625** | 0.655 | 0.081 |
| Trust | **7.352** | 1.302 | 0.470 | 0.636 | 0.079 |
| Anger | 3.536 | 0.921 | -0.548 | 0.644 | 0.084 |
| Fear | 4.737 | 1.321 | -0.563 | **0.665** | **0.086** |
| Sadness | 4.336 | 0.907 | -0.379 | 0.646 | 0.079 |
| Disgust | 2.278 | 0.559 | -0.427 | 0.610 | 0.080 |
| Surprise | 2.927 | 0.474 | 0.070 | 0.573 | 0.073 |
| Anticipation | 5.124 | 0.824 | 0.465 | 0.601 | 0.072 |

## 5. Results

### 5.1. Emotional Intensities

As ranges of VADER's valences and emotion intensities are not the same (-1 to 1 vs. 0 to 20) we cannot directly compare the two sets of distributions, but observe their behavior considering their own means.

First, applying VADER uncovers a subtle positivity-bias in our corpus, into which the distribution of specific emotions provides further insight. The emotions *trust* and *anticipation* might contribute to the right-skewed distribution, indicating a prevalence of positive emotional expressions, while joy appears to have a long right tail. Positive emotions may pull the skew towards higher values, which aligns with the prominence of positive valence in our corpus. Moreover, negative emotions tend to cluster towards moderately lower intensity levels, and the mean of emotions like disgust and anger have much lower mean values than more positive emotions like joy and trust (Table 1).The positive skew may be related to literary texts having high positive emotional content. Yet, it is essential to acknowledge that linguistic factors may influence this bias, as languages can exhibit inherent positivity [10]. Such biases can stem from cultural norms, semantic sectioning of the world, etc. Note, however, the high standard deviation in joy and trust, indicating that books in our corpus vary in terms of these emotions (Table 1).

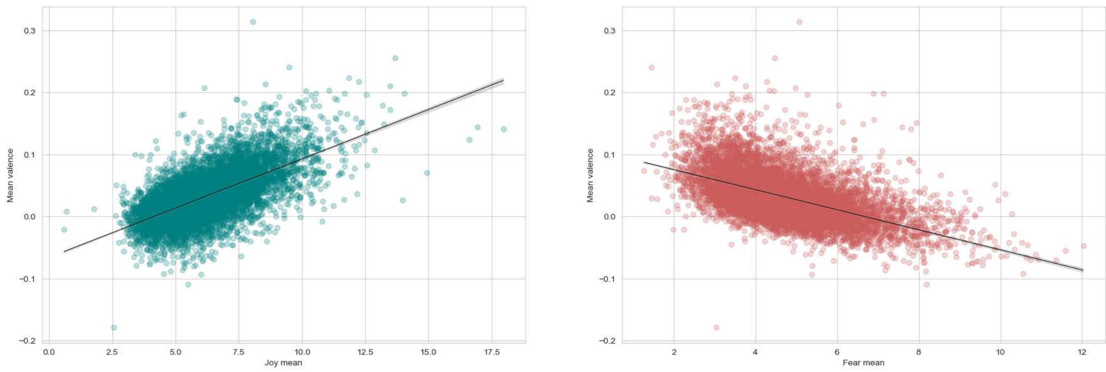


**Figure 3:** From left to right: 1) correlation of mean joy and mean valence, 2) correlation of mean fear and mean valence.

When directly correlating emotions with valence, we find the most significant negative correlation between valence and fear. This correlation suggests that VADER assigns lower sentiment scores to novels with higher intensities of fear, aligning with the expectation that fear, a negative emotion, would exhibit a stronger negative correlation with the overall sentiment analysis (Fig. 3). The opposite is true for joy, and most other emotions correlate with the corpus' overall valence in an expected way: trust and anticipation are positively correlated with valence, while anger, disgust, and sadness are negatively correlated (Fig. 4). These findings indicate a convergence between the sentiment analysis and emotion analysis. However, certain complex emotions, such as anticipation, trust and surprise, exhibit lower correlations with

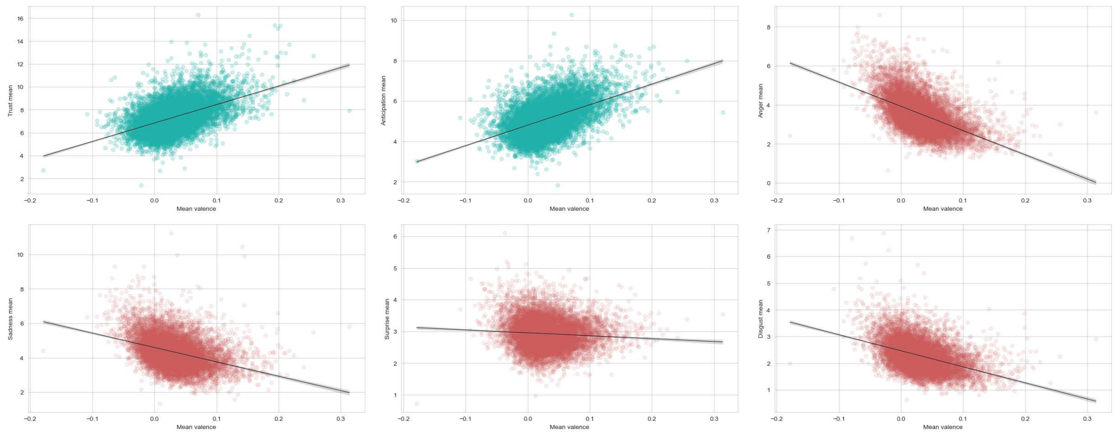

**Figure 4:** Correlation of remaining mean emotion values with mean valence.

valence (Table 1), suggesting divergence from what is captured the VADER annotation. These emotions are less intuitively and clear-cut positive or negative, which may contribute to the weaker correlation, why they may not be adequately captured via valences.

## 5.2. Emotion Arc Dynamics

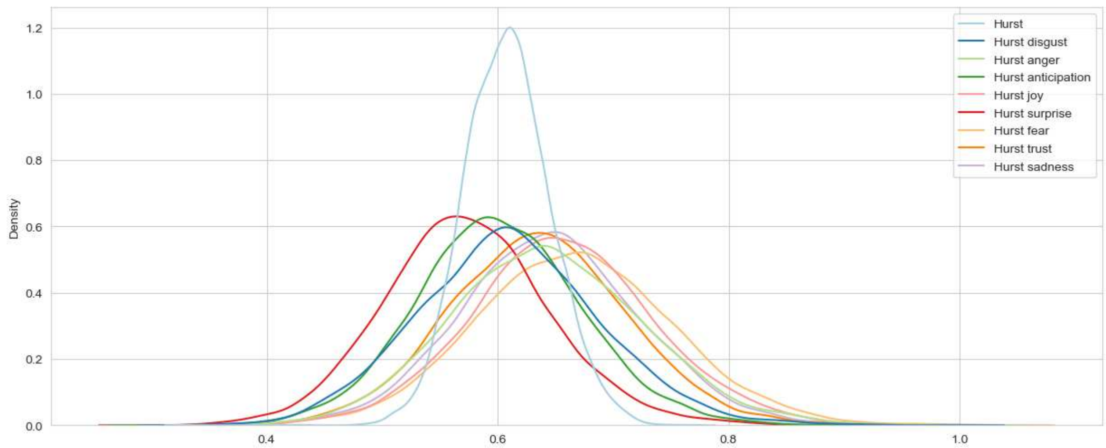


**Figure 5:** Distribution of Hurst exponents in our corpus, based on valence (light blue) and emotions. While Hurst based on valence peaks at 0.62, distributions of Hurst based on the emotions are more spread out, with surprise on the low and joy on the high end.

The analysis of the Hurst exponent of sentiment and emotion arcs reveals clusters in similar areas, suggesting a degree of interrelation between the Hurst exponent of sentiment and emotion arcs (Fig. 5). The overlap indicates that using emotion profiling enables a nuanced

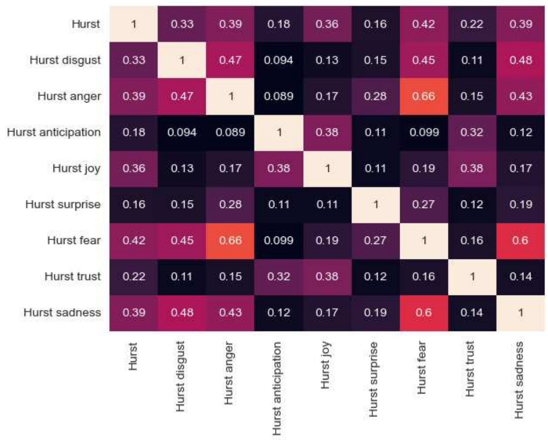


**Figure 6:** Correlation (Spearman's) between the Hurst exponent of sentiment and emotion arcs of novels.

understanding of the development of different emotional tones within the narratives, where VADER provides a more generalized and less transparent representation of the novels' "prismatic" internal dynamics. Looking at the distribution of Hurst exponent of titles in our corpus, we find that, tendentiously, surprise, anticipation and disgust cluster slightly below or at the distribution of the Hurst based on valence ("Hurst" in Fig. 5), while the remaining emotions cluster slightly above it. All distributions of Hurst based on emotion arcs are significantly smoother with longer tails than that based on valence.

Specifically, we see a tendency of the Hurst based on surprise being slightly lower and exhibiting a slightly different distribution than, e.g., fear and joy, at an average 0.57 (Table 1, Fig. 5). Note that the standard deviation is high (0.07), so that this difference should be regarded a tendency only, but may align with our intuition that it is easier to envision more progressive and linear increases or decreases in, e.g., joy than in surprise values – an emotion that may peak "surprisingly".

In sum, the dynamics of each individual emotion arc may offer a complex picture of a novel's progression, while the Hurst exponent of valence-based arcs offers a general outline that is more correlated to the Hurst of more clear-cut emotions like fear and joy (Fig. 6).

## 5.3. Manual Annotations of Emotion Arcs

To further assess the reliability of our emotion-based annotation, we inspected arcs of novels that showed strong values of Hurst: *The Old Man and the Sea* by Ernest Hemingway, which has one of the lowest Hurst exponents for joy in our corpus, as well as *A Portrait of the Artist as a Young Man* by James Joyce, which has one of the highest Hurst exponents for fear in our corpus. Fig. 7 shows our manual annotation of the correspondence of narrative events with emotion arcs in Hemingway's *The Old Man and the Sea.* Note that peaks in fear and joy seem to co-occur in this novel. While this may appear puzzling, our inspection confirms that this co-occurence of positive and negative emotions actually illustrates a central characteristic of Hemingway's prose style and the story overall: even in moments of crisis, Hemingway's protagonist continues to reflect on his love for the sea and on natural beauty, leading to complex feelings and contradictory emotional intensities in key scenes – love and hatred, fear and admiration (see, i.a., box 7 in Fig. 7). Such complexity is also reflected in the protagonist's character: his hardships and endurance, but essentially optimistic outlook on life.

Emotion arcs in James Joyce's *A Portrait of the Artist as a Young Man* parallel Jockers [22] valence-based arc of the same novel through the Syuzet package and human annotation, which has been called a "man in the hole" shape, with one central crisis.[3]

Here, only predominantly negative emotions are elevated in the main rise, and intensities of joy and fear do not co-occur as in Hemingway's prose (cf. Fig. 7). The more independently developing arcs are reflected in more varied Hurst exponents in the *Portrait*, which is among the top 50 books in our corpus with the highest standard deviation between their emotion arcs' Hurst exponents. In the *Portrait*, the Hurst of the negative emotions anger, fear, and sadness is > than 0.9, while the Hurst of anticipation, joy, and trust hovers around 0.8. The three negative emotions exhibit a clear and steady rise and decline around the central crisis. As the Hurst exponent measures persistence, i.e., whether increases are followed by increases or decreases by decreases, a very high Hurst exponent here adequately indicates the slow rise and decline of negative emotions. Arcs of positive emotions are less persistent but still exhibit more persistence than, for example, the arc of surprise, which here has a Hurst exponent of 0.58 and appears mean-reverting (cf. pink line in Fig. 8). Overall, the Hurst based on valence for the *Portrait* is 0.71, a value that may represent the average dynamics of various emotion arcs, and which does not capture the subtle but distinct difference between trends in positive and negative emotion arcs in the novel.

[3]Cf. Note that emotion arcs do not appear to show the ringing artifacts, the artificial positive trend in the beginning in Jockers [20], connected to Syuzet's low-pass filter (cf. Swafford [48]).

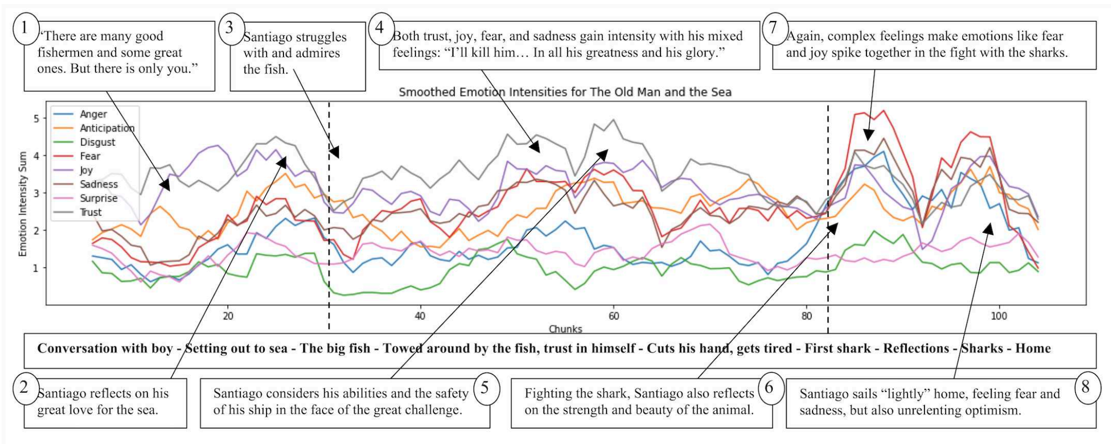


**Figure 7:** Emotion arcs of Hemingway's *The Old Man and the Sea*.

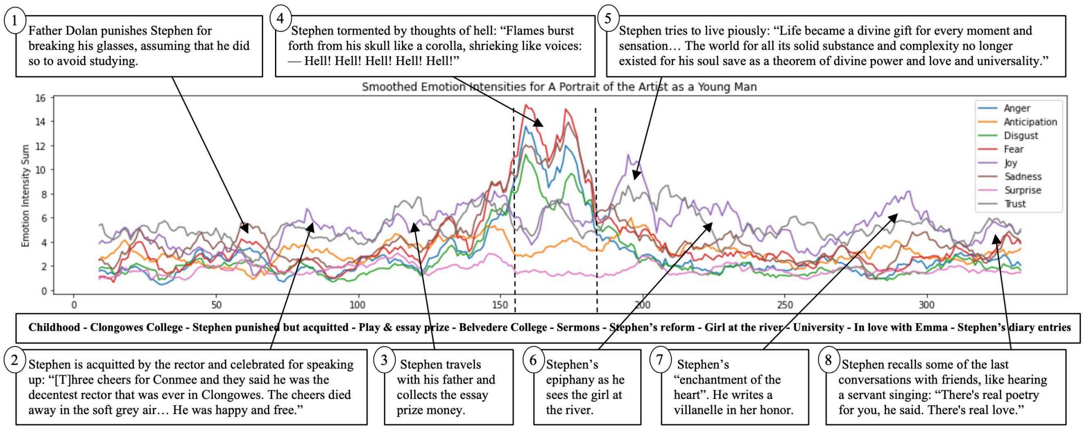


**Figure 8:** Emotion arcs in Joyce’s *A Portrait of the Artist as a Young Man*.

## 6. Conclusion and Future Works

Our analysis of the distribution of emotion intensities and Hurst exponents based on emotion- and sentiment arcs suggested that more valences subsume diverse emotional interactions. Some emotions are expectedly correlated to valence, while less clear-cut emotions like surprise seem to be less captured by valence annotation, and the Hurst of their arcs less correlated to that of the valence-based arc. Moreover, some emotion arcs, like surprise, are on average less persistent, and some, like fear, more persistent than valence-based arcs, which suggests that we may get a more nuanced understand the internal dynamics of novels, including progressions of emotions less clearly positive or negative, by analysing the Hurst of various emotion arcs.

Our inspection of individual titles suggested that the emotions and sentiments expressed in the text are *not* such that are explicitly felt by characters, nor transparently transmitted to readers. Rather, they are emotional textures narratives, from which the readers may derive complex (reading) experiences. The the co-occurrence of peaks in emotion arcs as seen in the case of Hemingway, as well as the difference between the Hurst exponents of emotion arcs as seen in the case of Joyce is not trivial, as it tells us something important about plotting narrative arcs in general: namely, that **sentiment valence** does not stand in direct relation to plot and narrative events, but rather subsumes trends in emotion evocation that pertain to both style and events. In other words, when plotting arcs based on emotions we are observing trends in a narrative event-style continuum, that is less well captured by valence annotation. In the future, we suggest studying the dynamics of emotion- and sentiment-based arcs closely, seeking to assess arc dynamics at both a local and global level, as well as linking represented emotion to the actual reader experience and appreciation.

## References

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

[2] Y. Bizzoni, P. Moreira, M. R. Thomsen, and K. Nielbo. "Sentimental Matters - Predicting Literary Quality by Sentiment Analysis and Stylometric Features". In: *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*. Toronto, Canada: Association for Computational Linguistics, 2023, pp. 11–18. URL: https://aclanthology.org/2023.wassa-1.2.

[3] 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* Nlp4dh (2022). DOI: 10.46298/jdmdh.9154.

[4] Y. Bizzoni, T. Peura, K. Nielbo, and M. Thomsen. "Fractality of sentiment arcs for literary quality assessment: The case of Nobel laureates". In: *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*. Taipei, Taiwan: Association for Computational Linguistics, 2022, pp. 31–41. URL: https://aclanthology.org/2022.nlp4dh-1.5.

[5] E. Cambria, A. Livingstone, and A. Hussain. "The hourglass of emotions". In: *Cognitive behavioural systems*. Springer, 2012, pp. 144–157.

[6] J. Cheng. "Fleshing out models of gender in English-language novels (1850–2000)". In: *Journal of Cultural Analytics* 5.1 (2020), p. 11652. DOI: 10.22148/001c.11652.

[7] J. Chun. *SentimentArcs: A Novel Method for Self-Supervised Sentiment Analysis of Time Series Shows SOTA Transformers Can Struggle Finding Narrative Arcs*. 2021. URL: http://arxiv.org/abs/2110.09454.

[8] A. S. Cowen and D. Keltner. "Semantic space theory: A computational approach to emotion". In: *Trends in Cognitive Sciences* 25.2 (2021), pp. 124–136. DOI: 10.1016/j.tics.2020.11.004.

[9] N. Z. Da. "The computational case against computational literary studies". In: *Critical inquiry* 45.3 (2019), pp. 601–639. DOI: 10.1086/702594.

[10] P. S. Dodds, E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth. "Human language reveals a universal positivity bias". In: *Proceedings of the National Academy of Sciences* 112.8 (2015), pp. 2389–2394. DOI: 10.1073/pnas.1411678112.

[11] P. Ekman. "Universals and cultural differences in facial expressions of emotion." In: *Nebraska symposium on motivation*. University of Nebraska Press. 1971.

[12] K. Elkins. *The Shapes of Stories: Sentiment Analysis for Narrative*. Cambridge University Press, 2022. DOI: 10.1017/9781009270403. URL: https://www.cambridge.org/core/product/identifier/9781009270403/type/element.

[13] K. Elkins and J. Chun. "Can Sentiment Analysis Reveal Structure in a Plotless Novel?" In: (2019). URL: https://arxiv.org/abs/1910.01441.

[14] V. Ganjigunte Ashok, S. Feng, and Y. Choi. "Success with Style: Using Writing Style to Predict the Success of Novels". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, 2013, pp. 1753–1764. URL: https://aclanthology.org/D13-1181.

[15] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas. "Sentiment analysis leveraging emotions and word embeddings". In: *Expert Systems with Applications* 69 (2017), pp. 214–224. DOI: 10.1016/j.eswa.2016.10.043.

[16] 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 (2020), pp. 322–332. DOI: 10.1093/llc/fqz092.

[17] C. Hutto and E. Gilbert. "VADER: A parsimonious rule-based model for sentiment analysis of social media text". In: *Proceedings of the international AAAI conference on web and social media*. Vol. 8. 1. 2014, pp. 216–225. DOI: 10.1609/icwsm.v8i1.14550.

[18] S. M. Islam, X. Dong, and G. de Melo. "Domain-Specific Sentiment Lexicons Induced from Labeled Documents". In: *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020, pp. 6576–6587. URL: https://aclanthology.org/2020.coling-main.578.

[19] 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.

[20] M. Jockers. *Revealing Sentiment and Plot Arcs with the Syuzhet Package*. 2015. URL: https://www.matthewjockers.net/2015/02/02/syuzhet/.

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

[22] M. Jockers. *That Sentimental Feeling*. 2015. URL: https://www.matthewjockers.net/2015/12/20/that-sentimental-feeling/.

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

[24] E. Kim, S. Padó, and R. Klinger. "Investigating the Relationship between Literary Genres and Emotional Plot Development". In: *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Vancouver, Canada: Association for Computational Linguistics, 2017, pp. 17–26. URL: https://aclanthology.org/W17-2203.

[25] S. Kiritchenko and S. M. Mohammad. "Capturing reliable fine-grained sentiment associations by crowdsourcing and best–worst scaling". In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics, 2016, pp. 811–817. URL: https://www.aclweb.org/anthology/N16-1095.

[26] J. Koljonen, E. Öhman, P. Ahonen, and M. Mattila. "Strategic sentiments and emotions in post-Second World War party manifestos in Finland". In: *Journal of computational social science* (2022), pp. 1–26.

[27] C. Koolen, K. van Dalen-Oskam, A. v. Cranenburgh, and E. Nagelhout. "Literary Quality in the Eye of the Dutch Reader: The National Reader Survey". In: *Poetics* 79 (2020), pp. 1–13. DOI: 10.1016/j.poetic.2020.101439.

[28] X. Li, L. Bing, W. Zhang, and W. Lam. "Exploiting BERT for End-to-End Aspect-based Sentiment Analysis". In: (2019), pp. 34–41. URL: https://aclanthology.org/D19-5505.

[29] H. Long and T. Roland. *US Novel Corpus*. Tech. rep. Textual Optic Labs, University of Chicago, 2016. URL: http://icame.uib.no/brown/bcm.html.

[30] 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*. Valencia, Spain: Association for Computational Linguistics, 2017, pp. 1217–1227. URL: https://aclanthology.org/E17-1114.

[31] S. Maharjan, S. Kar, M. Montes, F. A. González, and T. Solorio. "Letting Emotions Flow: Success Prediction by Modeling the Flow of Emotions in Books". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Volume 2, Short Papers*. New Orleans, Louisiana: Association for Computational Linguistics, 2018, pp. 259–265. URL: https://aclanthology.org/N18-2042.

[32] 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.

[33] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin. "Advances in Pre-Training Distributed Word Representations". In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), 2018. URL: https://aclanthology.org/L18-1008.

[34] S. Mohammad. "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 174–184. URL: https://aclanthology.org/P18-1017.

[35] S. M. Mohammad and P. D. Turney. "NRC emotion lexicon". In: *National Research Council, Canada* 2 (2013), p. 334.

[36] S. M. Mohammad. "Word Affect Intensities". In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), 2018, pp. 174–183. URL: https://aclanthology.org/L18-1027.

[37] E. Öhman. “The Validity of Lexicon-based Sentiment Analysis in Interdisciplinary Research”. In: *Proceedings of the Workshop on Natural Language Processing for Digital Humanities.* NIT Silchar, India: NLP Association of India (NLPAI), 2021, pp. 7–12. URL: https://aclanthology.org/2021.nlp4dh-1.2.

[38] E. Öhman, M. Pàmies, K. Kajava, and J. Tiedemann. “XED: A Multilingual Dataset for Sentiment Analysis and Emotion Detection”. In: *Proceedings of the 28th International Conference on Computational Linguistics.* Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020, pp. 6542–6552. URL: https://aclanthology.org/2020.coling-main.575.

[39] E. Öhman and R. H. Rossi. “Computational Exploration of the Origin of Mood in Literary Texts”. In: (2022), pp. 8–14. URL: https://aclanthology.org/2022.nlp4dh-1.2.

[40] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger. “Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 5.1 (1995), pp. 82–87. DOI: 10.1063/1.166141.

[41] R. Plutchik. “A general psychoevolutionary theory of emotion”. In: *Theories of emotion.* Elsevier, 1980, pp. 3–33.

[42] 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), p. 31. DOI: 10.1140/epjds/s13688-016-0093-1.

[43] A. J. Reagan, B. Tivnan, J. R. Williams, C. M. Danforth, and P. S. Dodds. “Benchmarking Sentiment Analysis Methods for Large-Scale Texts: A Case for Using Continuum-Scored Words and Word Shift Graphs”. In: arXiv:1512.00531 (2016). URL: http://arxiv.org/abs/1512.00531.

[44] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi. “Sentiment analysis based on improved pre-trained word embeddings”. In: *Expert Systems with Applications* 117 (2019), pp. 139–147. DOI: 10.1016/j.eswa.2018.08.044.

[45] F. N. Ribeiro, M. Araújo, P. Gonçalves, M. André Gonçalves, and F. Benevenuto. “SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods”. In: *EPJ Data Science* 5.1 (2016), pp. 1–29. DOI: 10.1140/epjds/s13688-016-0085-1.

[46] T. Schmidt and M. Burghardt. “An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing”. In: *Coling 2018.* Association for Computational Linguistics. 2018, p. 139.

[47] A. Swafford. *Problems with the Syuzhet Package.* 2015. URL: https://annieswafford.wordpress.com/2015/03/02/syuzhet/.

[48] A. Swafford. “Problems with the Syuzhet Package”. In: *Anglophile in Academia: Annie Swafford’s Blog* (2015). URL: https://annieswafford.wordpress.com/2015/03/02/syuzhet/.

[49] D. Teodorescu and S. M. Mohammad. “Generating High-Quality Emotion Arcs For Low-Resource Languages Using Emotion Lexicons”. In: *arXiv preprint arXiv:2306.02213* (2023).

[50] W.-w. Tung, J. Gao, J. Hu, and L. Yang. “Detecting chaos in heavy-noise environments”. In: *Physical Review E* 83.4 (2011). DOI: 10.1103/PhysRevE.83.046210.

[51] T. Underwood, D. Bamman, and S. Lee. “The transformation of gender in English-language fiction”. In: *Journal of Cultural Analytics* 3.2 (2018), p. 11035. DOI: 10.22148/16.019.