Exploring Sentiment Analysis for Plot Extraction in Literary Studies: Methodologies, Replication, and Critical Evaluation

Pouli, Jim

Exploring Sentiment Analysis for Plot Extraction in Literary Studies:
Methodologies, Replication, and Critical Evaluation

Jim Pouli
M.A. Book and Digital Media Studies
Faculty of Humanities
First reader: Dr. Peter Verhaar
Second reader: Dr. F.E.W. Praal
Word count: 27,407
Submission date: 20 June 2023 Netherlands

Abstract

Sentiment analysis is the method of computationally recognising opinions stated in a piece of text, particularly to identify if the writer has a positive, negative, or neutral attitude towards a given topic. Although sentiment analysis is commonly used to analyse short texts on social media platforms, its application in literary research has gained traction in recent years. The emergence of sentiment analysis tools, such as Syuzhet, has notably expanded the research possibilities in this field. However, despite these advancements, there remains a need to further explore and understand the strengths and limitations of sentiment analysis in the context of literary analysis.

The aim of this thesis is to further the body of knowledge about the use of sentiment analysis as a technique for plot extraction. Throughout this thesis, I experiment with a modified form of the social media analysis tool VADER. While this method proofs to work as an effective tool for extracting sentiment from linear stories, it still shows limitation on a sentence-to-sentence basis. Moreover, I use this tool to replicate an often-cited study by Reagan et al., where it was stated that the majority of stories can be categorized in six basic 'plot shapes'. I argue that while most of these shapes can still be identified using an alternative sentiment analysis technique, this technique occasionally classifies a story into a different shape than Reagan et al.'s analysis did.

I conclude by giving a critical evaluation of sentiment analysis as a tool for plot extraction. Since the 'plot' of a story is a multifaceted concept, we cannot simply argue that a sentiment analysis graph displays the progression of plot. Nonetheless, modifications can be made to get a fuller understanding of the narrative of a story.

Table of contents

1. Introduction

1.1 Background

The digital humanities may have begun as a niche subfield within the humanities, but the discipline has since come a long way. These days the boundaries of the digital humanities field are as distinct as those of many other disciplines.[1] The field's primary focus is the use of digital resources, tools, and approaches in humanities studies. In using such tools, it reflects a more exploratory and less qualitative approach than, let's say, social sciences, but it also represents novel attempts to model complex human wisdom. Among these attempts is the analytical practice known as 'sentiment analysis'.

Sentiment analysis is the method of computationally recognizing opinions and emotions expressed in a text, particularly to understand if the writer has a positive, negative, or neutral viewpoint. Thanks to the growth of internet platforms and social media, we now have a colossal and ever-expanding selection of opinions in the form of digital text. This data means new and interesting possibilities for research. For example, researchers have used data from Twitter and YouTube to explore public opinions on the Covid-19 vaccination.[2] But apart from academia, the results from such analyses have found their way to news outlets and in corporate practices as well. Online newspapers write pieces during an election-period to demonstrate that negativity is the key to a successful tweet.[3] Companies have applied sentiment analysis to determine public opinion on their products based on a selection of consumer reviews.[4] A popular tool in extracting such information is called VADER and is part of the Natural Language Toolkit written in the Python programming language.

Another relatively new application of sentiment analysis has emerged in narratology and literary studies, and focuses on analysing the sentiments associated with stories. Whereas the examples mentioned above mostly focus on short pieces of text (tweets, comments, reviews etc.), literary researchers focus on the progression of sentiment throughout a longer text, e.g. a novel or a play. This method offers a computational approach to a long-explored

---

[1] J. Luhmann and M. Burghardt, 'Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape', *Journal of the Association for Information Science and Technology*, *73* (2022), pp.148-171.

[2] J. Al-Garaady and M. Mahyoob, 'Public sentiment analysis in social media on the SARS-CoV-2 vaccination using VADER lexicon polarity', *Humanities and Educational Sciences Journal* (2022), pp.591-609.

[3] N. Grover, 'Just say no: negativity is secret of political tweet success, study finds', *The Guardian*, 13 April, 2021. <https://www.theguardian.com/technology/2021/apr/13/just-say-no-negativity-is-secret-of-political-tweet-success-study-finds>

[4] L.D. Ma, 'Evaluation of product conceptual design based on Pythagorean fuzzy set under big data environment', *Scientific Reports*, *12* (2022), pp.1-20.

distant reading notion in narratology; that all stories can be categorised into a fixed set of plot arcs.[5] Ever since Aristotle described the basic triangle-shaped plot structure, there have been countless scholarly and non-scholarly discussions regarding the idea that there might be a limited number of fundamental or archetypal story structures. Inspired by this debate, Reagan et al. used the sentiment analysis tool *Hedonometer* for a large-scale study. By analysing over 1,300 texts, Reagan et al. argued that most stories can indeed be categorized under a fixed set of 'plot shapes': a set of six core emotional arcs which form the essential building blocks of complex emotional trajectories.[6]

Similarly motivated by the debate about plot structures, Matthew Jockers created the sentiment analysis package Syuzhet in 2015. Jockers' goal with Syuzhet, as explained in a post on his personal blog, was to extract sentiment and plot information from prose, resulting in a model of the plot arc of any given story.[7] Whereas VADER and the Hedonometer find their origins in the analysis of social media texts, Syuzhet was designed with the specific application of story analysis in mind. Syuzhet quickly became the most downloaded sentiment analysis package in R, demonstrating Jockers' work to be an intriguing contribution to an already active and expanding interest in narratology research in the digital humanities.[8]

Nevertheless, Syuhet has also been the object of fierce criticism since then. Around the time Syuzhet was released in 2015, Annie Swafford criticized the tool for a number of factors.[9] Syuzhet had difficulties recognizing satire and sarcasm, and Swafford believes that the smoothing functions (like the Fourier transform) generate a degree of distortion in the results. Moreover, the lexicon used by Syuzhet seems to be less successful when analysing a text that is not written in present-day English. Numerous updates have been made since then to address some of these concerns. Hoyeol Kim noted in 2022, however, that, despite numerous updates and the popularity of this sentiment analysis tool, many of Swafford's criticisms are still valid today.[10]

---

[5] S. Onega and J.A.G. Landa, *Narratology: an introduction* (Abingdon, Oxfordshire: Routledge, 2014), p. 4.
[6] 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', *EPJ Data Science*, 5 (2016), pp.1-12.
[7] M. Jockers, 'Revealing Sentiment and Plot Arcs with the Syuzhet Package', *matthewjockers.net*, 2 February, 2015. <https://www.matthewjockers.net/2015/02/02/syuzhet/> (12 September, 2022).
[8] H. Kim, 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.' *DHQ*, 16 (2022), n.pag.
[9] A. Swafford, 'Why Syuzhet Doesn't Work and How We Know', *Anglophile in Academia: Annie Swafford's Blog*, March 30, 2015. <https://annieswafford.wordpress.com/2015/03/30/why-syuzhet-doesnt-work-and-how-we-know/> (5 September, 2022).
[10] Kim, H., 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.' *DHQ*, 16 (2022), n.pag.

Reagan et al.'s findings on the fixed plot shapes continue to motivate similar research in the field of the digital humanities as well. A recent study by Boyd, Blackburn and Pennebaker, for example, explored if it was possible to reveal other narrative structures, such as staging, through a similar methods.[11] However, Kim has also shown in his critical assessment of Syuzhet that differences between sentiment analysis tools, such as the lexicon and the analysis method, can cause significant dissimilarities between the eventual results, and can mean the difference between a happy ending and a tragic one. Furthermore, Jockers' work also still motivates narratology research, as well as research beyond literature. Nonetheless, it is important to remember that sentiment analysis research, specifically on literary texts, is still a relatively new research practice, and is still far from being a perfected method. There are still uncertainties around this research approach, as Swafford and Kim have indicated.

Many questions on the use of sentiment analysis tools in narratology and the digital humanities are worth exploring. Do sentiment analysis tools provide sufficient data on the progression of sentiment throughout a story, despite their shortcomings? Do different tools, with different lexicons and smoothing methods, provide similar results for the same texts? And ultimately, do these results actually give us useful information about the plot of a story?

1.2 Research objectives and research questions

The aim of this thesis is to further the body of knowledge about the use of sentiment analysis as a technique for plot extraction. While there is more and more research on the progression of sentiment throughout a story, it is still a relatively new research practice. It is worth to revaluate, and where possible improve, upon the research by Jockers and by Reagan et al. The main research question of this thesis, which corresponds to its main objective, is as follows: 'How can sentiment analysis be effectively utilized for plot extraction in literary studies, and what improvements can be made to existing methods?'

This thesis will be subdivided into three main objectives, each with its corresponding subquestion, in order to answer the main question. The first objective is to develop a sentiment analysis tool that works similarly to Syuzhet. When it comes to literary analyses for NLP research in R, Syuzhet is currently the most popular tool to perform sentiment analysis.[12]

---

[11] R.L. Boyd, K.G. Blackburn and J.W. Pennebaker 'The narrative arc: Revealing core narrative structures through text analysis', *Science advances*, *6* (2022), n.pag.
[12] H. Kim, 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.' *DHQ*, 16 (2022), n.pag.

However, this tool only functions with R as a programming language. Three factors led to the choice to experiment with Python rather than R as the programming language for this thesis. The first of which is to expand the possibilities of computational narratology research to a different programming language, with the intent of creating a method similar and as effective as Syuzhet. Second, by altering the analytical tool VADER to function similarly to the sentiment analyses tool implemented in Syuzhet, it is possible to test if two of its core characteristics—its lexicon and its analysis method—will also hold true in the context of a literary study, and might even improve upon the original method. Finally, developing a sentiment analysis program in Python offers the advantage of transparency and customization. By creating our own software, we gain direct access to its methods, providing a clear opportunity to critique the methodology used in existing tools like Syuzhet. This transparency allows for modifications to be made, enabling experimentation with enhanced techniques for analysing the plot of a story. These three factors will be elaborated upon further in the methodology chapter (chapter 3). The subquestion corresponding to this objective is as follows: 'How can a sentiment analysis tool be developed in Python that functions similarly to Syuzhet and effectively contributes to computational narratology research?'

The second objective of this thesis is to try and replicate the results found by Reagan et al. in 2016. Reagan et al. used the Hedonometer, a sentiment analysis programme intended for the analysis of social media text, in a novel way by analysing the progression of sentiment throughout a large corpus of texts. What they found was that a large number of these texts can be classified into six 'basic shapes'. Researchers are still citing the results from these findings to this day. For example, in *Stories in Action*, Walsh et al. explore the influence of stories on policy-making, arguing for their impacts on learning and persuasion.[13] They cite Reagan et al.'s results and state that, despite the need for uncertainty to keep the audience engaged and interested, there is also a simultaneous desire for predictability, as seen in the prevalence of specific emotional arcs in storytelling. Dale et al. were also inspired by Reagan et al.'s research and attempted to uncover these emotional arc structures in films.[14] However, while these results are still being cited, the Hedonometer is merely one of many sentiment analysis tools. Would these same results be found when using a different analysis tool? One of the most important ways to increase the trust in the validity of these findings is through

[13] J. Walsh, N. Vaida, A. Coman and S.T. Fiske, 'Stories in Action', *Psychological Science in the Public Interest*, 23 (2022), pp. 99-141.
[14] K.R. Dale, J.T. Fisher, J. Liao and E. Grinberg, 'The Shape of Inspiration: Exploring the Emotional Arcs and Self-Transcendent Elicitors Within Inspirational Movies', *Media Psychology*, 2023, pp. 1-23.

replication. When the outcome of a study is verified to be consistent, it is more likely to indicate a reliable claim to new knowledge. After the efficiency of the Python sentiment analysis program is determined, the analysis method will be used to replicate Reagan et al.'s study. The subquestion corresponding to this objective is as follows: 'Can the Python sentiment analysis program replicate the findings of Reagan et al. regarding the classification of texts into six emotional arcs, and what insights can be gained from this replication?'

The third objective of this thesis will be to provide a critical evaluation of sentiment analysis as a tool for plot visualization in literary studies as a whole. In their studies, Jockers, Reagan et al., and others frequently use the term "plot", but to what extent do the results of such analyses actually provide a visual depiction of the narrative of a story? Following the study's findings, this thesis will offer a critical assessment and suggest modifications to the methodology that might result in a more accurate depiction of a story's plot. In giving these suggestions, the goal is to not only address a gap in the research, but also to provide a possibility for practical relevance. The subquestion corresponding to this objective is as follows: 'To what extent do sentiment analysis results in literary studies provide an accurate visual depiction of a story's narrative or plot, and what modifications can be suggested to improve its accuracy and portrayal?'

1.3 Thesis structure

Chapter two will function as the theoretical framework for this thesis. First, the concept of sentiment analysis will be covered in this chapter; how do sentiment analysis tools function, and what is this tool used for in and outside of the digital humanities? This section will also contain a more detailed look at the R package Syuzhet, which functions as inspiration for the Python program that will be created for this thesis, in order determine how it functions and what its strengths and limitations are. Next, the literature more specifically focused on the use of sentiment analysis in narratology will be discussed. This section will also go into Reagan et al.'s research on the six shapes or emotional arcs that most stories fall into it, which is the study that will be replicated for this thesis. Lastly, this chapter will end with a section on the importance of replication in the field of the humanities, and how this differs from reproducing a study. Subsequently, the method will be discussed in the third chapter. The methodology chapter will go further into how the Python program functions to carry out the sentiment analysis, as well as how the shapes or 'emotional arcs', as Reagan et al. describe them, will be determined. In the fourth chapter the results are presented. Subsequently, a chapter will be dedicated as critical evaluation of sentiment analysis as a tool for plot extraction. The literary framework, as well as the results of the sentiment analysis, will be used to determine whether the results of sentiment analysis actually tell us something about the progression of plot throughout a story. Moreover, it will discuss the practical use of sentiment analysis, as well as methods for potentially enhancing these tools. Finally in the fifth chapter a conclusion is drawn.

2. Theoretical framework

2.1 What is sentiment analysis?

Sentiment analysis is an approach to natural language processing (NLP) and, simply stated, identifies the emotional tone behind a body of text. Bing Liu defines it as the computational study of opinions, sentiments, evaluations, attitudes, appraisal, affects, views, emotions, subjective, etc. expressed in text.[15] It can be applied to a variety of texts such as reviews, blogs, news, comments or any other type of document. Sentiment analysis creates a wide range of possibilities for researchers to analysis large clusters of text. In such research, targeted elements are often being analysed. These are usually single words, phrases, or sentences. Sometimes whole documents are studied as a sentiment unit, but it is generally agreed that sentiment resides in smaller linguistic units.[16] Whereas a method like topic modelling may deal with whole taxonomies of topics, sentiment classification usually deals with two classes: positive and negative.

This approach to analysing text mostly reached traction after the year 2000, primarily because relatively low quantities were recorded in digital form before then. [17] Due to the web and social media's explosive rise in the last twenty years, we now have a continuous flow of opinionated data captured in digital formats.[18] Other reasons are technological advancements, particularly the development in 'big data', and the advancement of natural language processing tools. These tools have caused a shift in focus to more challenging and, consequently, more conceptual problems.[19]

Sentiment analysis has become an extremely popular tool in computational science, in large due to its huge potential for practical applications, mostly for industrial purposes. One of the most intriguing NLP challenges for industrial applications nowadays is mining product reviews.[20] The long-term objective of being able to automatically identify consumer input in

---

[15] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions* (Cambridge: Cambridge university press, 2020), p. 1.
[16] B. Pang and L. Lillian, 'Opinion mining and sentiment analysis', *Foundations and Trends® in information retrieval*, 1–2 (2008), pp. 1-135.
[17] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions* (Cambridge: Cambridge university press, 2020), p. 3.
[18] H. Kennedy, 'Perspectives on sentiment analysis', *Journal of Broadcasting & Electronic Media*, *56* (2012), pp. 435-450.
[19] J. Sonntag and S. Manfred, 'Sentiment analysis: What's your opinion?', in C. Biemann and A. Mehler, (eds.), *Text mining*, *From Ontology Learning to Automated Text Processing Applications*, (New York City: Springer, 2014), pp. 177-199.
[20] H. Kennedy, 'Perspectives on sentiment analysis', *Journal of Broadcasting & Electronic Media*, *56* (2012), pp. 435-450.

significant amounts would aid merchandisers and manufacturers in creating specialized marketing campaigns catered to the standing a product has among its customers, and also contribute to improving the items. In addition, sentiment analysis has gained interest in academic research as well, and, more particularly, within the Social Sciences. Some researchers have examined the ways in which opinions are expressed during parliamentary discussions for example, while others are curious about automatically gathering opinions on political problems from newspapers or social media.[21]

The earliest articles that focus on the use of natural language processing to analyse the sentiment of a certain text dates back to the early 2000s. Motivated by topic modelling techniques, similarly developed around the turn of the millennium, Bo Pang and Lillian Lee considered the classification of documents by another standard: sentiment.[22] Pang and Lee used movie reviews as data and found that the results from machine learning are quite good in comparison to the human generated baseline. The machine learning techniques, in other words, were often successful in recognizing a positive and a negative movie review. Nonetheless, this early study also revealed a number of shortcomings which proved difficult to resolve to the present day. For example, the program had difficulties understanding a deliberate contrast to an expectation or to another review. By mentioning a positive sentiment in a negative review, the classifiers failed to comprehend the overall sentiment, whereas a human would easily have detected the true sentiment of the review.

In 2002, Peter Turney published a similar study in which he proposed a simple unsupervised learning algorithm for classifying reviews.[23] Whether a review contained a positive semantic orientation or a negative semantic orientation decided its classification as 'recommended' or 'not recommended' respectively. Turney's algorithm, tested on a variety of reviews, achieved an average accuracy of 74%.

Both of these early applications of what was later termed 'sentiment analysis' show a clear foundation for the use of this technique. Many present-day applications still focus on analysing reviews, but the possibilities of using this tool has since expanded significantly. Sentiment analysis has also been used to analyse changing opinions on political manners over

---

[21] Kennedy, 'Perspectives on sentiment analysis', pp. 435-450.

[22] B. Pang, L. Lee and S. Vaithyanathan, 'Thumbs up? Sentiment Classification using Machine Learning Techniques', *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (2002), pp. 79–86.

[23] P. Turney, 'Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews', *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02),* (2002), pp. 417-424.

time[24]; it has been used to predict the success of a novel or movie[25]; and more recently, it has been used to analyse the impact of the coronavirus on social life.[26]

When sentiment analysis is used to analyse texts such as reviews, tweets or surveys, the underlying presumption is made that the emotion in the text has a set value; it is either positive or negative. However, we naturally assume the sentiment to change across larger texts, like a novel. How sentiment analysis has been employed to analyse such lengthy literary texts will be further discussed in chapter 2.3 and 2.4. First, chapter 2.2 will provide an overview of how the majority of sentiment analysis tools function.

2.2 How do sentiment analysis tools function?

In 'Sentiment Analysis in Literary Studies; A Critical Survey', Simone Rebora present a critical evaluation of how most sentiment analysis tools function.[27] Rebora mainly focuses on three interconnected aspects:

- The emotion theory which the tool adopts
- The method used to build the emotion dictionary
- The technique adopted to accomplish the analysis

As for the emotion theory, sentiment tools tend to follow one of two structures:

- A dimensional representation of emotions
- A discrete representation of emotions

The dimensional representation of emotions is based mostly around Russell's theory from 1980 in which he proposed a bi-dimensional system able to chart all emotional states.[28] His theory states that any human emotion can be logically defined by combining the two dimensions of valence (positive vs. negative) and arousal (calm vs. intense). Numerous sentiment analysis tools include this principle by further simplification. In most cases this means reducing it to valence alone, on a continuous scale that ranges between two extreme values. Nonetheless, this analytical simplification (positive vs. negative) suffices in most commercial applications, as the core concept of sentiment analysis is to mine the contrast

[24] Y. Matalon, O. Magdaci, A. Almozlino and D. Yamin, 'Using sentiment analysis to predict opinion inversion in Tweets of political communication', *Scientific reports*, 11 (2021), pp. 1-9.
[25] V. Jain, 'Prediction of movie success using sentiment analysis of tweets', *The International Journal of Soft Computing and Software Engineering*, 3 (2013), pp. 308-313.
[26] M. Singh, A.K. Jakhar and S. Pandey, 'Sentiment analysis on the impact of coronavirus in social life using the BERT model', *Social Network Analysis and Mining*, 11 (2021), pp. 1-11.
[27] S. Rebora, 'Sentiment Analysis in Literary Studies. A critical Survey', *Digital Humanities Quarterly*, In press 2023
[28] J.A., Russell, 'A circumplex model of affect', *Journal of personality and social psychology*, 39 (1980), pp. 1161-1178.

between approval and disapproval, or between endorsement and rejection. For literary studies however, this distribution might pose a problem. Sprugnoli et al. demonstrated in 2016 that annotating historical texts between positive and negative was an almost impossible task for humans, posing the question if machine-annotated texts would be reliable at all.[29]

Next, sentiment analysis tools can measure the sentiment of a text by means of a dictionary, or lexicon. This lexicon contains a list of individual words with an emotional value attributed to them. These dictionaries play a major role in how much emotion or sentiment will be measured overall. The ways in which these dictionaries are created can be distinguished in three main approaches:

- Word lists
- Vector space models
- Hybrid approaches

Worst lists are the simplest of these approaches: a list of words is created where each one is attributed in terms of sentiment. Such lists can be created by means of crowdsourcing, where the final values are then assigned via majority consent. While the concept of such a word list is simple, it required extensive precatory work. Vector space models are more frequently used. Here, words are converted into multidimensional vectors that contain data on semantic similarity. These dictionaries adapt to a specific linguistic context, through distributional semantics and word embedding. This technique starts with a selection of what is known as 'seed words' such as 'good', 'bad', or any word related to a basic emotion. It is possible to instantly assign a value to every word in a dictionary using these seed words. Since both word lists and vector space models have their strengths and shortcomings, hybrid approaches are aimed at reaching a compromise by combining the two.

The final aspect of sentiment analysis tools that Rebora mentions relates to the approach used to complete the analysis. Again, three approaches are recognized:

- Simple or advanced wordcount
- Syntactic structure analyses
- Machine learning techniques

The wordcount method produces a 'bag of words' analysis by ignoring sentence structure and word order. To determine a score, the words are counted and their values are added. The more extensive and lengthy the texts are, the more dependable this becomes. Syntactic structure

[29] R. Sprugnoli, S. Tonelli, A. Marchetti and G. Moretti, 'Towards sentiment analysis for historical texts', *Digital Scholarship in the Humanities*, 31 (2016), pp. 762-772.

analysis is more advanced than working exclusively with word counts. It examines the structure of the sentence and makes adjustments for complexities in sentences. Such complexities arise in sentences that contain negation: 'He was not a bad person.' gets recognized as a positive statement, despite containing the word 'bad'. Machine learning places itself as the most successful approach of these three methods. It uses a bottom-up approach to create a knowledge model through a trial-and-error process. This approach can be used to solve even the most difficult problems, such as understanding irony and sarcasm.

2.3 Sentiment analysis and narratology

Since the digital humanities is focused on using computational techniques to study traditional humanities subjects such as history and literature, sentiment analysis has demonstrated to be an interesting research tool. Humanities researchers have used sentiment analysis to revisit results found by means of distant reading, or to find new results altogether. This chapter will focus on how researchers have applied sentiment analysis in the field of digital humanities.

In addition to the studies conducted by Jockers and Reagan et al. on sentiment analysis in literary texts (as discussed in chapter 1), it is worth noting that sentiment analysis has found various other applications within this domain. In 2018, Kim and Klinger conducted a study on sentiment analysis for computational literary studies.[30] They identified five key areas of applications:

1. Classification of literary works according to the emotions they portray
2. Classification of genres
3. Modelling sentiments and emotions from books from earlier periods
4. Character network research using emotional connections
5. Additional and broader uses

The research done by Reagan et al. and Jockers, together with Kim and Klinger's list of applications, highlights various applications of sentiment analysis within narratology, including emotional classification of literary works and genre classification. Narratology is the discipline that studies society and culture through literary analysis. It revolves around the presentation of a series of events, whether real or fictional, to the reader.[31] Since substantial

[30] E. Kim, and R. Klinger, 'A survey on sentiment and emotion analysis for computational literary studies', *arXiv preprint arXiv:1808.03137*, (2018).
[31] S. Min and J. Park, 'Narrative as a Complex Network: A Study of Victor Hugo's Les Misérables', *Proceedings of HCI Korea*, (2016), pp. 100-107.

part of the study of narratology is dedicated to the concept of ‘plot’, it is important to establish what exactly the definition of ‘plot’ is.

Because the term ‘plot’ has been employed in so many different situations, narratologists find it difficult to define its scope.[32] Karin Kukkonen has defined the many ways of which plot can be conceived of in the *Handbook of Narratology*, which is still one of the most established theorizations on plot. In here, she identifies plot as a fixed structure; plot as a progressive structuration and plot as a part of the authorial design.[33] Kukkonen states that in practice these different conceptualisations are often combined, and describes the overall definition of plot as ‘the ways in which the events and characters’ actions in a story are arranged and how this arrangement in turn facilitates identification of their motivations and consequences.’

The notion that there are only a finite number of plots and that every new story is only a variation on them is very old, going at least as far back as Aristotle's Poetics.[34] He outlines the fundamental triangle-shaped plot structure with a start, middle, and end. Since then, more plot structures have been described which can be recognized in many stories. The hero's journey’, also known as the ‘monomyth’, is a frequently mentioned structure in narrative theory and comparative mythology. It describes stories in which a hero embarks on an expedition, overcomes a life-or-death crisis, and returns home altered or transformed.[35] Joseph Campbell, who used the monomyth to compare mythology, popularized this structure, which is divided into four thresholds and 17 stages. Russian Formalist Vladimir Propp is well known for his work on morphology of the folktales, which examined the underlying structure and the functions of Russian folktales. In his analysis on folktales, he distinguished between two terms: ‘fabula’ (the chronological order of events in a story) and ‘syuzhet’ (the specific arrangement and presentation of those events in the actual telling or presentation of the story). Eventually, Jockers would name his sentiment analysis programme after the latter term.

American author Kurt Vonnegut, known for Slaughterhouse-Five, among many other novels, wrote his master’s thesis for anthropology on the notion that ‘stories have shapes which can be drawn on graph paper, and that the shape of a given society’s stories is at least

---

[32] M. Blythe, Research fiction: storytelling, plot and design. ‘*Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*’, pp. 5400-5411.
[33] K. Kukkonen, ‘Plot’, in P. Hühn, J. Meister, J. Pier and W. Schmid, *Handbook of narratology*, (Berlin: de Gruyter, 2014), pp. 706-719.
[34] M. Finkelberg, ‘Aristotle and episodic tragedy’, *Greece & Rome*, *53* (2006), pp. 60-72.
[35] Campbell, J., *The hero with a thousand faces*, (New York City: Pantheon Books, 1949), p. 23.

as interesting as the shape of its pots or spearheads'.[36] Vonnegut famously considers this theory his greatest contribution to the literary culture. One of the shapes that Vonnegut describes is the 'Cinderella' shape, which is visualized in figure 2.1.

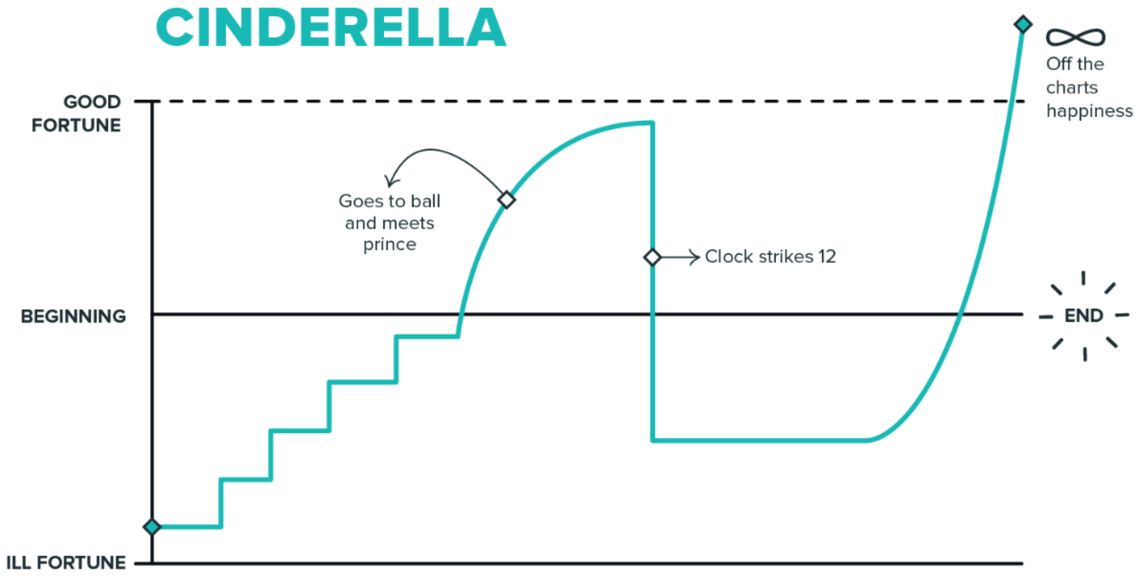

Figure 2.1: A map of the story structure described by Vonnegut as 'Cinderella', adapted to the fairy-tale of *Cinderella*. Via Visage.[37]

Vonnegut describes the story's progression as a staircase-like ascent in good fortune that corresponds to Cinderella's fairy godmother's entrance, culminating in a high point at the ball, before a sharp descent back into bad luck as the clock strikes midnight. The graph concludes with a swift return to good fortune, as Cinderella fits the glass slipper and lives happily ever after.

While Campbell's work on the monomyth has been met with fierce critique (often being accused of source bias),[38] and while Vonnegut's master thesis on the basic shapes of stories was rejected, the concept that stories follow a fixed structure is still being explored in the humanities to this day. Computational methods, such as sentiment analysis, have introduced a new method of extracting these structures.[39] With this specific purpose in mind,

---

[36] K. Vonnegut, 'Kurt Vonnegut on the Shapes of Stories', *Youtube*, 30 October, 2010. < https://www.youtube.com/watch?v=oP3c1h8v2ZQ&ab_channel=DavidComberg> (14 September, 2022).
[37] K. French, 'Kurt Vonnegut Graphs the Shapes of Stories', *Visage*, 13 August, 2014. < https://visage.co/kurt-vonnegut-shows-us-shapes-stories/> (14 September, 2022).
[38] O. Rank, F. Raglan and A. Dundes, *In quest of the hero* (Princeton: Princeton University Press, 1990), p. 11.
[39] 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', *EPJ Data Science*, 5 (2016), pp.1-12.

Matthew Jockers created a sentiment analysis tool entitled Syuzhet, a package created for R that analyses the sentences of literary texts. This way, Jockers aimed to reveal the emotional and affective shifts that act as a proxy for the narrative movement between conflict and conflict resolution.[40] With Syuzhet, Jockers has initiated a wave of digital humanities research that aims at extracting sentiment data from novels, plays, movies etc. Chapter 2.5 will focus more on how Syuzhet operates, as well as take a critical look at its methods. But first, specific cases of researchers using sentiment analysis as a tool for extracting and visualizing data on the emotional trajectory of stories will be discussed in chapter 2.4.

2.4. Extracting and visualizing sentiment from literary texts

A key study for this thesis was done by Reagan, Mitchell, Kiley, Danforth and Dodds in 2016, in which they use sentiment analysis to uncover patterns in the dramatic arcs of stories.[41] They argued that because of advances in computing power, natural language processing and text digitization, it has been made available to examine the structure of stories through a ‘big data’ lens. Our ability to communicate is based in part on a shared emotional experience, with stories often taking different emotional trajectories and generating meaningful patterns for us. By using sentiment analysis on a subset of 1,327 stories from Project Gutenberg’s fiction collection, they found a set of six fundamental emotional arcs that represent the essential building block of complicated emotional trajectories. On top of that, they examined the most similar contemporary stories and discovered that certain emotional arcs have more success than others (as measured by downloads). Essentially, stories that follow specific emotional arcs are more likely to resonate with readers and attract a larger audience. By analysing sentiment and emotional patterns, Reagan et al. were able to uncover these correlations, indicating the influence of emotional progression on readers’ reception of stories.

The main takeaway from the article by Reagan et al. is that their research finds broad support that, from a corpus of 1,327 stories, all follow one of six basic emotional arcs. These emotional arcs are, for example ‘Rags to riches’, ‘Tragedy’, etc. In the rags to riches story arc, a poor or rejected main character usually acquires something (money, love etc.), loses it, and then gains it again by the conclusion of the story. Reagon et al. describe this arc simply as

[40] Jockers, M., ‘Revealing Sentiment and Plot Arcs with the Syuzhet Package’, *matthewjockers.net*, 2 February, 2015. <https://www.matthewjockers.net/2015/02/02/syuzhet/> (12 September, 2022).
[41] A.J. Reagan, L. Mitchell, D. Kiley, C. Danforth P. and Dodds, ‘The emotional arcs of stories are dominated by six basic shapes’, *EPJ Data Science*, 5 (2016), pp.1-12.

‘rise’. This rise can be observed in novels such as *The Winter’s Tale*. Tragedy is describes as ‘fall’ and contains perhaps the most famous tragedy: *Romeo and Juliet*. Both of these shapes are visualized in figure 2.2. Reagan et al. argue that visualizing the emotional trajectory of a story using sentiment analysis is a great indicator of the dramatic course of the story.

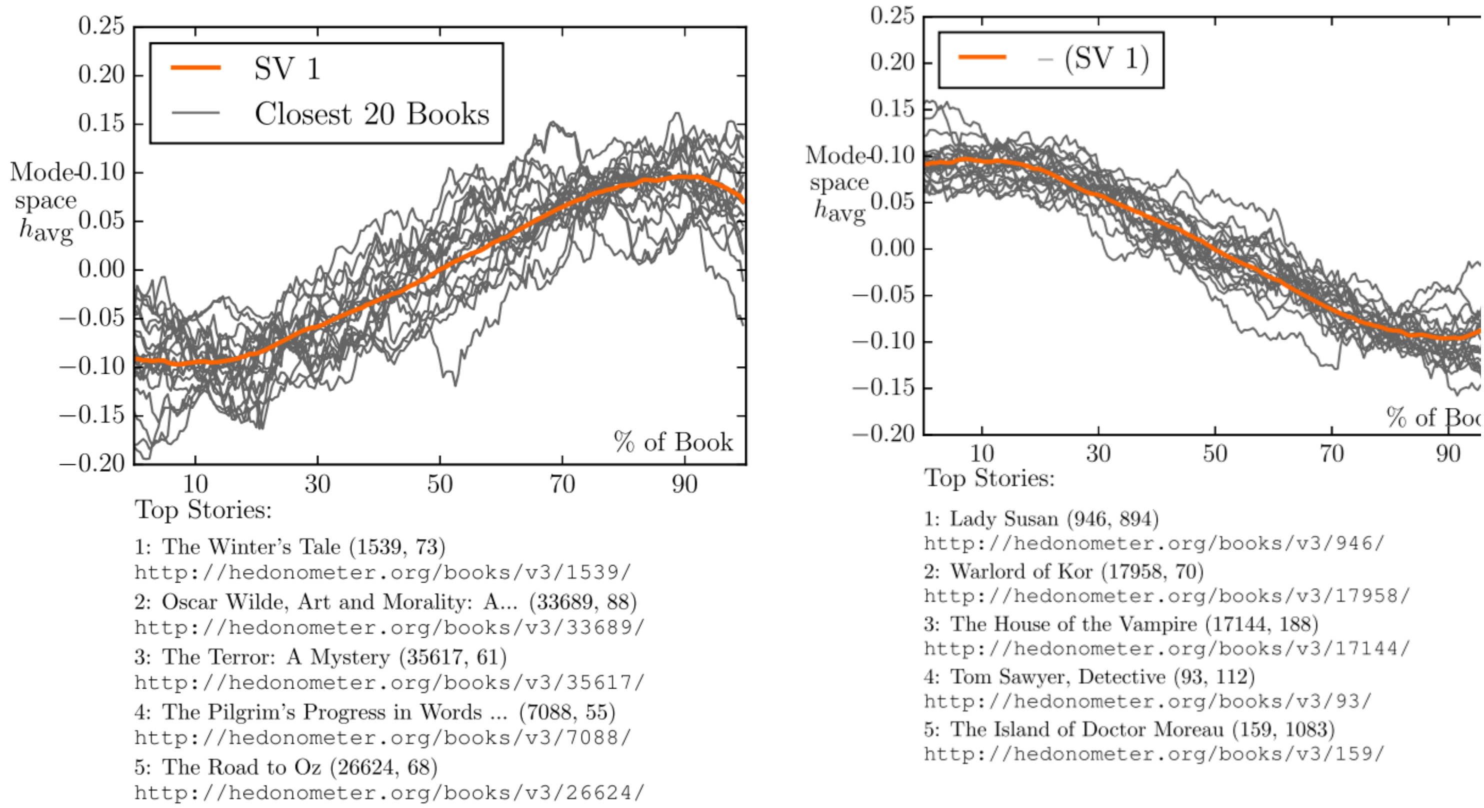

Figure 2.2: Left shows the graph which visualizes Reagan et al.’s shape described as ‘rise’, more commonly known as ‘Rags to riches’. Right shows the shape of ‘fall’, more commonly known as ‘Tragedy’. Via Reagan et al.[42]

Continuing on this research Reagan et al., Elins and Chun wanted to explore if they could apply this method on a modernist novel, one that breaks with traditional plot structure.[43] By applying Syuzhet on Virignia Woolf’s *To the Lighthouse*, Elins and Chun found that while *To the Lighthouse* does not follow a storyline with a typical hero at the centre, it does have an underlying emotional structure that is scattered among its characters, which they refer to as a ‘distributed heroine model’. They argue that, as a result of this analysis, they are able to confirm what certain critics have already determined about *To The Lighthouse*. That is, that

[42] A. J. Reagan et al., ‘The emotional arcs of stories are dominated by six basic shapes’.
[43] K. Elkins and J. Chun, ‘Can Sentiment Analysis Reveal Structure in a Plotless Novel?’, *arXiv preprint arXiv*, (2019).

while the numerous characters are introduced to us in a somewhat disjointed way, a pattern emerges in their individual emotions as we move through them. This pattern is visualized through the graphs of the Syuzhet package.

In 2013, Nalisnick and Baird presented an automatic method for analysing sentiment dynamics between characters in plays.[44] Similar to Reagan et al., Nalisnick and Baird presented their results in the form of a line graph, which displays the sentiment score on the y-axis and the full narrative time of the story on the x-axis. However, in contrast to Reagan et al., they measured the sentiment score per character. First, Nalisnick and Baird attempted to interpret to which character a line in a text is directed towards. Subsequently, the emotions in a character's speech can be assigned appropriately, allowing them to create lists of the character's foes and allies, as well as identifying situations that are crucial to the character's emotional growth. The analysis was done using AFINN: a set of lexicons used for sentiment analysis developed by Finn Årup Nielsen. A number between -5 (negative) and +5 (positive) is assigned to a list of English terms in order to rate the sentiment of a text. By using the analysis on some of Shakespeare's plays, Nalisnick and Baird were able to visualize the sentiment which two characters in a play felt towards each other. In the case of Hamlet and Gertrude, the two lines that represented the sentiment of both characters changed drastically in the third act, crossing each other and moving in opposite directions, as visualized in figure 2.3.

[44] E.T. Nalisnick and H.S. Baird, 'Character-to-character sentiment analysis in Shakespeare's play', *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2 (2013), pp. 479-483.

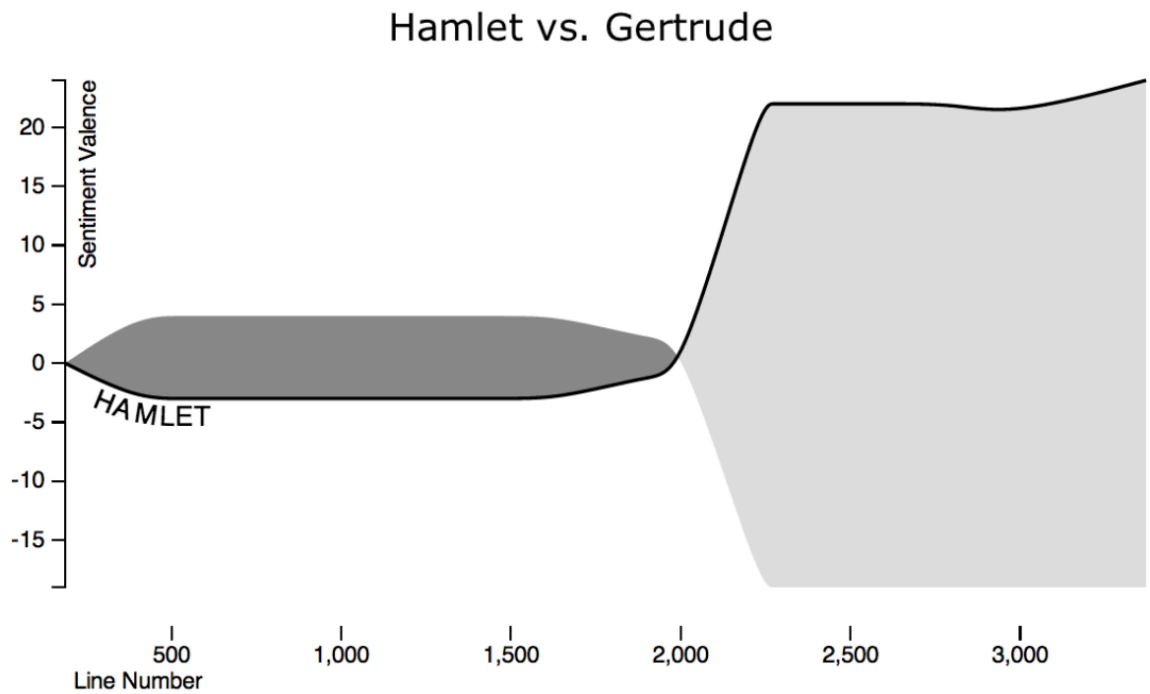

Figure 2.3: Character-to-character sentiment analysis of Hamlet. The black line represents Hamlet's sentiments for Gertrude, and the opposing line of the grey area represents Gertrude's sentiments for Hamlet. Via Nalisnick and Baird.[45]

Nalisnick and Baird proved with their study that interpersonal relationships can be analysed using sentiment analysis. This character-to-character sentiment analysis gives a different visualization from Reagan et al.'s sentiment analysis. Whereas Reagan et al.'s method shows the sentiment throughout an entire story at one glace, this technique provides the possibility to generate more graphs, each showing the sentiment that two characters have towards each other. It combines the plotting aspect of sentiment analysis with an interpersonal relationship analysis.

2.5. Syuzhet and VADER; attributes and limitations

As discussed earlier, sentiment analysis tools have conveyed a variety of commercial functions. VADER is a popular lexicon- and rule-based sentiment analysis tool designed specifically for social media sentiment analysis.[46] Texts like reviews or tweets may be

---

[45] Nalisnick and Baird, 'Character-to-character sentiment analysis in Shakespeare's play'.
[46] C.J. Hutto, 'vaderSentiment', *github*, 1 april, 2022. < https://github.com/cjhutto/vaderSentiment> (1 June, 2023).

examined with VADER.[47] Syuzhet distinguishes itself as a dictionary-based tool by being primarily created to analyse the arcs of sentiment in literary texts. The name 'Syuzhet' comes form the Russian Formalists Victor Shklovsky and Vladimir Propp. They divided narrative into two components: the 'fabula' and the 'syuzhet'. The latter of which refers to the device or technique of a narrative.[48] In short, the distinction is between the chronological sequence of events on the one hand, and the manner in which these events are narrated on the other.

Hutto and Gilbert's released VADER in 2014. Matthew Jockers designed Syuzhet one year later, in 2015. The Syuzhet package comes with four sentiment dictionaries: the Syuzhet, Bing, Afinn, and NRC lexicons. How these lexicons are constructed can be seen in figure 2.4.

	Syuzhet	Bing	Afinn	NRC
No. of Positive Words	3587	2006	878	2312
No. of Negative Words	7161	4783	1598	3324
No. of Other Words	-	-	1	8265
Total	10748	6789	2477	13901

Figure 2.4: The amount of sentiment words used in the Syuzhet package's lexicon. Via Digital Humanities Quarterly.[49]

The Syuzhet lexicon is a custom-made sentiment dictionary by the Nebraska Literary Lab and functions as Syuzhet's default dictionary. As the words were selected from a collection of 165,000 human-coded sentences, collected from a small corpus of modern books, the default lexicon should be more suited to fiction.[50] VADER's default lexicon, like Syuzhet's, is custom built. Nevertheless, a different method for creating the lexicon was used for it to be better suited to analyse social media content. The VADER dictionary was created based of existing well-established lexicons (such as UWC and ANEW), but supplemented with additional

[47] V. Bonta and N. Janardhan, 'A comprehensive study on lexicon based approaches for sentiment analysis', *Asian Journal of Computer Science and Technology*, 8 (2019), pp.1-6.
[48] M. Jockers, 'Syuzhet Package in R', RDocumentation, December 14, 2017. < https://www.rdocumentation.org/packages/syuzhet/versions/1.0.4> (2 September, 2022).
[49] H. Kim, 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.' *DHQ*, 16 (2022), n.pag.
[50] M. Jockers, 'Revealing Sentiment and Plot Arcs with the Syuzhet Package', *matthewjockers.net*, 2 February, 2015. <https://www.matthewjockers.net/2015/02/02/syuzhet/> (12 September, 2022).

lexical features commonly used to express sentiment in social media text such as emoticons, acronyms and slang.[51]

As figure 2.5 shows, Syuzhet has been the most frequently downloaded package for sentiment analysis in R since 2016, making it the most popular sentiment analysis tool for this programming languag. Kim argues that Jockers' Syuzhet is the software that originated a wave of interest towards sentiment analysis in literary studies, as a result of its distinct functionalities in comparison to other tools.[52] Syuzhet contains a series of visualization algorithms, which use various smoothing techniques to produce a visualization of 'plot arcs'. It gains its efficiency by combining speed and visualization power. Based on the classification of methods made in chapter 2.2 of this thesis, Syuzhet's operating can be classified as follows:

- It uses a dimensional representation of emotions
- It makes use of a lexicon (or dictionary) created through crowdsourcing
- The analysis is run via simple wordcount

While VADER can be run in R as well, its program is primarily used in Python. Its operating is very similar to that of Syuzhet; it too uses a dimensional representation of emotions, and makes use of a lexicon created through crowdsourcing. However, where Syuzhet makes use of simple wordcount, VADER uses a syntactic structure analysis, which is a slightly more advanced method.

[51] 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*, 8, (2014), pp. 216-225.
[52] Kim, 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.'

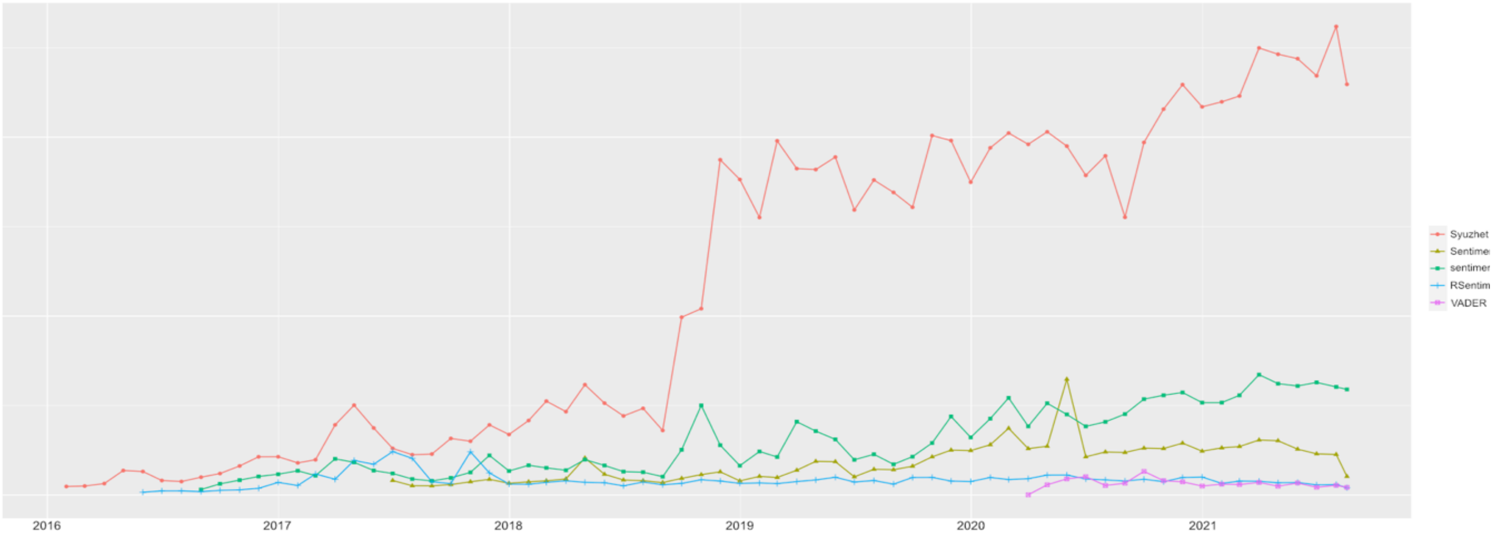

Figure 2.5: The number of sentiment analysis R packages downloaded per month (created on August 17, 2021). Via Digital Humanities Quarterly.[53]

Both Syuzhet and VADER are two of the most popular sentiment analysis options today. Nevertheless, both of these sentiment analysis systems have inherent flaws due to the way they function. Firstly, Syuzhet and VADER both operate with a dimensional representation of emotions, primarily focusing on the positive versus negative valence. This approach limits the analysis of human emotion to a binary framework, which can potentially result in the loss of relevant information. Sentiment analysis often encompasses a broader spectrum beyond these two extremes, incorporating nuances such as aesthetic admiration (e.g., beauty versus ugly) and differentiating them from physical reactions (e.g., pleasure versus pain). However, it is important to acknowledge that the choice to employ this dichotomous split in sentiment analysis tools may be driven by pragmatic reasons, as it allows for simpler implementation and practical usage. The fundamental idea of these analysis tools is to mine opinions, meaning that a distinction between positive and negative valence becomes sufficient in accomplishing this task.[54] The argument can be made that this simplified distinction is more useful in commercial applications where the opinions of reviews are being analysed, in comparison to analysing the often complex sentiment of literary texts. As proven by Sprugnoli et al., the distinction of literary texts in positive and negative is a very complex task, and sentiment analysis tools made with the intent to analyse

[53] H. Kim, ‘Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.’ *DHQ*, 16 (2022), n.pag.
[54] S. Rebora, ‘Sentiment Analysis in Literary Studies. A Critical Survey’, *Digital Humanities Quarterly*.

lengthy texts would benefit from a unique framework with a broader set of categories.[55] Nonetheless, the distinction between positive vs. negative is still used by many researchers (such as Reagan et al.), for its simple effectiveness.

Secondly, the main dictionary is built through crowdsourcing, which is not a foolproof method. The trustworthiness of the annotators must be taken into question. Moreover, the resulting lexicons might prove to be a good representation of the emotions experiences by current-day Internet users, but not of the value system that supports a Shakespearean play.[56]

Thirdly, the use of simple wordcount by Syuzhet is not the most accurate form of analysis available.[57] Especially because of this, Kim argues that it is one of the least advanced types of software to conduct sentiment analysis, despite its popularity.[58] When dealing with shorter texts, wordcount is highly ineffective. In the case of Syuzhet, it should also be mentioned that the algorithm's performance is also affected by the fact that, instead of counting the number of occurrences, it counts the unique words. This means that if a positive word occurs twice in the same sentence, the overall sentiment score will be the same if it had only occurred once. This is where VADER's analysis method is ahead of that of Syuzhet; the syntactic structure analysis is able to recognise when a positive or negative word occurs more often, as well as valence shifters and intensifiers. This means that Syuzhet, which assigns a score to each word independently, will conclude that a statement like "I don't love you" is positive. VADER on the other hand can detect the negation in this statement and assigns it a negative score. Nevertheless, while VADER can detect sentences that contain negation, both analysis methods still have difficulty recognizing other specific uses of language, such as sarcasm or irony.[59] As a result, Kim has argued that the humanities should use deep learning approaches for sentiment analysis, in contrast to wordcount or syntactic structure, as this method has proven to be the most advanced option.[60]

All together, both Syuzhet and VADER reveal a fairly similar structure. Kim argues that, because of its wordcount method, Syuzhet is the lesser-advanced choice of software to perform sentiment analysis. Nonetheless, the package is verified by literary academics

[55] R. Sprugnoli, S. Tonelli, A. Marchetti, and G. Moretti, 'Towards sentiment analysis for historical texts', *Digital Scholarship in the Humanities*, 31 (2016), pp. 762-772.
[56] S. Rebora, 'Sentiment Analysis in Literary Studies. A Critical Survey', *Digital Humanities Quarterly*.
[57] Rebora, 'Sentiment Analysis in Literary Studies. A Critical Survey'.
[58] Kim, 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.'
[59] A. Swafford, 'Why Syuzhet Doesn't Work and How We Know', *Anglophile in Academia: Annie Swafford's Blog*, March 30, 2015. <https://annieswafford.wordpress.com/2015/03/30/why-syuzhet-doesnt-work-and-how-we-know/> (5 September, 2022).
[60] Kim, H., 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.' *DHQ*, 16 (2022), n.pag.

through a comparison of computational sentiment evaluations.[61] It produces effective results, and its transparency and adaptability make it a popular and efficient tool. It also contains a bigger lexicon than VADER, and an accessible way for text preparation and visualization.[62] Of course, the purpose with which both lexicons were crated should be taken into account as well when deciding to use one or the other.

2.6. The need for replication in the humanities

This thesis will partly function as a replication study of *The emotional arcs of stories are dominated by six basic shapes* by Reagan, Mitchell, Kiley, Danforth and Dodds.[63] It will replicate the findings using the same data (a selection of 60 literary works of the 1,357 selected by Reagan et al.), but with a different research protocol. The aim of replicating this study is to add information about the reliability of the conclusions and estimates drawn from the data by Reagan et al. Before the significance of replication studies in the humanities is explained, the difference between replication and reproduction, two similar processes, will be defined.

The word 'reproducibility' refers to the capacity to accurately reproduce each step of the workflow that was followed in a research using the data and techniques that the original author had either recorded or made available.[64] When the same procedure is followed, results can be obtained repeatedly with a high degree of agreement. The humanities, a discipline whose results and analyses are frequently reliant on interpretation and qualitative methodologies, have up to now been less concerned with the idea of reproducibility.[65] Humanities academics are often interested in the significance or worth of cultural or historical material, and how these are interpreted frequently depends heavily on the prior knowledge and theoretical viewpoint of scholars. Even when every stage of a study has been properly documented, it is still possible that different researchers will come to different interpretations of the same data while using the same analytical techniques and applying them to the same

---

[61] K. Elkins and J. Chun, 'Can Sentiment Analysis Reveal Structure in a Plotless Novel?', *arXiv preprint arXiv*, (2019).
[62] Kim, H., 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.' *DHQ*, 16 (2022), n.pag.
[63] 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', *EPJ Data Science*, 5 (2016), pp.1-12.
[64] S.N. Goodman, D. Fanelli, and J.P. Ioannidis, 'What does research reproducibility mean?', *Science translational medicine*, 8 (2016), pp. 1–6.
[65] Holbrook, B., Penders, B., de Rijcke, S., 'The humanities do not need a replication drive', Centre for Science and Technology Studies, 21 January, 2019. < https://www.cwts.nl/blog?article=n-r2v2a4&title=the-humanities-do-not-need-a-replication-drive> (7 July, 2022).

data. Nonetheless, the aim for reproducibility encourages a meaningful condition to scientific research. Making sure research is repeatable demands the release of the raw data and enables others to do thorough analyses that are unaffected by earlier ones. Reproducibility makes it possible to avoid missing the crucial phase of data processing.[66] Creating transparency is therefore the primary goal of computational reproducibility.

Given the notion that transparency for the digital humanities serves as a tool to promote collaboration and reuse rather than serving as a goal in and of itself, it might be argued that replication is ultimately more important than reproduction. Replication refers to a re-implementation of the experiment as opposed to reproduction, which effectively means an identical duplicate of an experiment based on the same data and code.[67] It necessitates a careful revaluation of the study to ascertain whether the conclusions of the prior analysis were accurate. Replication entails a critical evaluation of the data and of the methodology in addition to an effort to verify the results of prior studies. In such a replication study, the decision could be made to reassemble or gather the dataset that was described in the preceding study.[68] This kind of replication can be driven by the need to validate prior findings and guarantee that subsequent studies can build on reliable outcomes. Other kinds of replication can be separated from this strict form of replication, in which researchers try to accurately imitate the original methods. For instance, researchers may test the generalizability of a method by applying it to a new collection of data, or they might try to address the same topic using the same data and a different technique. Thanks to this they can assess the reliability of certain findings.

Based on a large number of scientists who have been speaking of a replication crisis in multiple scientific disciplines, Rik Peels wrote on article arguing that replication in the humanities is worth pursuing.[69] Since the humanities employ a wide range of empirical methods, Peels argues that replication studies are possible and desirable, since such studies help in giving greater validity to scientific findings. Peels and Bouter also argue that the current state of affairs in the humanities is that they lack studies that are explicitly designed as replication studies.[70] Since the need for replication studies in other scientific fields (such as

---

[66] S.N. Goodman, D. Fanelli, and J.P. Ioannidis, 'What does research reproducibility mean?', *Science translational medicine*, 8 (2016), pp. 1–6.
[67] Goodman, Fanelli and Joannidis, ''What does research reproducibility mean?'.
[68] R. Peels, 'Replicability and replication in the humanities', *Research Integrity and Peer Review*, 4 (2019), pp.1-12.
[69] Peels, 'Replicability and replication in the humanities'.
[70] R. Peels and L. Bouter, 'The possibility and desirability of replication in the humanities', *Palgrave Communications*, 4 (2018), pp. 1-4.

biomedical, natural and social sciences) originated from failed attempts at replication, Peels and Bouter argue that more replication studies need to be carried out in order to assess the need for these.

3. Methodology

This thesis has three main objectives: to create and evaluate a Python program which functions as an alternative to Syuzhet, to use this program to replicate the research set up by Reagan et al. in which the emotional arc shapes of texts are determined, and to provide a critical evaluation of sentiment analysis as a tool for plot visualization in literary studies as a whole. This methodology chapter will first motivate the decision to experiment with VADER using the programming language Python, rather than Syuzhet. Next, the method of the sentiment analysis program used for this thesis will be explained. This method focuses on the use of aggregate scores in order to answer the question whether we can use aggregates of sentiments scores to represent the plot of a literary text. Lastly, this chapter will discuss the texts that will be analysed, as well as how the shapes of these stories will be determined to replicate Reagan et al.'s research.

3.1 Experimenting with the VADER analysis method and lexicon

In 2022, Syuzhet.R was the most popular sentiment analysis tool based on number of downloads.[71] Syuzhet was created by Jockers with the specific intent in mind of creating a sentiment analysis tool that could measure and illustrate the progression of sentiment throughout a literary text. As a result, such literary research is more often than not done using Syuzhet, in contrast to the numerous other sentiment analysis programs that are available. Nevertheless, Syuzhet only functions using the programming language R. The decision to experiment with the programming language Python for this thesis, in contrast to R, is based on three decisions.

Firstly, Natural Language Processing (NLP) is a field of study that can be pursued using various programming languages, including R and Python. While R has expanded beyond its origins as a statistical package and is commonly used for NLP tasks, Python offers certain advantages in versatility and broader applicability. As a general-purpose programming language, Python can be employed for a wide range of applications, including web development, data manipulation, and machine learning. This flexibility makes Python an attractive choice for researchers and practitioners looking to explore different domains and leverage NLP techniques within a larger ecosystem of tools and frameworks. Python is often

[71] H. Kim, 'Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.' *DHQ*, 16 (2022), n.pag.

the de-facto programming language for text processing, with a lot of built-in capability that makes it easy to use and fairly quick, as well as a variety of feature-rich packages like NLTK.[72] A quick online search on popular programming platforms also indicates that Python is often recommended when it comes to NLP research because of its readability, extensive machine-learning environment, and deep learning APIs. Although much of this depends on individual preferences, the fact remains that adapting Syuzhet to a different programming language will allow NLP researchers to choose between different languages, based on their objectives and requirements. Secondly, by modifying the analytic tool VADER to work similarly to Syuzhet, it is possible to examine if two of its fundamental properties will work in a literary research context as well; its lexicon, and its analysis method. As discussed in chapter 2.5, Syuzhet's default lexicon is most often used for literary research. For this thesis, the lexicon for VADER will be used. Both the default Syuzhet lexicon and the VADER lexicon share similarities; they both limit the emotional range of a text to valence, and they both use word lists which have been built through crowdsourcing. VADER's lexicon is based on words found in a social media context, and therefor incorporates numerous common lexical features, such as emoticons, acronyms and slang.[73] Syuzhet's lexicon on the other hand was selected from a collection of 165,000 human-coded sentences, collected from a small corpus of modern books.[74] While this gives the impression that Syuzhet's lexicon is more suited to fiction, it is still worth testing whether this is actually true. Another lexicon might produce outcomes that are comparable. In comparison to other lexicons based on social media texts, VADER is not only very popular, but also highly accurate.[75] This makes for an interesting dictionary to experiment with on literary texts. Apart from this, VADER's analysis methodology is slightly more advanced than that of Syuzhet. By means of syntactic structure analyses, in comparison to simple word count, VADER can detect valence shifters and intensifiers. In practice, this will make VADER recognize linguistic forms such as negation. Whereas Syuzhet measures the words in 'I don't love you anymore' individually, and therefor gives it a positive sentiment score, VADER will recognize the overall sentiment of the

---

72 C.D. Larose and D.T. Larose, *Data science using Python and R* (New Jersey: John Wiley & Sons, 2019).
73 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*, 8, (2014), pp. 216-225.
74 M. Jockers and, R. Thalken, 'Sentiment analysis', *Text Analysis with R: For Students of Literature*, (2020), pp. 159-174.
75 M.A. Al-Shabi, 'Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining' *IJCSNS*, *20* (2020), p.1-7.

sentence as negative. Using a more advanced analysis method might prove to give better results in the overall sentiment graph, in comparison to Syuzhet.

Lastly, building a sentiment analysis method step-by-step not only enhances transparency but also promotes critical evaluation when there is a change of language. By constructing the method from scratch, it becomes more transparent, allowing for a clear understanding of the underlying processes. Simultaneously, when transitioning from one language to another, such as moving from R to Python, programmers are compelled to engage with the logic of the method rather than resorting to mere code replication. This combination of step-by-step construction and language transition fosters a comprehensive approach, enabling a deeper evaluation of the sentiment analysis process based on the aggregation of sentiment scores.

3.2 Building and evaluating the sentiment analysis programme

This chapter will go into further detail about how the sentiment analysis tool utilised for this thesis will work. The method will be very similar to how Syuzhet operates using R. However, in this thesis, Python will be employed as the programming language, and the sentiment analysis will be conducted using the NLTK module VADER, which was originally developed as a standalone package and later incorporated as an add-on to NLTK. The following section will discuss how VADER recognises and extracts sentiment found in a text, as well as how the data will be visualised.

Firstly, a text file (novel, play, etc.) will be loaded in as a single long string. Since the purpose of performing sentiment analysis is to uncover the progression of sentiment throughout a text, it is necessary to tokenize this string into individual sentences. Jockers argues that sentences are the fundamental units of composition of a story, so in order to understand the sentiment of an entire story, it is useful to measure the sentiment along these components.[76] For both R.syuzeht and the syuzhet implementation in Python, a similar method is used to extract the sentences from a long string of text. However, the Syuzhet package in R and the nltk.tokenize package in Python give slight differences in the amount of sentences that are calculated. However, this amount is often very small, making the differences negligible.

[76] M. Jockers, 'Revealing Sentiment and Plot Arcs with the Syuzhet Package', *matthewjockers.net*, 2 February, 2015. <https://www.matthewjockers.net/2015/02/02/syuzhet/> (12 September, 2022).

Using the VADER lexicon, a sentiment score is then assigned to each sentence, logically returning a list with the same length as the number of sentences that have been created by tokenizing the text. While this list provides some information on the sentiment of a text, they don't reveal anything about the narrative's organization of the distribution of positive and negative terminology in the text. The next step in creating a more comprehensive illustration of the progression of sentiment throughout the story, the sentiment values can be plotted in a bar graph, as exemplified for Oscar Wilde's *The Picture of Dorian Gray* (fig 3.1).

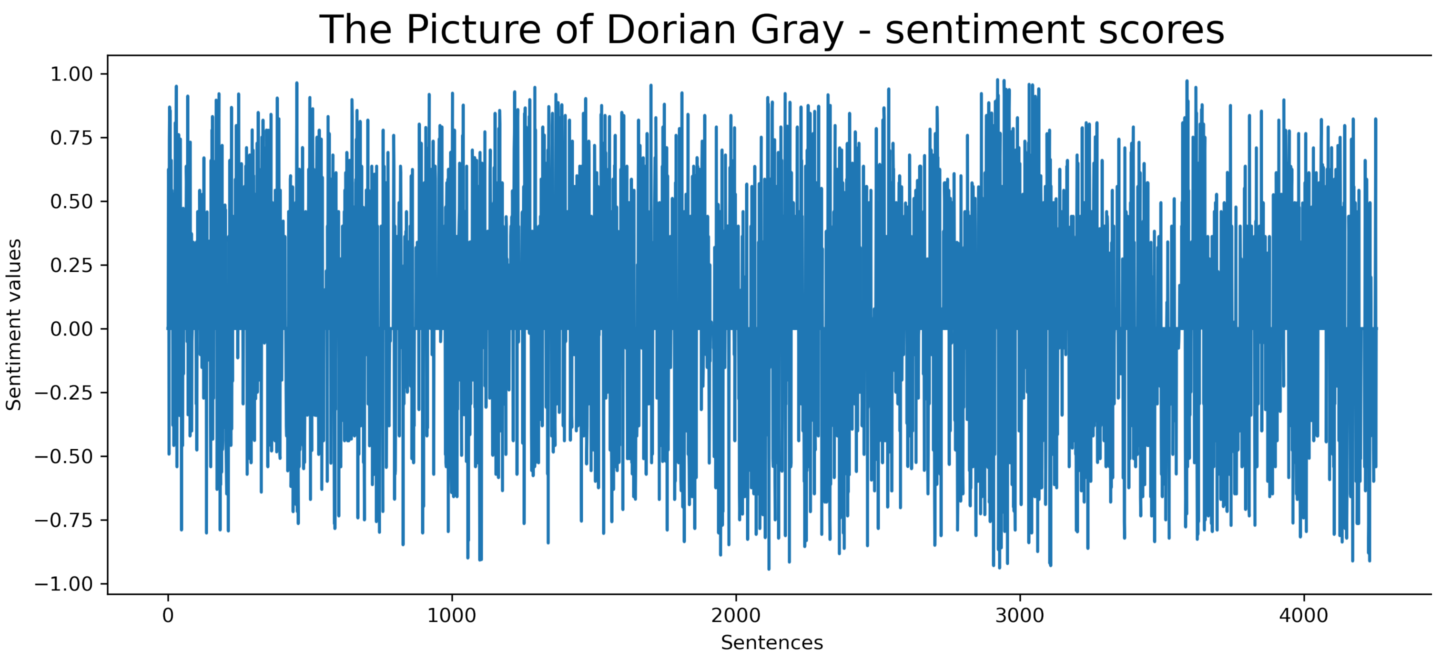

Figure 3.1: Sentiment scores for The Picture of Dorian Gray by Oscar Wilde. The sentiment is measured using the VADER lexicon.

As illustrated in figure 3.1, the scores that are produced vary strongly. It is hard to interpret the progression of sentiment throughout the novel meaningfully by means of this graph, since there is too much visual information. While this method of plotting may be informative when plotting sentiment for a short text sample, it is useful to smooth the data to display overall trends for a longer text. With R.Syuzhet, three distinct forms of smoothing algorithms, including a rolling average (also known as a 'moving average' or 'rolling mean'), *loess*, and a discrete cosine transformation, are used and compared in the result. The three methods of smoothing the data enable the possibility for comparing the various degrees of information. The later two smoothing functions still receive criticism as a result of their inaccuracy and simplification of results; the rolling averages approach is regarded to produce the most accurate results.[77] In order to test its effectiveness with a different lexicon (VADER) and with a different analysis method (M2), this thesis will adopt the rolling averages method.

Although the number of sentences differs between literary texts, it can be argued that the rough narrative structure of texts are similar in the sense that all narratives have starts, middles, and endings. They are intended to be linear narratives that should be read from the beginning to its conclusion. Since the texts that will be researched for this thesis differ in length, a method for comparison is required to recognize how similar one plot form is to another. This can be accomplished by further 'decimating' the rolling averages. Essentially, this means that some observations are chosen at certain intervals, and all others are ignored. With this method, some loss of information is inevitable, but the end product still allows for tracking when the scores increase and decrease. For this thesis, the rolling averages are decimated to return a list of 100 values. Plotting these values illustrates the progression of sentiment as shown in figure 3.2.

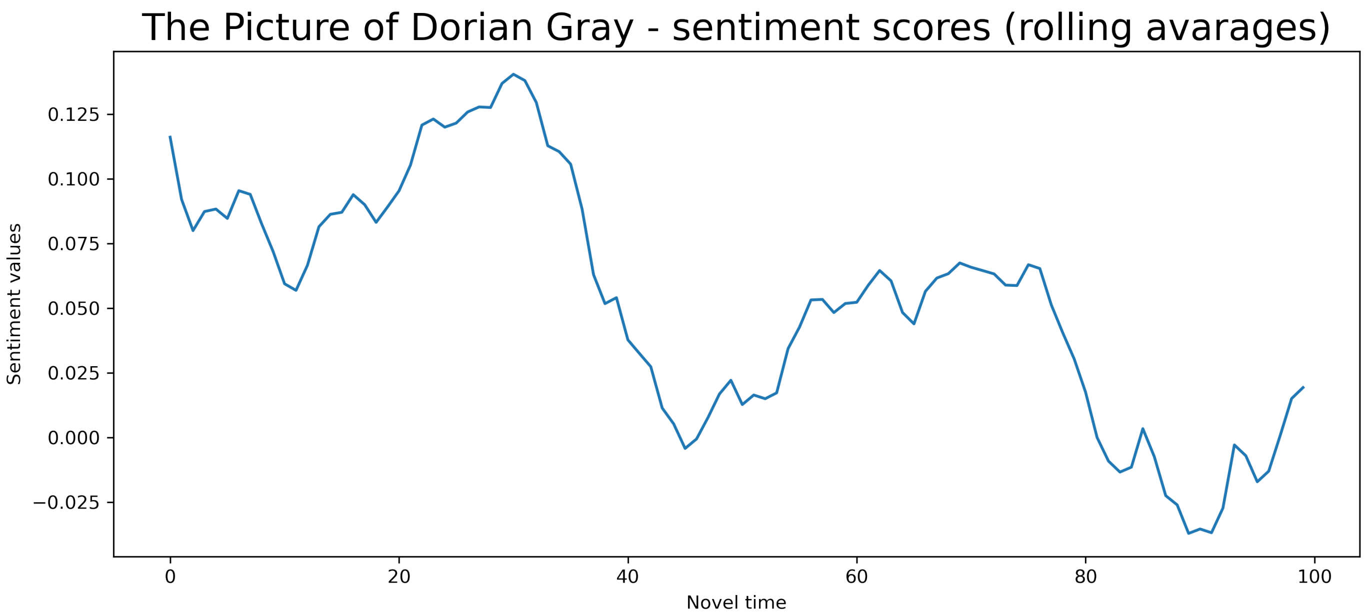

Figure 3.2: Sentiment scores for The Picture of Dorian Gray by Oscar Wilde, generated using the Syuzhet replication in Python. The sentiment scores are altered to show the decimated values of the rolling averages of the original scores.

[77] A. Swafford, 'Why Syuzhet Doesn't Work and How We Know', *Anglophile in Academia: Annie Swafford's Blog*, March 30, 2015. <https://annieswafford.wordpress.com/2015/03/30/why-syuzhet-doesnt-work-and-how-we-know/> (5 September, 2022).

To test whether the approach employed for this thesis can accurately recognize the progression of sentiment throughout a story, this thesis will first analyse five distinct stories, backed by close reading to control the analysis's efficacy. These five stories will all be plays by English playwright William Shakespeare, for multiple reasons. The plays that Shakespeare wrote in the late 16th and early 17th century are still a popular topic for research today, both in the humanities and the digital humanities. Shakespeare's importance in the digital humanities canon matches his renown in other fields such as education and the arts. As a results, almost all of his work is readily available online; from their original scanned and digitized versions, to versions that are marked up in XML. The broad availability of these plays, often freely accessible to be altered by means of text manipulation tools, is one reason to use these plays. The other reason is the variety of genres in which Shakespeare wrote. Shakespeare's plays are traditionally grouped into three categories: comedies, histories, and tragedies, all with their own distinct characteristics. A Shakespearean tragedy commonly includes an unhappy ending; the main character's demise must be included in the finale. A comedy, on the other hand, contains the polar opposite: the plays are mostly centred on love and passion, and feature a happy ending. By choosing plays from all three categories, the sentiment analysis programme designed for this thesis can be tested to recognise and correctly interpret varying degrees of emotional valence. Other than having a variation of the three categories into which Shakespeare's plays are generally categorised, the plays are randomly selected. The two comedies that will be analysed are *A Midsummer Night's Dream* and *The Taming of the Shrew*, the two tragedies are *Romeo and Juliet* and *Hamlet*, and the history play is *Henry V*. These five plays will be evaluated separately by means of close reading, as well as comparing the results to the sentiment analysis to that of Syuzhet, in order to determine the accuracy of the method employed for this thesis.

Given what Rebora stated, it becomes evident that the VADER and Syuzhet dictionary, which was constructed through crowdsourcing, is inherently less effective when it comes to analysing the sentiment of Shakespearean plays. While it may accurately capture the emotions experienced by present-day Internet users, its ability to convey the emotional nuances of Early Modern English plays is compromised. The English language has undergone significant transformations over time, particularly in semantics. Numerous words have evolved or fallen out of use in contemporary English. Consequently, comprehending Shakespeare's works can pose a challenge for modern readers due to these semantic shifts, and the same applies to the limitations of VADER and Syuzhet. Acknowledging this disadvantage, the study will take it into account and critically evaluate the extent to which the analysis tool can accurately interpret these 17th-century English plays. By employing a combination of close reading and sentiment analysis, the research aims to determine the method's accuracy and its potential shortcomings in grasping the true emotional essence of Shakespeare's works. The results will be thoroughly discussed, shedding light on the tool's effectiveness in capturing the intended sentiments expressed within these plays written in Early Modern English.

After determining whether the method used for this thesis is suitable for recognizing the progression of sentiment throughout a story, the study by Reagan et al. will be replicated. For each of the six shapes, as determined by Reagan et al., ten texts will be analysed and visualised.

3.3 Replicating Reagan et al.'s six basic emotional shapes

Reagen et al. have determined that a group of six fundamental emotional arcs serve as the fundamental buildings blocks of complex emotional trajectories. They found that, within a corpus of 1,327 texts acquired from project Gutenberg, three main sentiment structures can be established, with either a positive or a negative mode coefficient. This results in six 'bases' of emotional arcs. While not all 1,327 analysed stories could be clustered among these six shapes, the majority fell in the three modes that were determined by the principal component analysis. To replicate the study done by Reagan et al., 60 of the 1,357 researched texts will be analysed. For every emotional arc that Reagan et al. have identified (a total of six), ten texts will be chosen. For every emotional arc, the ten closest stories will also be plotted in the same graph in order to identify whether these stories follow a uniform shape, similar to the results by Reagan et al. These six graphs (together with their five closest text) can be seen in figure 3.3.

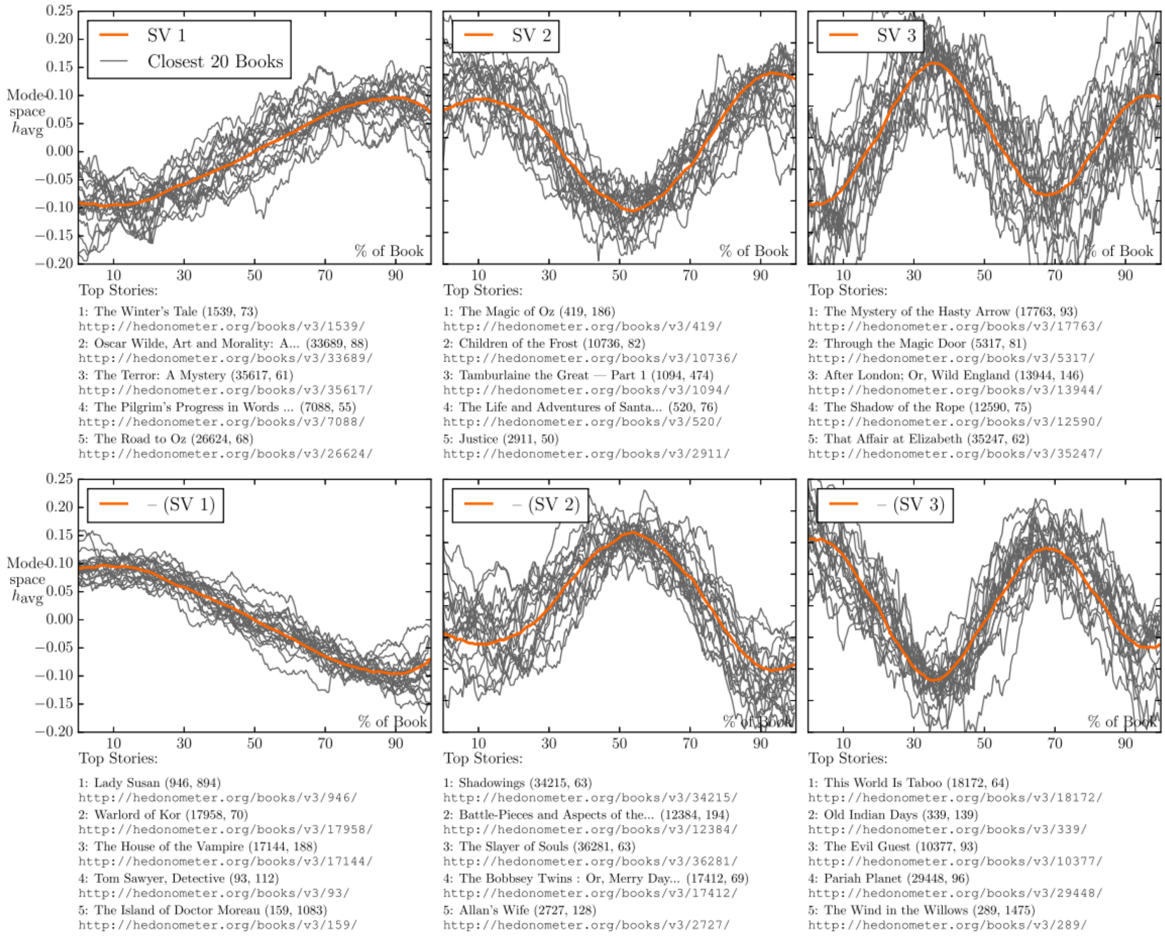

Figure 3.3: The six emotional arcs and their five closest stories. Via Reagan et al.

It must be mentioned that Reagan et al. made a selection from Project Gutenberg that consisted of more than just novels and plays. They also included texts that consisted of a collection of short stories, poems, essays etc. For this essay, an emphasis will be placed on texts that follow a linear storyline, as it is primarily interested in examining whether the aggregate sentiment scores of linear stories give information about the plot progression and emotional arc shape of a story. Since some of the ten closest stories that Reagan et al. found to be closets to certain emotional shapes are not linear stories, they will be excluded from the study. For example, this means that the second closest story to emotional arc 1, *Oscar Wilde, Art and Morality: A Defence of "The Picture of Dorian Gray"* by Christoper Sclater Millard will not be taken into the analysis, since this is a collections of reviews and letters on Oscar Wilde and his novel *The Picture of Dorian Gray*. In the appendix A, a complete list of the stories that are included in this thesis can be found.

The texts will be downloaded from Project Gutenberg, an online volunteer effort to digitize, archive and distribute literary works. Every text downloaded from the Project

Gutenberg website opens and closes with a so-called ‘boilerplate’, specifically modified to texts available on Project Gutenberg. Before the sentiment of the texts will be analysed, the boilerplates of these texts will be removed automatically, in order to prevent a distorted graph. Since the lines ‘start of the Project Gutenberg eBook’ and ‘end of the Project Gutenberg eBook’ are included in every text on the site, Python code has been written that can determine where these boilerplate sentences are located in each text and exclude them from the analysis.

To determine whether the sample of stories that have been selected for this thesis follow the same shape as depicted by Reagan et al., this thesis will make use of curve fitting. Curve fitting seeks for the ideal set of parameters for a specified function that best matches a given set of data points. To achieve this, the functional form of the mapping function will first be defined. The objective is to identify the model that most effectively captures the data that has been generated by the sentiment analyses. Three different mapping functions will be used in order to fit the shapes as described by Reagan et al. (table 3.1).

Emotional arc	**Shape characteristics**
Rags to riches	Rise
Tragedy	Fall
Man in a hole	Fall-rise
Icarus	Rise-fall
Cinderella	Rise-fall-rise
Oedipus	Fall-rise-fall

Table 3.1: The six emotional arcs as describes by Reagan et al., together with the characteristics that define their shape

The following definition of a straight line between inputs and outputs will be used to examine whether the first two emotional arcs (‘rags to riches’ and ‘tragedy’) follow an ascending or descending shape:

$$y = ax + b$$

This is a linear equation that results in straight line that is either increasing or decreasing. It is likely that a large number of these stories do not perfectly fit a linear style with a constant rise in sentiment values; some graphs in the first two emotional arcs will probably better fit in a polynomial line of degree two, in which a curve can be observed. However, the aim of this thesis is not to find a perfect fitting line, but to fit these data points to a model that recognizes whether the stories actually fit the 'rise' or 'fall' shape as described by Reagan et al. A linear polynomial is sufficient in providing this information, and will therefore be used for the first two emotional arcs.

The next two emotional arcs are 'man in a hole' and 'Icarus'. In this case, a polynomial regression of degree two will be used, in order to recognize whether the texts in these arcs actually follow the 'fall-rise' and 'rise-fall' shape as described by Reagan et al. For these texts, the following equation will be used to see whether these stories follow the graph of a quadratic polynomial:

$$y = ax2 + bx + c$$

By adding squared terms to the objective function, a polynomial regression can be created which follows the shapes of 'man in a hole' and 'Icarus' more fittingly. This shape will take into account the 'fall-rise' and 'rise-fall' sequence as described by Reagan et al. This is a second-degree polynomial function, the graph of which will be parabolic of shape.

For the last two emotional arcs, 'Cinderella' and 'Oedipus', a third degree polynomial will be used, also known as a cubic function:

$$y = ax^3 + b2 + cx + d$$

This function will be employed in order to analyse whether the last two emotional arcs described by Reagan et al. follow a 'rise-fall-rise' and a 'fall-rise-fall' shape.

By plotting the curve fitted line, it is possible to explore whether the analysed stories actually follow the shape as indicated by Reagan et al. However, it is also useful to examine how closely these stories follow this shape. In order to do so, this thesis will make use of the root-mean-square error (or RMSE). The RMSE is the square root of the residuals' variance. It represents the model's absolute fit to the data—how close the observed data points are to the predicted values of the model. RMSE is an absolute measure of fit, whereas R-squared is a relative measure.

To calculate the RMSE, the first step involves calculating the sum of squared errors (SSE). The SSE is computed by comparing each data point in the measured sentiment scores with the corresponding value from the fitted curve. It starts by setting the SSE to zero and then goes through each data point. For each point, the difference between the actual value and the value from the fitted curve is calculated, squared, and added to the SSE. The SSE represents the total accumulated squared differences between the observed data points and the fitted curve.

Once the SSE is obtained, the RMSE can be calculated by dividing the SSE by the number of data points, and then taking the square root of the result. The RMSE quantifies the average magnitude of errors between the original sentiment scores and the fitted sentiment scores. A lower RMSE value indicates a better match between the curve and the data.

The RMSE is a useful indicator of how well the model predicts the response. It allows for model evaluation and comparison, with lower RMSE values indicating better performance and a closer match between the model's predictions and the actual data.[78] A value of zero would indicate a perfect fit to the curve fitted shape. By means of plotting the shape of all 60 stories, as well as calculating the RMSE scores of each graph, it will be possible to indicate what stories follow the same shape as argued by Reagan et al., as well was which stories follow these shapes the most similar. These findings will reveal the extent to which Reagan's study can be replicated using a different approach of sentiment analysis.

[78] T. Chai and R.R. Draxler, 'Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature', *Geoscientific model development*, 7 (2014), pp.1247-1250.

4. Results

This section will present the findings of the sentiment analyses. The sentiment analysis results for five plays by Shakespeare will first be assessed, which will function as a performance analysis. Here, using the VADER sentiment lexicon, five well-known plays by Shakespeare are analysed in a method inspired by that of R.Syuzhet. The aggregate scores of the sentiment found throughout each play are visualized. Every sentiment graph for each play will be examined separately in respect to the plot to determine whether the graphs emotional peaks and troughs accurately depict the play's key moments. Next, the accuracy of these graphs will be discussed and interpreted, as well as compared to the analysis results retrieved with Syuzhet. To conclude this section, the shortcomings of analysing Shakespeare's plays by means of sentiment analysis will be discussed.

After that, once the sentiment analysis approach for this thesis has been determined to be functional, it will be utilised to replicate Reagan et al.'s study. For each sentiment 'shape', as determined by Reagan et al., the ten stories closest to this shape will be plotted in order to determine whether similar results are produced when a different analysis method is used.

4.1 Plotting the sentiment of Shakespeare's plays

Five of Shakespeare's well-known plays—two comedies, two tragedies, and one history—have been analysed in order to determine whether the VADER analysis method can correctly interpret the valence throughout a story. The examination of the two comedies, *A Midsummer Night's Dream* and *The Taming of the Shrew*, will be covered first. This will be followed by the tragedies *Romeo and Juliet* and *Hamlet*, and lastly by the sentiment graph of the history *Henry* V. Every graph depicts the progression of the sentiment throughout the play, with significant moments in the stories being annotated.

4.1.1 A Midsummer Night’s Dream

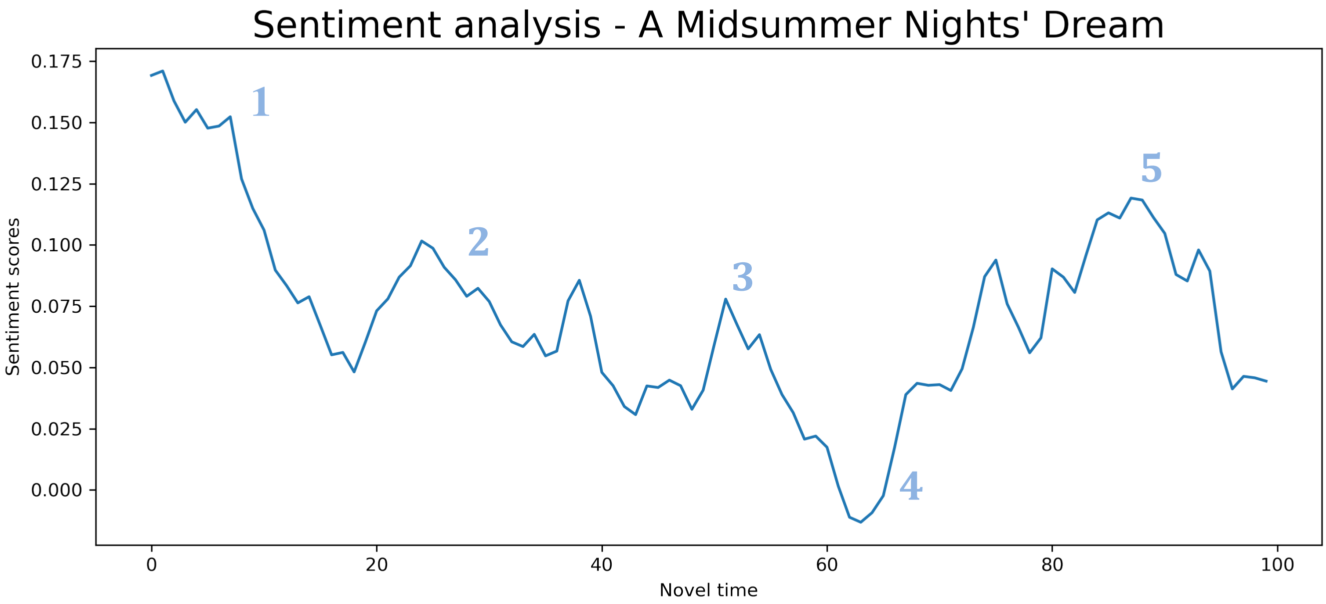

Figure 4.1: Progression of sentiment throughout *A Midsummer Nights’ Dream*.

1. For days before their wedding, Theseus and Hyppolyta are having a discussion. Theseus would rather marry now than keep waiting. The two talk lovingly to each other.
2. As a result of the potion, Lysanders fall deeply in love with Helena and praises her beauty.
3. Demetrius trips over while wandering in the woods. Since Helena is the first person he sees, the potion makes him instantly fall in love with her. A humorous scene ensues.
4. Hermia lashes out at Helena after being called ‘small’ for the second time, and a verbal and physical fight between Helena and Hermia ensues. In a fit of rage, Hermia threatens to scratch out Helena’s eyes with her fingernails, to which Helena flees.
5. The rude mechanicals perform *Pyramus and Thisbe*. The actors are doing their parts so poorly that the audience members laugh as if it were a comedy.

4.1.2 The Taming of the Shrew

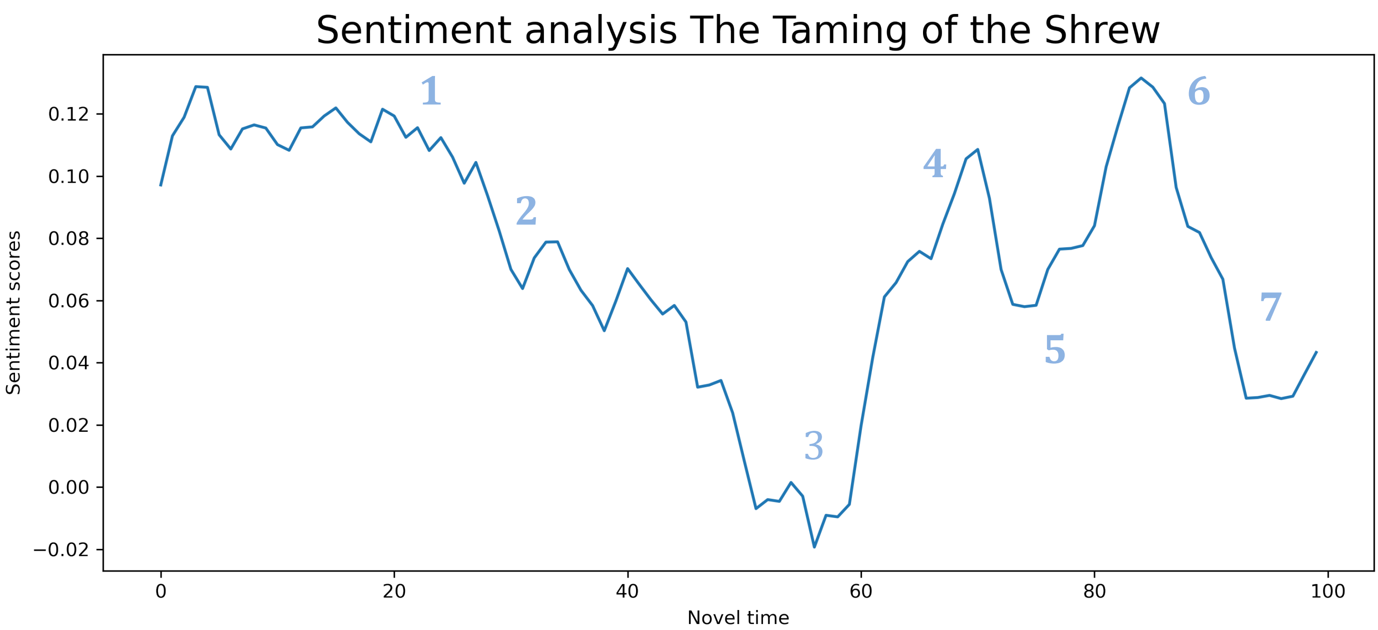

Figure 4.2: Progression of sentiment throughout *The Taming of the Shrew*.

1. Eager and confidant, Petruchio has travelled to Padua from Verona in search of a wealthy wife.
2. Tensions rise at Baptista's home as Katherine follows Bianca while furiously yelling at her. Because Bianca won't tell Katherine who of the suitors she likes, Katherine has tied Bianca's hands together and is attempting to beat her sister.
3. Petruchio starts “taming" his new wife in Verona. She is denied clothing and food since, in Petruchio's opinion, nothing is good enough for her. Petruchio scolds at his servants that the perfectly cooked meat is overcooked and a lovely outfit doesn't fit properly.
4. In Padua, Lucentio and Bianca flirt and kiss, while Hortensio and Tranio spy on them.
5. The play shifts to Petruchio and Katherine again, who are arguing on their way to the wedding. The “taming” of Katherine comes in to force, as she agrees with everything Petruchio says; ‘And be it moon, or sun, or what you please’.
6. Katherine appears to be completely captivated by Petruchio. Petruchio answers Katherine’s compliance lovingly.
7. A friendly argument between the three men regarding whose wife is the most obedient breaks out because it's still believed that Petruchio is married to a shrew. To the other

men’s surprise, Katherine is the only one of the three who comes when called, winning the bet for Petruchio.

4.1.3 Romeo and Juliet

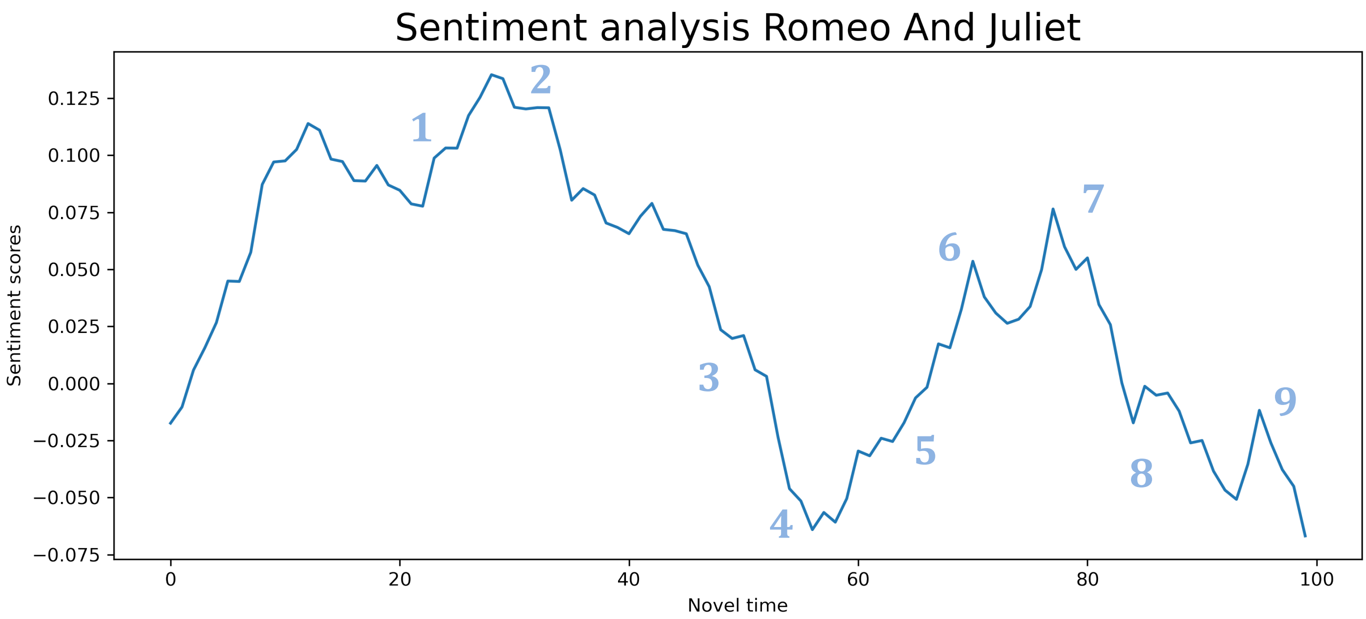

Figure 4.3: Progression of sentiment throughout *Romeo and Juliet*.

1. Romeo meets Juliet for the first time at the masquerade ball and the two fall in love.
2. The balcony scene: Romeo and Juliet declare their love for each other.
3. In a fit of rage, Romeo kills Tybalt. Romeo immediately proclaims his regret, as he realizes that killing his new wife’s cousin will have him banished.
4. Juliet learns that Tybalt has been killed, and Romeo banished; a low point in the sentiment graph.
5. Romeo and Juliet spend the night together, causing a rising line.
6. Juliet meets with Paris, who speaks lovingly of her. With a knife in her hand and threatening to commit suicide rather than wed Paris, Juliet begs Friar Lawrence for assistance.
7. Juliet drinks the potion that will make her fall asleep for 24 hours after speaking hopefully about the outcome of her and Friar Lawrence’s plan.
8. Juliet’s family finds her asleep, and presumes she’s dead.
9. Romeo kills Paris, and shortly after commits suicide in the tomb where Juliet lays. When Juliet finally wakes up and realizes what has happened, she too kills herself.

4.1.4 Hamlet

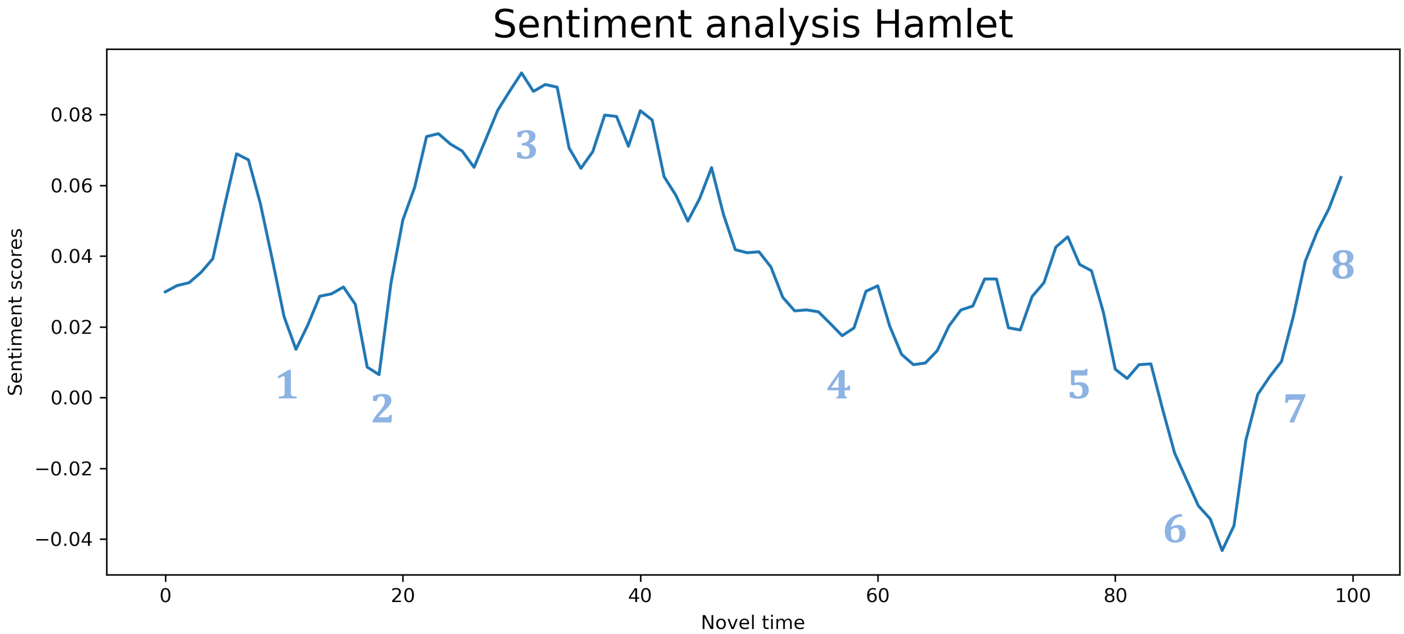

Figure 4.4: Progression of sentiment throughout *Hamlet*.

1. The ghost of Hamlet's father first appears to Bernardo, Marcellus and Horatio. The soldiers urge Horatio to confront the ghost as they draw their swords and cower in fear.
2. When Hamlet discovers the truth from the ghost of his recently departed father, he is furious: the late monarch was actually murdered by his own brother, the current King Claudius.
3. Plonius describes how he believes that Hamlet has gone crazy as a result of his love for Ophelia.
4. Hamlet kills Polonius.
5. Hamlet realizes that Ophelia is dead, and the grave he's came across is meant for her corpse. Laertes, overcome with grief, accuses Hamlet of being the reason of her death.
6. Horatio is told by Hamlet in the castle's hall how he exposed the king's plan against him and defeated Rosencrantz and Guildenstern.
7. During his duel with Laertes, Hamlet gets cut by a poisoned blade. After the death of the Queen, Laertes and Claudius, Hamlet dies in Horatio's arms.

8. Fortinbras, having just arrived at the palace, realizes that the entire Danish royal family has died. He seizes the throne for himself and requests a military funeral for Hamlet

4.1.5 Henry V

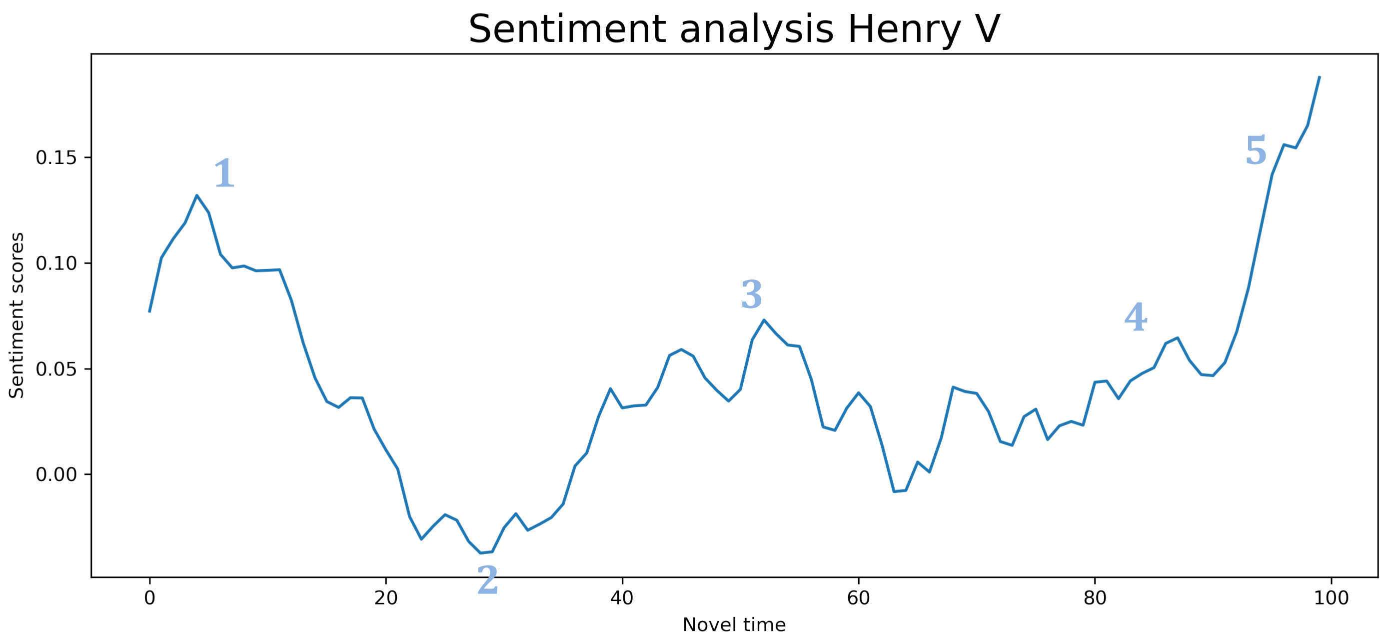

Figure 4.5: Progression of sentiment throughout *Henry V*.

1. The Canterbury bishops outline King Henry's character and explain how, almost miraculously, he has transformed to a graceful monarch. They offer him money to fund a war.
2. The action shifts to a London pub where Henry's old friends have a discussion. Nym and Bardolph talk about Pistol's recent marriage to Mistress Nell. Pistol and Mistress enter and Nym and Pistol draw swords on each other. This shift in the setting of the play shows the effect of the war around the country.
3. The night before battle, Henry disguises himself in a long cloak and visits his soldiers to check on their morale after realizing the disadvantage his army is in. He visits every one of his soldiers, addresses them as brothers, and lifts their spirits.
4. After battle, Henry finds out about the numerous French casualties and the small number of English. He claims that God alone is responsible for the English victory.
5. King Henry has travelled to King Charles VI's palace in France to speak with him and discuss a lasting peace. When Henry and Catherine, Charles' daughter, are left alone together with only Catherine's maid left to help translate, a comic scene ensues. Henry

makes an effort to convince her to marry him. While he lacks in French vocabulary, Catherine ultimately understands and accepts his offer.

4.2 Evaluation

Based on a comparison of the sentiment scores to a close reading of each play, the sentiment analysis utilised for this thesis appears generally successful at illustrating the progression of sentiment throughout a text in the form of a graph. Nonetheless, there are some outliers in the graphs that are worth mentioning. This section will focus on how the analysis successfully depicted the sentiment across these stories, as well as where the study revealed surprising results.

4.2.1 Assessing the sentiment scores of Shakespeare's plays

Table 4.1 shows the average sentiment score for each play. It is worth noting that these scores have been calculated by dividing the summation of all sentiment scores per play by the amount of analysed sentences. These analysed texts are all plays, and every section of dialogue starts with the name of the character that is speaking. In tokenizing the texts, every character name is being recognized as a sentence, and results in a sentiment score of 0. The scores attached to the character names have a minimal effect on the sentiment graphs, and have therefore not been removed from the final analysis. However, including these "sentences" into the calculation of the entire play results in a lower overall sentiment score than when these character names would have been eliminated, since the summation of scores remains the same.

Play title	Overall sentiment score
A Midsummer Night's Dream	0.07698
The Taming of the Shrew	0.07420
Henry V	0.05456
Hamlet	0.03419
Romeo and Juliet	0.02454

Table 4.1: Overall sentiment score for the five analysed plays in descending order

As can be expected, the two comedies, *A Midsummer Night's Dream* and *The Taming of the Shrew*, show the highest sentiment score. The sentiment starts high during the exposition of the characters and setting at the beginning of both plays. As action starts to rise, the sentiment starts to decrease. As the plays reach their conclusion, the sentiment starts to rise again, and ends in a happy ending. Nevertheless, both comedies show a decrease in sentiment at the end of each graph, indicating that the happy ending is not the highest point of sentiment in both plays. The explanation for this surprising decrease in sentiment differs between both plays. While *A Midsummer Night's Dream* does have a happy ending (as the majority of the characters has a positive conclusion) the sentiment decreases quite a bit towards the end as a result of the use of language during the play performed at the end: *Pyramus and Thisbe*. In the narrative of *Pyramus and Thisbe*, two young lovers are forbidden from being together because their families are rivals. The tragic end of this play results in the use of the words such as 'sad', 'death' and 'die'. This results in an abundance of negative sentiment scores, reducing the overall average. As for *The Taming of the Shrew*, the decrease in sentiment results from the bet between the three married men. Since Katherine has won the bet for Petruchio, the other two men are extremely disappointed. Their use of language in this final scene is what decreases the sentiment of this happy ending.

Henry V, although not a comedy, shows a similar progression of sentiment. The sentiment starts to lower as Henry foils an assassination attempt, delivers inspiring speeches, and outduels his opponent. The sentiment reaches a high as Henry's marriage unites the two nations. For these three plays, all peaks and troughs in the graph can be logically attributed to the use of language occurring at those points in the play.

Romeo and Juliet and *Hamlet*, the two tragedies, show the lowest measured sentiment scores. Surprisingly, these plays still show a positive sentiment score on average. The sentiment scores assigned by VADER range from -1 to 1, with 1 being the most positive and -1 being the most negative. For a play like *Romeo and Juliet*, which is known for its tragic story arc, an overall negative sentiment score would not have been unexpected. Some occasionally surprising high sentiment scores can also be seen in figure 4.3 as well. Point 6 and 7 take place between the first and third scene of the fourth act. Crucial events in these scenes are Juliet's threat to kill herself if Friar Lawrence cannot save her from marrying Paris, and Juliet drinking the potion which would help her set her death in scene. In spite of these dramatic events, the analysis gives a positive sentiment for these scenes, and even appears as two peaks in the sentiment graph. There are multiple reasons for this seemingly mismatch between the action in the plot and the measured sentiment score. For one, a large portion in act four consists of conversations between Juliet and the nurse, and Juliet and Paris. During the first conversation, it becomes apparent that the nurse does not mourn the banishment of Romeo, and tries to convince Juliet that Paris is an absolute gentleman and a much better suitor. During her conversation with Paris, Paris talks to her in a kindly but slightly arrogant manner. Juliet answers neutrally, expressing neither love nor hate towards him. When taking a closer look at the sentiment score per sentence, it becomes clear that the main positive sentiment comes from Paris' lines, whereas Juliet's lines mainly present a neutral score. Therefore, while these scenes present a personal low point for one the protagonists of *Romeo and Juliet*, the dialogue from the supporting characters lift up the overall sentiment score. Secondly, apart from the dialogue increasing the overall sentiment score, the VADER sentiment analysis also appears to struggle in recognizing the overall negative sentiment in the Shakespearean use of language in these scenes. Table 4.2 shows the sentiment score per sentence for the section in which Juliet threatens to kill herself to Friar Lawrence if he refuses to help her.

Lines said by Juliet (Act 4, scene 1 *Romeo and Juliet*)	Sentiment score
Tell me not, Friar, that thou hear'st of this, Unless thou tell me how I may prevent it.	0.0258
If in thy wisdom, thou canst give no help, Do thou but call my resolution wise, And with this knife I'll help it presently.	0.8508
God join'd my heart and Romeo's, thou our hands; And ere this hand, by thee to Romeo's seal'd, Shall be the label to another deed, Or my true heart with treacherous revolt Turn to another, this shall slay them both.	0.9520
Therefore, out of thy long-experienc'd time, Give me some present counsel, or behold 'Twixt my extremes and me this bloody knife Shall play the empire, arbitrating that Which the commission of thy years and art Could to no issue of true honour bring.	0.5147
Be not so long to speak.	0.0000
I long to die, If what thou speak'st speak not of remedy.	-0.5994

Table 4.2: Sentiment score per sentence for Juliet's conversation with Friar Lawrence

While this section of dialogue presents a significant moment of despair for Juliet, the overall score shows a positive sentiment. The third sentence, in which Juliet states that she will kill both her hand and heart before being joined to another in marriage, results in a particularly positive score (0.9520). Only her last sentence, in which she states her intentions in a more literal manner ('I long to die'), gives a negative sentiment score.

An essential principle in writing is 'Show, Don't Tell,' which urges authors to vividly depict emotions and experiences through actions, descriptions, and dialogue rather than explicitly stating them. Shakespeare frequently employs this technique, often expressing characters' inner turmoil and intentions through poetic and metaphorical language. Consequently, sentiment analysis tools like VADER may struggle to discern the true sentiment of individual sentences in Shakespearean plays. The emotions of literary characters are intricately intertwined with the poetic style of the work, necessitating a deeper understanding of the literary techniques employed to accurately interpret their motivations and intents.

A similar occurrence can be found in Hamlet. In a particularly grim moment for Hamlet, he expresses his longing to die in the following manner: 'O, that this too solid flesh would melt. Thaw and resolve itself into a dew! Or that the Everlasting had not fix'd his canon 'gainst self-slaughter! O God! God!' This soliloquy too receives a positive sentiment score, despite its sombre content.

This demonstrates how it can be challenging for VADER to analyse the sentiment for individual sentences in Shakespearean plays, since the motivations and intents of characters are frequently conveyed in a more poetic style of writing. Shakespeare's use of 'Show, Don't Tell' allows readers to experience the characters' emotions through vivid imagery and expressive language, evoking a profound connection. However, it poses difficulties for sentiment analysis tools that rely on more direct and literal language to determine sentiment.

Similar to *Romeo and Juliet*, *Hamlet* has an overall positive sentiment score, despite being a tragedy (table 4.1) Furthermore, despite the play's tragic climax, Hamlet shows an unexpectedly steep rise in sentiment at the play's conclusion (figure 4.4). This rise occurs in the second scene of the fifth act, a particularly violent scene in the play. Characters are stabbed, poisoned, and killed off one after the other as the theme of revenge comes to a close. The average sentiment score of this final scene is 0.0405, which is higher than the average of the complete play. When inspecting the play sentence for sentence, it becomes clear that a large portion of this high sentiment score results from the dialogue between Hamlet and other supporting characters. While Hamlet and Horatio discuss the execution of Rosencrantz and Guildenstern, Hamlet appears rather indifferent, as he believes they deserved it. Moreover, the frequent mention of positive words such as 'heaven', 'heart', and 'love' give an overall positive score to this section. In the following conversation, Osric makes an effort to flatter Hamlet by agreeing with everything he says, and shortly after, he ecstatically praises Laertes. In his conversation with Laertes, in which Hamlet asks for forgiveness, words such as 'pardon', 'honour' and 'love' again cause high sentiment scores. The duel largely renders neutral scores, and while the death of the Queen, Claudius and Hamlet results in some negative scores, these scores are low in quantity. This shows that even though a story ends tragically, this does not necessarily correspond with a particularly low sentiment score.

It is worth mentioning that the form in which Hamlet is written may also have an effect on this seemingly contradicting result. Since the vast majority of a play consists of dialogue, less descriptions and characterisations are written than in, for example, a novel. It is likely that a lower sentiment score would have been found at the conclusion of Hamlet if the deaths of the characters had been written with more use of description and exposition. However, more study must be conducted to conclusively determine this.

4.2.2 Comparing the results produced by VADER to Syuzhet

The method used for this thesis is largely based on the R package Syuzhet, created by Matthew Jockers. One of the objectives of this thesis is to check whether a similar method to Syuzhet can be created using the VADER tool in Python. Syuzhet is the most popular sentiment analysis tool when it comes to literary research, and has been verified by literary academics through a comparison of computational sentiment evaluations.[79] In this section, it will be checked whether the aggregates of the sentiment scores retrieved by VADER portray a similar progression of sentiment as the analysis made in Syuzhet. A side-by-side comparison of the graphs resulting from Python and from R can be found in the appendix (Appendix B).

As can be seen in figure 8.1 to figure 8.10, the graphs show a very similar progression of sentiment throughout all five plays. While the default lexicon for VADER was meant to analyse social media texts, and while this lexicon is significantly smaller than the default Syuzhet lexicon, both give a nearly identical progression of sentiment throughout all plays. Nonetheless, the differences in both lexicons do give some noteworthy outliers.

For one, the Syuzhet sentiment analysis finds a (slight) dip in sentiment at around act three scene two in *A Midsummer Night's Dream*, whereas VADER finds a slight peak here. The difference is negligible, but it is still interesting to observe how the two different lexicons may give the same statement a different interpretation. In this specific scene, Oberon wonders if Titania has awakened yet and who or what she may have first seen. When Puck shows up, he informs him reluctantly that Titania has fallen in love "with a monster". Oberon's plan is working better than he initially expected. The results from these lines can be seen in table 4.3.

---

[79] K. Elkins and J. Chun, 'Can Sentiment Analysis Reveal Structure in a Plotless Novel?', *arXiv preprint arXiv*, (2019).

Lines said by Puck (Act 3, scene 2 of *A Midsummer Night's Dream*)	VADER sentiment score	Syuzhet sentiment score
My mistress with a monster is in love.	0.6369	-0.25
Near to her close and consecrated bower, While she was in her dull and sleeping hour, A crew of patches, rude mechanicals, That work for bread upon Athenian stalls, Were met together to rehearse a play Intended for great Theseus' nuptial day.	0.2023	-1.15
The shallowest thick-skin of that barren sort Who Pyramus presented in their sport, Forsook his scene and enter'd in a brake.	0.0000	-0.5
When I did him at this advantage take, An ass's nole I fixed on his head.	0.3071	0.25
Anon, his This be must be answerèd, And forth my mimic comes.	0.0000	0
When they him spy, As wild geese that the creeping fowler eye, Or russet-pated choughs, many in sort, Rising and cawing at the gun's report, Sever themselves and madly sweep the sky, So at his sight away his fellows fly, And at our stamp, here o'er and o'er one falls; He murder cries, and help from Athens calls.	0.8126	-3.1
Their sense thus weak, lost with their fears, thus strong, Made senseless things begin to do them wrong; For briers and thorns at their apparel snatch; Some sleeves, some hats, from yielders all things catch.	0.7783	-3
I led them on in this distracted fear, And left sweet Pyramus translated there.	0.4497	-0.1
When in that moment, so it came to pass, Titania wak'd, and straightway lov'd an ass.	0.5423	-0.5

Table 4.3: sentiment scores assigned by VADER and Syuzhet for *A Midsummer Night's Dream*

Overall, the VADER lexicon assigns positive sentiments to Puck's speech, whereas Syuzhet assigns almost exclusively negative sentiment. The question is, which lexicon interprets these lines correctly? Puck's dialogue results from his and Oberon's devious plan in making Titania fall in love with a donkey. Puck is speaking with excitement since he followed Oberon's instructions correctly. Nevertheless, his text is full of words that can be interpreted in negative sense, such as 'creeping', 'murder' and 'senseless'. This specific section of the play shows how the analysis of emotion can be difficult for a program to interpret: it cannot separate narratological context from use of language the same way a human reader could.

The other striking difference between two graphs occurs in Hamlet. Right after Hamlet accidentally kills Polonius, VADER's graph only dips slightly. Syuzhet's graph on the other hand takes a significant decline. What follows after Polonius' death is a conversation between Hamlet and his mother. He passionately asks her why she chose to wed a corrupt man like Claudius and makes angry remarks about the superiority of his father to his uncle. His mother begs him to stop. Hamlet continues to criticise her and rage against Claudius when all of a sudden, his father's ghost reappears in front of him. This is one of the few instances in which Syuzhet recognizes that the situation as more negative than VADER does: a likely result from the smaller lexicon that VADER uses. VADER finds a lot more neutral sentences, whereas Syuzhet recognises a lot more negative sentences, since VADER cannot recognize as many words as Syuzhet.

While the differences in both lexicons can produce a contrast in the analysis graphs (as shown in *Hamlet*), the results show that this rarely happens. Nevertheless, the difference in lexicon size (around 7,500 for VADER compared to 14,000 words for Syuzhet) does show their effects on the maximum and minimum score for the sentiment graphs. Taking *Romeo and Juliet* as an example, the Y-axis shows a sentiment range from -0.075 to 0.125 for the VADER graph, and -0.3 to 0.2 for Syuzhet. While the overall shape is the same, these differences become more apparent when both datasets are plotted in the same graph. The overlaid plots of the sentence based sentiment values for *Romeo and Juliet* in figure 4.6 display very comparable distributions, means and variances. However, the VADER histogram shows substantially smaller ranges, with the exception of the region surrounding neutral emotion. Since any word not found in VADER's lexicon will be considered neutral, *Romeo and Juliet* has a larger amount of sentences being labelled as such when using VADER in comparison to using the Syuzhet lexicon (60,8% for VADER versus 47,6% for Syuzhet).

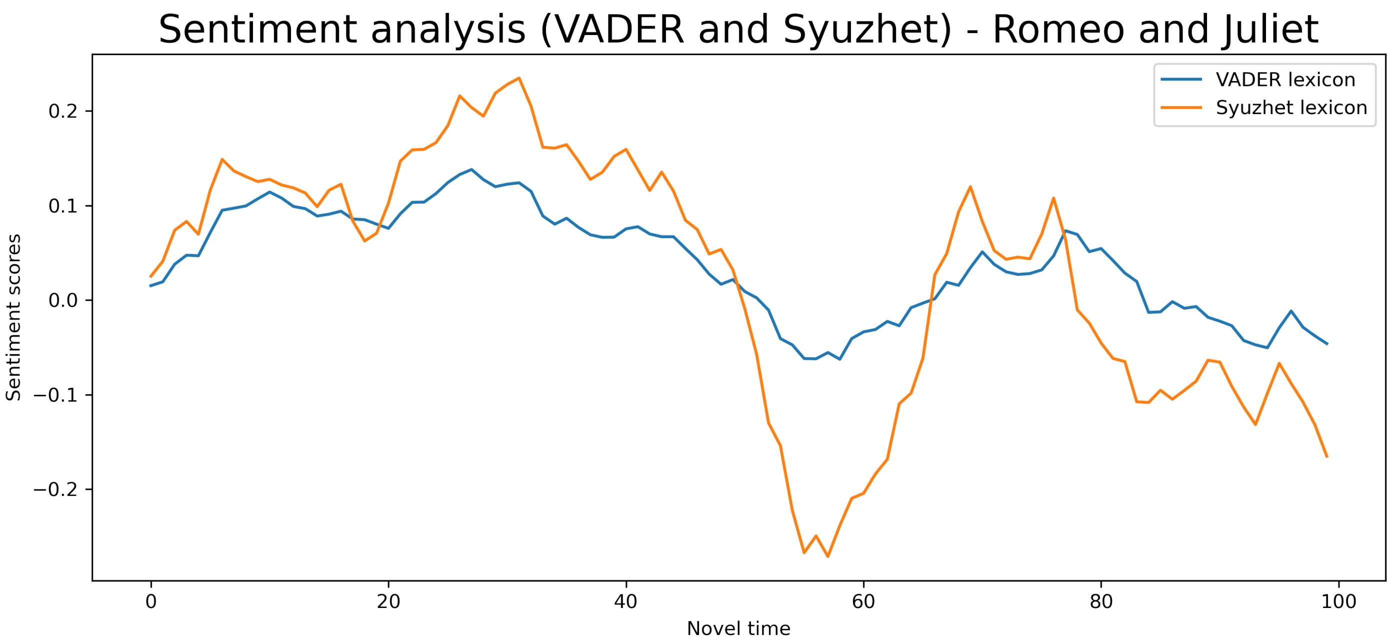

Figure 4.6: a comparison of the sentiment found by VADER and Syuzhet in *Romeo and Juliet*

In addition to using a different lexicon for this thesis, VADER additionally uses a different method to conduct sentiment analysis. By using syntactic structure analyses, in comparison to simple word count, VADER can detect valence shifters and intensifiers. A sentence such as 'I don't love you' will be determined to be positive by Syuzhet, since it gives a score to each word individually. VADER can however recognize the negation present in this sentence, and gives the sentence a negative score. This thesis proposed that, since VADER has a slightly more advanced method for analyzing sentiment, it could produce a more accurate result from the sentiment present in the analyzed plays. In reality, these occurrences of negation are so rare that these present hardly any difference on the eventual sentiment graph. Nevertheless, when analyzing the stories more closely, on a sentence-to-sentence basis, these differences become more apparent, as shown in table 4.4.

Play and act	Line	VADER sentiment score	Syuzhet sentiment score
***Hamlet* (Act 1 scene 2)**	'My father's spirit in arms! All is not well'	-0.1043	1.3
***Romeo and Juliet* (Act 2 scene 5)**	'I'faith, I am sorry that thou art not well.'	-0.2764	0.9
***The Taming of the Shrew* (Act 1 scene 1)**	'I' faith, sir; you shall never need to fear'	0.6628	-0.5

Table 4.4: Scores given by Syuzhet and Vader for lines that contain negation

Figure 4.4 demonstrates that, in contrast to Syuzhet, VADER does in fact identify the negation in these sentences. This can make the difference between a phrase being positive or negative. This means that VADER will yield more accurate results if the goal is to analyse these plays on a more detailed basis (sentences in contrast to the entire play).

Overall, these differences between the VADER and Syuzhet graphs are very minimal. Based on these findings, it can be stated that both tools produce almost identical shapes when it comes to the progression of sentiment throughout a story. Nevertheless, it should be emphasised that, when examining the findings sentence-by-sentence, the disparities between both methods become more apparent. Syuzhet can often accurately recognize the sentiment of more sentences as a result of its larger lexicon, whereas VADER labels a sentence more often as neutral. The VADER analysis method is on the other hand better in recognizing negation in sentences, which Syuzhet cannot do.

4.2.3 Limitations

As already mentioned in §4.2.1, the VADER method does show some weaknesses when it comes to assigning scores to the events occurring in Shakespeare's plays. Table 4.2 shows that it can be difficult for VADER to give deeper meaning to a sentence beyond the words that it contains. Shakespeare's poetic writing style can often conceal the literal connotation of a sentence, making it difficult for a sentiment analysis tool to extract the emotion expressed by a character. When it comes to more literal sentences, VADER has less difficulty in assigning a sentiment score. The difficulty in interpreting this use of language becomes apparent in sections such as the fourth act of *Romeo and Juliet*, and the climax of *Hamlet*.

Moreover, the different levels of focalisation can occasionally lead to quick variances in the sentiment scores of the narrative. While these striking changes in variances are not an inherent flaw of sentiment analysis, they are worth mentioning as a point of attention for correctly interpreting these graphs. For a large part of the plays, this change in focalisation does not get falsely interpret by VADER; after all, we ask the program to analyse the entire play linearly. Nevertheless, Shakespeare regularly transitioned between characters' points of view in his plays. Without taking these transitions in mind, it is hard to make sense of the graphs of *Henry V* and *A Midsummer Night's Dream*. Point 2 in in figure 4.5 shows a striking decrease in sentiment in *Henry V*. This trough is the result of diverging from Henry V as the main protagonist; this section of the play takes place in a pub in London with a variety of different side-characters, who discuss the on-going war. The same goes for *A Midsummer Night's Dream*; part of the tragedy *Pyramus and Thisbe* (essentially a play within a play) decreases the sentiment of the end of this comedy, which is known for their happy endings. While the tragic 'play within a play' consists of many words that are measured with low sentiment scores, the overall mood of the scene is positive; it is a humorous moment. Since the sentiment analysis tool can merely measure the sentiment of the words present, the context can often get lost, and these moments can appear to be more negative than they are in reality.

Aside from these limitations, the sentiment analysis method used in this thesis' study of Shakespeare's plays has a few other drawbacks that are less apparent when seen in a graph but are nonetheless important to note when analysing the results sentence-by-sentence.

As Rebora has stated, the method that was used to build the VADER dictionary makes it an inherently less effective tool for analysing the sentiment of Shakespeare's plays. The VADER dictionary, which was created by means of crowdsourcing, may successfully represent the emotions experienced by current-day Internet users, but is less effective in capturing the emotional representation of a an Early Modern English play. Over time the English language has changed in multiple aspects, such as in semantics. There are many words whose meaning have evolved over time, or have stopped being used in present-day English all together. Shakespeare may sometimes be challenging to understand for contemporary readers due to these semantic shifts, and the same applies to VADER. Some examples can be found in *Romeo and Juliet*. When Romeo has killed Mercutio in the third act, the Prince decides to spare his life, as Romeo was not the one who incited the violence. Nonetheless, he does decide to banish Romeo: 'But I'll amerce you with so strong a fine, That you shall all repent the loss of mine.' The word 'amerce' is no longer used in today's English and essentially means 'to punish with a fine'. This line said by the Prince receives an overall positive sentiment score as the VADER lexicon does not recognise the word 'amerce'.

Substituting this word with 'punish' immediately lowers the overall sentiment of this line, as this word is recognised as negative by VADER's lexicon. The same applies for the term 'betossed' as Romeo states the following line: 'What said my man, when my betossed soul, Did not attend him as we rode?' In this case, 'betossed' means concerned or stressed, yet VADER recognizes this sentence as neutral.

Since the analysis is reliant on the terms found in VADER's lexicon, it is unable to identify the valence of sentences in which the words are not recognized. But even when VADER does recognise all the words in a phrase, it may have trouble understanding its tone. As indicated by Rebora, sentiment analysis tools have difficulties recognizes certain rhetorical devices such as sarcasm or irony. Hamlet uses sarcasm on several occasions throughout the play. By doing this, he is able to convey his feelings of contempt for his uncle Claudius. One famous example is the very first line said by Hamlet in the play: 'A little more than kin, and less than kind.' Hamlet says this to Claudius, and he basically states that the two are a little more than just blood relatives now that Claudius has married is mother. However, they do not belong to the same kind, as Hamlet is concerned. Hamlet also expresses his anger at the new king for his hasty marriage to the queen. Here, 'kind' has the meaning of 'polite' or 'respectful'. This line shows the snarky and opinionated side of Hamlet, and serves as a means of showing that Hamlet sees Claudius as far from friendly. Nevertheless, this sentence receives a positive sentiment score, despite its sarcastic tone. Later in the play, during a conversation with Ophelia, Hamlet says the following: 'O God, your only jig-maker! What should a man do but be merry? For look you how cheerfully my mother looks, and my father died within's two hours.' Using sarcasm once more, Hamlet implies that Ophelia is ignorant for failing to see that he is unhappy and that it is only reasonable for him to feel this way. 'What should a man do but be merry?' is a sarcastic question that VADER is unable to recognise. Given that this sentence is taken literally, it receives a high sentiment score.

Overall, the VADER sentiment analysis appears to be fairly accurate in visualizing the progression of sentiment throughout these plays. The results are almost identical to the results found by Syuzhet, which is a text analysis tool that has been proven by academics to accurately reflect the development of emotion throughout the course of a text. The general highs and lows of the sentiment scores can be recognised in key-moments in the play by comparing the graphs to the plot. Nevertheless, some unexpected results were observed as well. These unexpected results are largely influenced by Shakespeare's approach to writing: his lyrical and poetic style of writing frequently hides a sentence's plain meaning to VADER. As a result, VADER has difficulties in recognizing the underlying tone, which gives unexpected results in the fourth act of *Romeo and Juliet* and the climax of *Hamlet*. The 'happy endings' of the two comedies are not recognized in the sentiment analysis graph either. While the ending of these two plays can considered being positive, these dips in the graph are not the results of the analysis failing to correctly interpret the sentiment of the sentences. The 'happy ending' in *A Midsummer Night's Dream* consists largely of negative words as a result of the 'play within a play' that takes plays towards the conclusion; while the ending of this play is still joyful, the context or overall mood of a scene cannot be measured beyond the words used. Furthermore, VADER performs less well when individual lines are considered. This is based on the use of archaic words, the use of sarcasm and irony, and again the Shakespearean style of writing. This demonstrates once more that VADER performs better with lengthier texts and when seen from a broad perspective as opposed to a zoomed-in one.

4.3 Replication of Reagan et al.'s sentiment arc shapes

According to Reagan et al., who examined a large number of Project Gutenberg manuscripts, the bulk of these texts can be categorized into six basic shapes. Reagan et al. achieved this result by first identifying the most prevalent shapes (fall-rise, rise-fall-rise, etc.), and then designing a line that best matched these shapes. As determined in §4.1 and §4.2, the sentiment analysis method employed for this thesis is, despite its flaws, a functional method in analysing the progression of sentiment throughout a story. This VADER analysis method will now be used to replicate Reagan et al.'s study, in order to determine whether identical results can be obtained using a different analysis method.

4.3.1 ‘Rise’ and ‘fall’ shapes

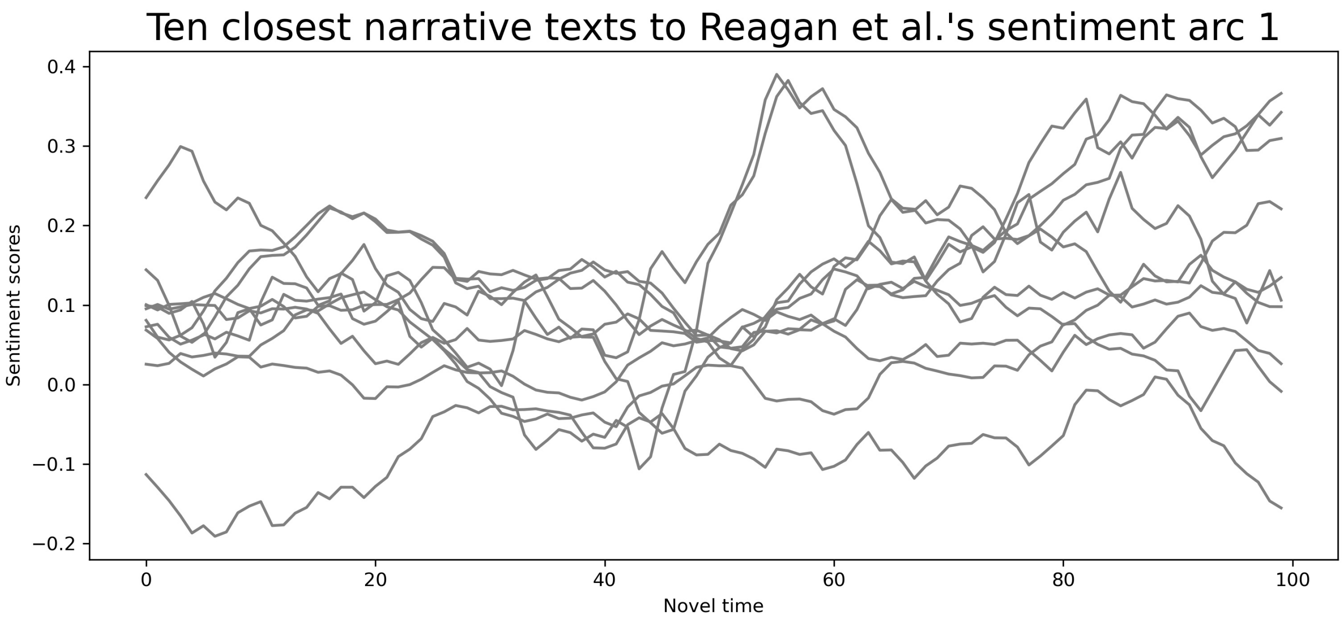

Figure 4.7 : The ten closest narrative texts to Reagan et al.’s arc shape 1, named ‘Rags to Riches’ (rise), analysed by VADER

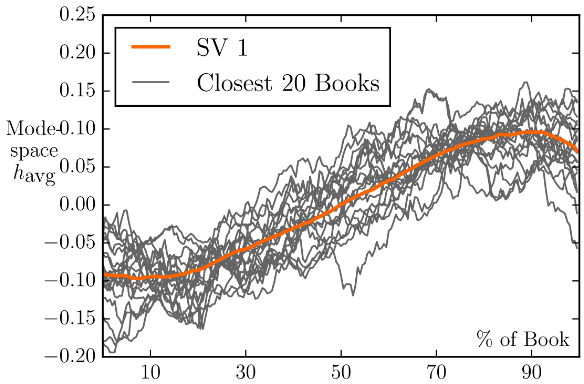

Figure 4.8: Reagan et al.’s ‘Rags to Riches’ arc shape and the 20 closest texts

Figure 4.7 shows the ten narrative texts, analysed using VADER, that fit Reagan et al.’s first sentiment arc shape (‘Rags to Riches’ or a ‘rise’ shape) the closest. It becomes apparent that these ten texts do not fit a similar line as well as figure 4.8, in which Reagan et al.’s results are presented. The graph lines show rather different trajectories, with the final data points showing a wide range of sentiment scores between -0.2 and 0.4. This shows that the scores resulting from the VADER analysis differ quite a lot from those made with the Hedonometer by Reagan et al.

The subsequent step involves determining which analysed stories by VADER closely align with the 'rise' shape, as described by Reagan et al., and which deviate the most. To accomplish this, curve fitting is utilized, aiming to find the most suitable mathematical function to fit a group of data points. The goal is to fit the sentiment values of the stories to different types of curves in order to identify recognizable patterns in emotional arcs. It is important to note that while the stories may not precisely match the curves, the primary focus is to determine if they generally follow the shapes defined by Reagan et al.

Different polynomial functions (linear, quadratic, and cubic) were selected based on the expected shapes of emotional arcs, as discussed in chapter 3. The linear function (y = ax + b) represents the first two emotional arcs, 'rags to riches' and 'tragedy.' In this equation, 'x' represents the novel time, and 'a' and 'b' are parameters that require optimization. The linear function aids in understanding the relationship between time and emotion scores, indicating whether scores increase, decrease, or remain relatively stable over time. Thus, it helps determine if the stories classified by Reagan et al. as 'rising' or 'falling' exhibit the expected trajectory.

To fit the data points, the 'curve_fit' function in Python is employed. It optimizes the parameters (a and b) of the linear polynomial equation (ax + b) to best match novel time and mean sentiment scores. The optimized parameters are used to generate fitted sentiment scores by applying the linear polynomial equation to the 'novel time' array. The resulting curve-fitted line is plotted alongside the original scores, visualizing the progression of sentiment for each individual story. Two examples are depicted in figures 4.9 and 4.10. Although the stories may not perfectly align with these curves, they generally exhibit an upward trend in sentiment. To assess the goodness of fit, the root-mean-square error (RMSE) is calculated, for which first the SSE should be calculated. To do this, Python code calculates the sum of squared errors (SSE) by comparing each data point in the measured sentiment scores with the corresponding value from the fitted curve. It starts by setting the SSE to zero, then goes through each data point. For each point, it calculates the difference between the actual value and the value from the fitted curve, squares this difference, and adds it to the SSE. The SSE represents the total accumulated squared differences between the observed data points and the fitted curve. The root mean square error (RMSE) is then calculated by dividing the sum of squared errors (SSE) by the number of data points, taking the square root of the result. The RMSE quantifies the average magnitude of errors between the original sentiment scores and the fitted sentiment scores. A lower RMSE value indicates a better match between the curve and the data. The RMSE is a widely used and understood measure of the goodness of fit, and helps evaluation of how well the curves fit the sentiment data.

By plotting the sentiment shape to a curve fitted line of the same data, and calculating the RMSE score of these lines, it can be deduced which story fits the best with the ‘Rags to Riches’ story shape. The results show that not all stories appear to follow a linear rising line. The end result shows that nine out of the ten analysed stories follow an overall continuous rise in sentiment. *The Spanish Tragedy* by Thomas Kyd actually follows a linear decreasing line.

The results for the best and least fitted story can be seen in figures 4.9 and 4.10.

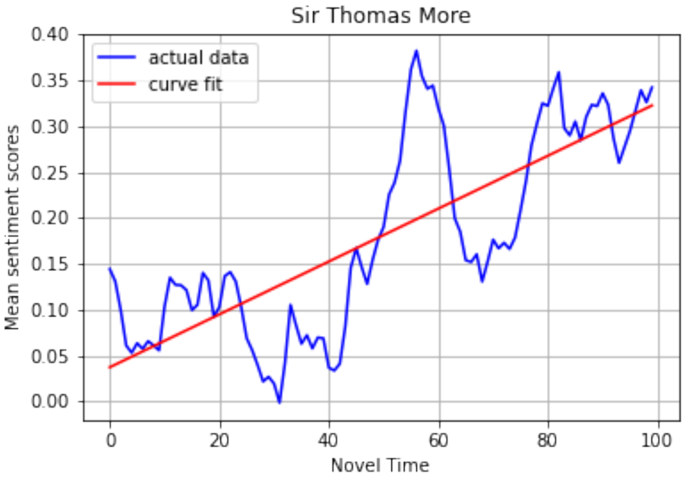

Figure 4.9 : Sentiment shape and curve fit for Sir Thomas More by William Shakespeare

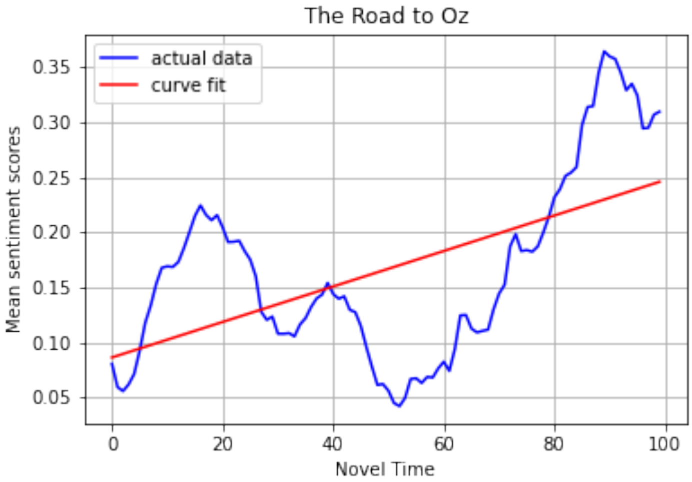

Figure 4.10 : Sentiment shape and curve fit The Road to Oz by L. Frank Baum

As shown in these graphs, *Sir Thomas More* fits the best with the 'Rags to Riches' plot shape as described by Reagan et al., as it has the lowest RMSE score with 0.1763. Nevertheless, based on the graph, it must be stated that the VADER analysis of *Sir Thomas More* is far from a straight line, especially based on the steep rise at around halfway through the story. *The Road to Oz* by L. Frank Baum, which has an RMSE score of 0.2221, fits the least well with the 'rise' shape. These findings present a contrast to the conclusions drawn by Reagan et al., who identified *The Winter's Tale* by William Shakespeare as the most suitable match for arc shape 1. Additionally, they found that *The Road to Oz* exhibited a stronger alignment with this arc shape compared to *Sir Thomas More*.

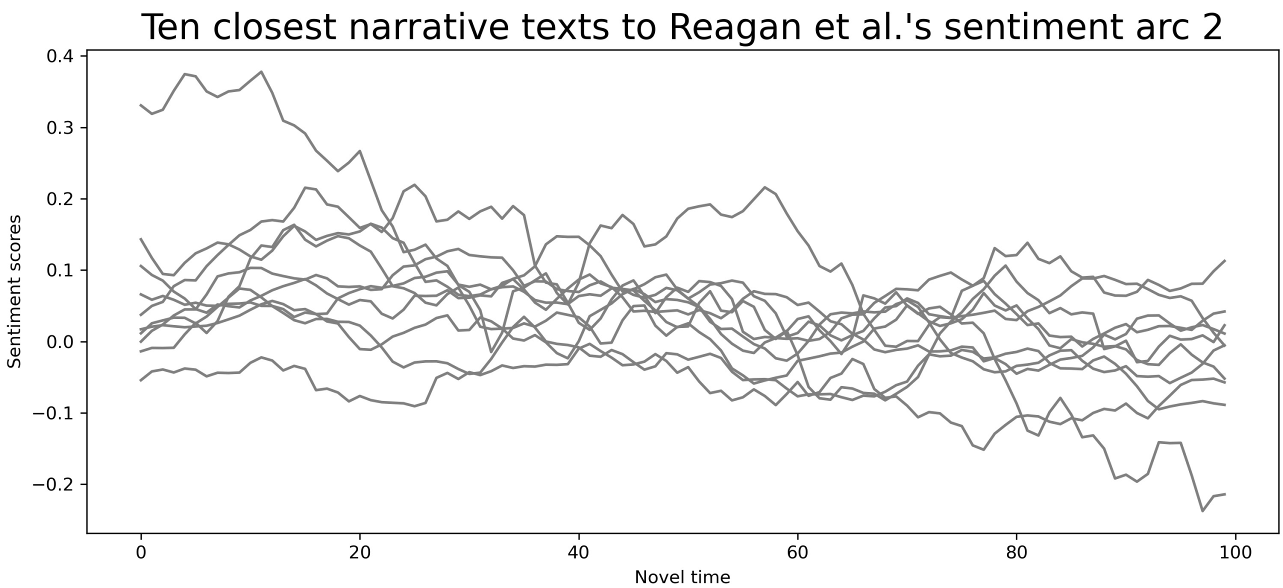

Figure 4.11: The ten closest narrative texts to Reagan et al.'s arc shape 1, named 'tragedy' or 'riches to rags' (fall), ar

The other linear model found by Reagan et al., the 'fall' shape, is characterised by continuous decreasing sentiment data. The ten closest stories to this emotional arc shape are illustrated in figure 4.11.

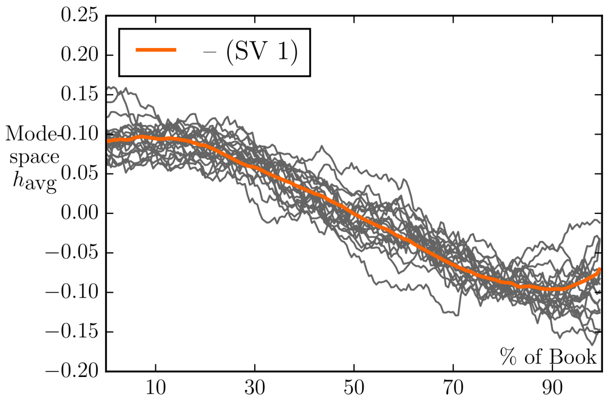

Figure 4.11: Reagan et al.'s 'Tragedy' or 'Riches to Rags' arc shape and the 20 closest texts

The ten stories that are cateogrized in the 'Riches to Rags' shape appear to follow the linear declining line considerably closer than the sentiment shape from 'Rags to Riches'. This is further supported by the RMSE scores, which are consistently lower than those from figure 4.7. Of these ten stories, *Tom Sawyer, Detective* by Mark Twain follows the line the closest, with an RMSE score of 0.1557 (figure 4.12). The story that follows this linear line the least close is *The House of the Vampire* by George Sylvester Viereck, with an RMSE of 0.2743 (figure 4.13). While these results show that the results from VADER and from Reagan et al. are closer to each other than those from arc shape 1, they still produce different results for the best fitting story to this shape. In contrast to *Tom Sawyer, Detective*, Reagan et al. found *Lady Susan* by Jane Austen to be the best fit for arc shape 2.

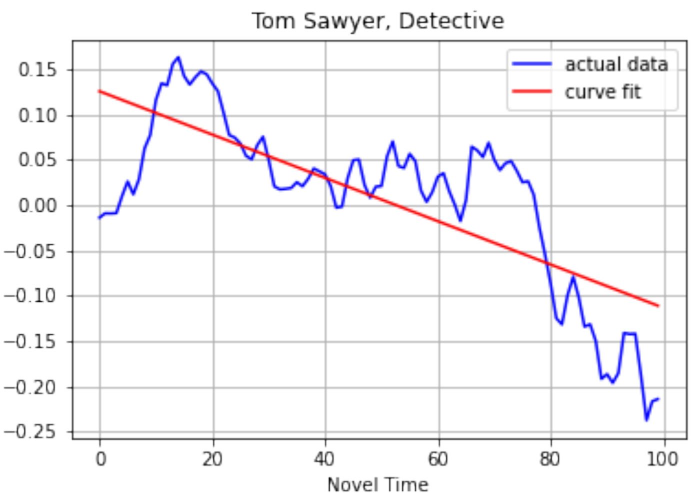

Figure 4.12: Sentiment shape and curve fit for Tom Sawyer, Detective by Mark Twain

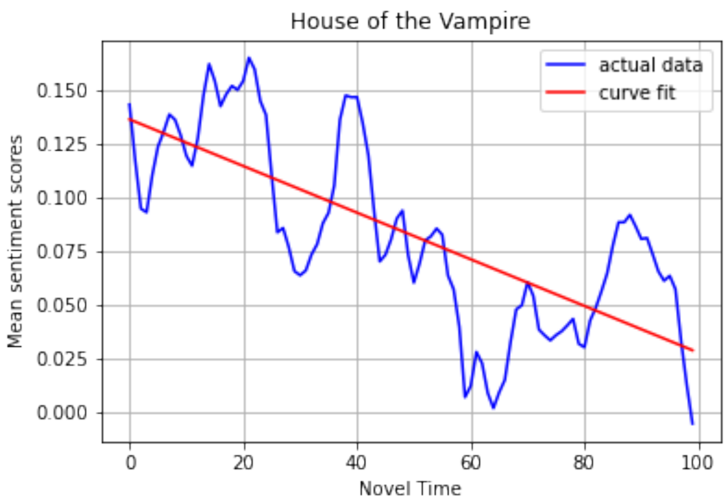

Figure 4.13 : Sentiment shape and curve fit for House of the Vampire by George Sylvester Viereck

4.3.2 ‘Fall-rise’ and ‘rise-fall’ shapes

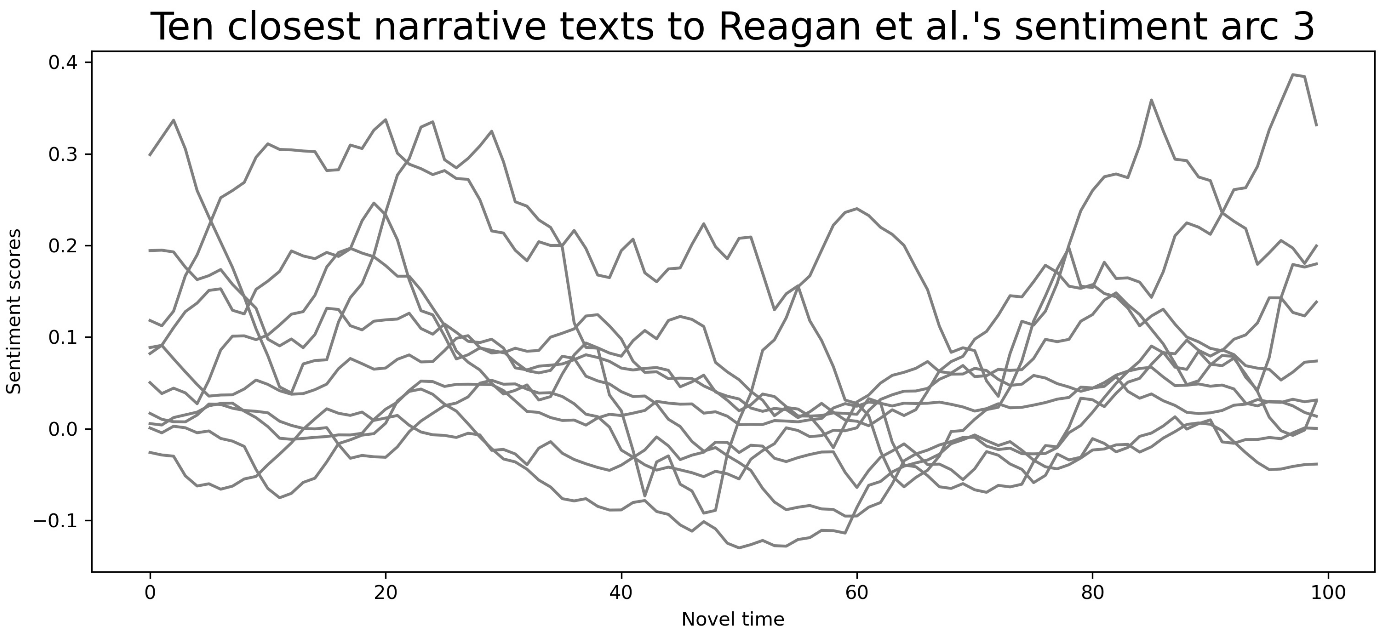

Figure 4.14 : The ten closest narrative texts to Reagan et al.’s arc shape named ‘Man in a Hole’ (fall-rise), analysed by VADER

The results from VADER’s sentiment analysis of the ten stories which Reagan et al. found to be closest associated with a ‘fall-rise’ shape can be seen in figure 4.14. While there are some outliers, the ‘fall-rise’ shape can still be recognized. Nevertheless, the shape is far from being as uniform as Reagan et al.’s results (figure 4.15).

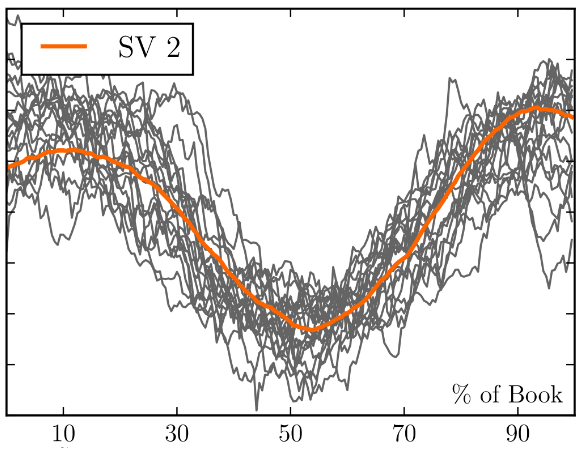

Figure 4.15: Reagan et al.’s ‘Man in a Hole’ arc shape and the 20 closest texts

For the 'Man in a hole' and 'Icarus' shapes, a quadratic polynomial regression (y = ax^2 + bx + c) was chosen to model the emotional arcs. Once again, in this equation, 'x' symbolizes the novel time, while 'a', 'b', and 'c' are parameters that need to be optimized. This function aids in understanding the relationship between time and emotion scores, allowing us to determine if the stories classified by Reagan et al. as 'Man in a hole' or 'Icarus' exhibit the expected trajectory. A quadratic function can help in recognizing a concave shape, where the sentiment initially decreases, reaches a minimum point, and then increases. On the other hand, a quadratic function can also exhibit a convex shape, where the sentiment initially increases, reaches a maximum point, and then decreases. It will therefor help determine if the stories classified by Reagan et al. as 'fall-rise' or 'rise-fall' exhibit the expected trajectory.

In a similar manner to the method described for the first two shapes in chapter 4.3.1, to fit the data points to the quadratic function, the 'curve_fit' function in Python is employed. The optimized parameters resulting from this function are used to generate fitted sentiment scores by applying the quadratic polynomial equation to the 'novel time' array. The resulting curve-fitted line, representing the sentiment progression, is plotted alongside the original scores for each individual story. This visualization allows us to observe the emotional arc and the alignment with the expected 'Man in a hole' or 'Icarus' shape.

Plotting the fitted curve for these graphs shows that only seven stories follow a shape that can be classified as ‘fall-rise’. One graphs shows a ‘fall’ shape (*Tamburlaine the Great Part I* by Christopher Marlowe), and two graph lines show a ‘rise’ shape (*Justice* by John Galsworthy and *A Yankee Flier Over Berlin* by Rutherford George Montgomery). Of the seven stories that follow a story that can be classified as a ‘man in a hole’ emotional arc, *R. Holmes & Co.: Being the Remarkable Adventures of Raffles Holmes, Esq., Detective and Amateur Cracksman by Birth* by John Kendrick Bangs scored the lowest RMSE score with 0.1695 (figure 4.16), and therefor fits the closest to this shape. Much like the first two arc shapes, this differs from Reagan et al.’s results where *The Magic of Oz* fits the closets the arc shape 3. The highest scoring story, and therefor the least close to a ‘fall-rise’ shape, was *The Sky Is Falling* by Lester Del Rey, with an RMSE score of 0.2298 (figure 4.17).

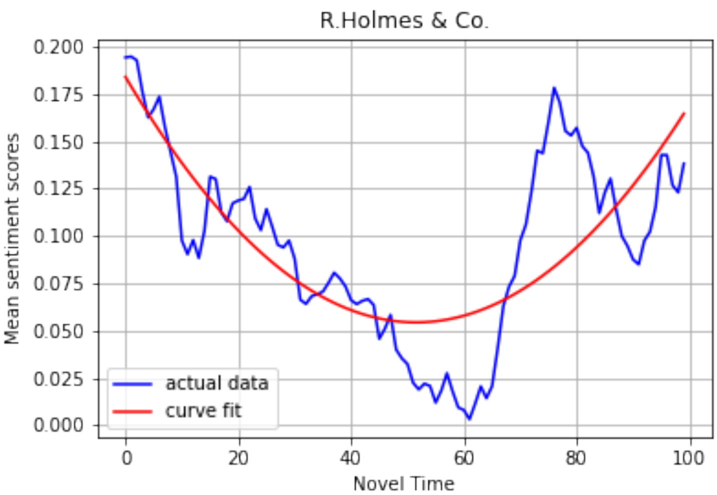

Figure 4.16: Sentiment shape and curve fit for R. Holmes & Co.: Being the Remarkable Adventures of Raffles Holmes, Esq., Detective and Amateur Cracksman by Birth by John Kendrick Bangs

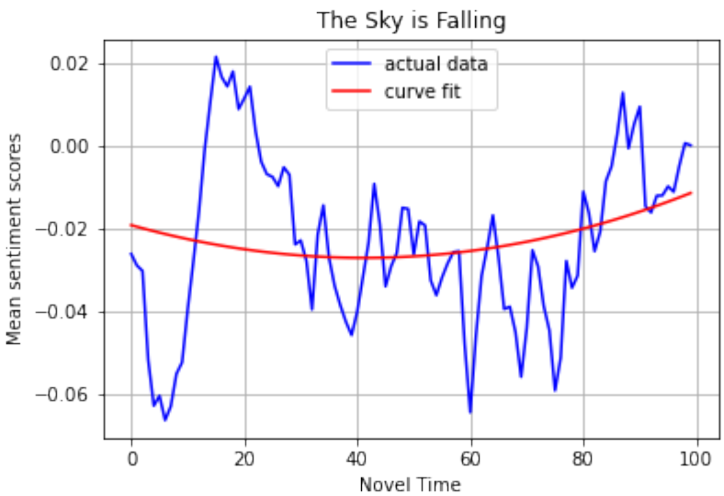

Figure 4.17: Sentiment shape and curve fir for The Sky is Falling by Leste Del Rey

Figure 4.18 displays the findings of VADER's sentiment analysis of the ten stories that Reagan et al. discovered to be most closely related to a ‘rise-fall’ shape. Of these ten stories, eight showed to have a similar shape. One of the remaining graphs had a ‘rise’ shape (*The Slayer of Souls* by Robert W. Chambers) and the other had a ‘fall’ shape (*The Oakdale Affair* by Edgar Rice Burroughs). Again, while the overall form of a rising and falling graph line can be recognized in figure 4.18, the graph shows a lot of noise, and is not as uniform a shape as Reagan et al.’s results (figure 4.19).

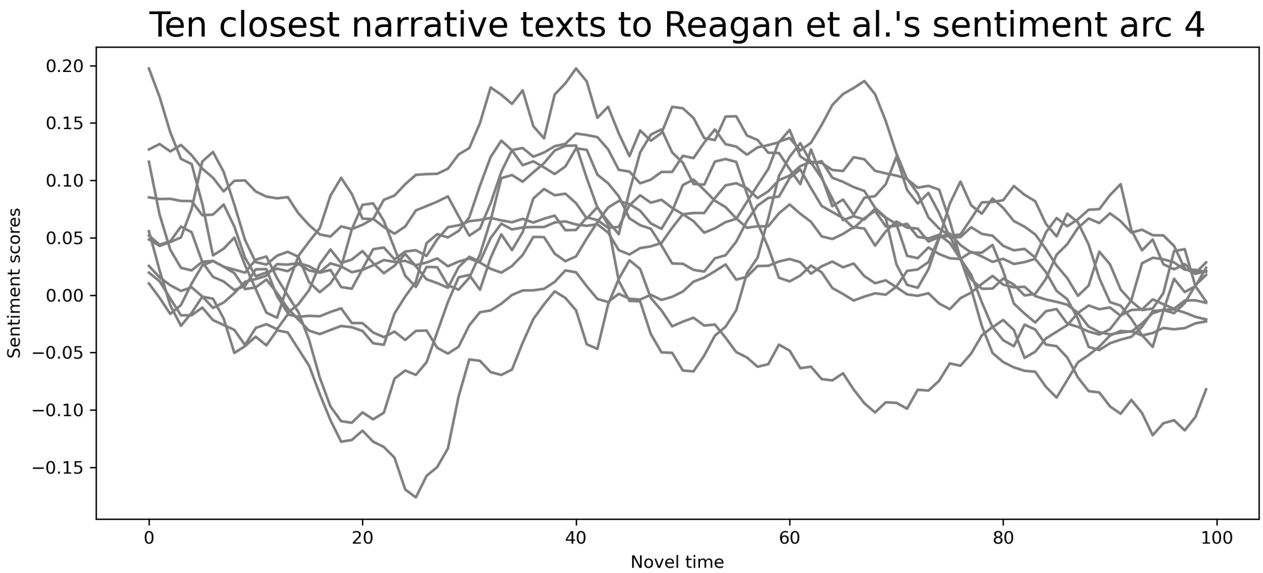

Figure 4.18: The ten closest narrative texts to Reagan et al.’s arc shape 4, named ‘Icarus’ (rise-fall), analysed by VADER

The story that resembles the ‘rise-fall’ shape the closest is according to the VADER analysis is *The Food of the Gods and How It Came to Earth* by H. G. Wells, with a normalised RMSE score of 0.1391 (figure 4.20). As for Reagan et al.’s results, the story that closest fit to arc shape 4 is *The Slayer of Souls*, which interestingly enough did not display a ‘rise-fall’ shape at all in VADER’s analysis, but rather a ‘rise’ shape. The story that resembles this shape the least well is *The Way of the World* by William Congreve with a score of 0.2385 (figure 4.21).

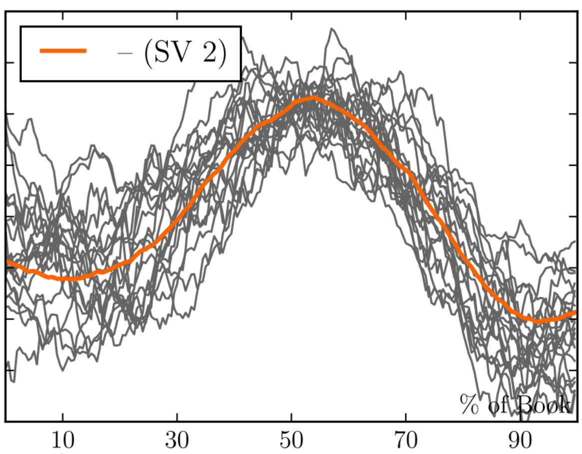

Figure 4.19: Reagan et al.’s ‘Icarus’ arc shape and the 20 closest texts

Figure 4.20: Sentiment shape and curve fit for The Food of the Gods and How It Came to Earth by H. G. Wells

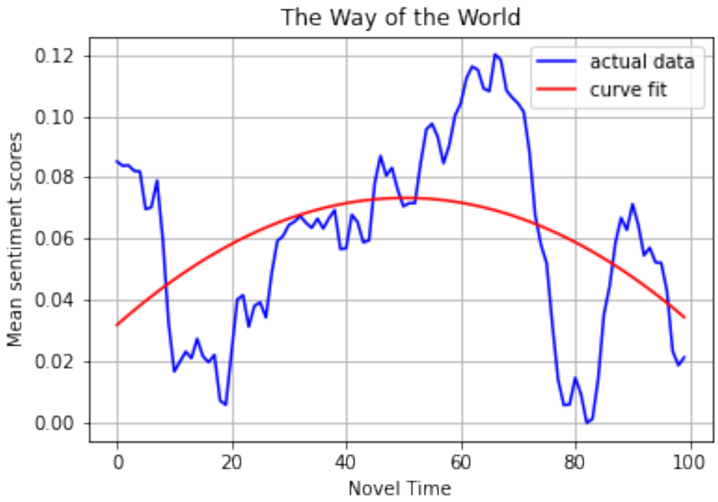

Figure 4.21: Sentiment shape and curve fit The Way of the World by William Congreve

4.3.3 ‘Rise-fall-rise’ and ‘fall-rise-fall’ shapes

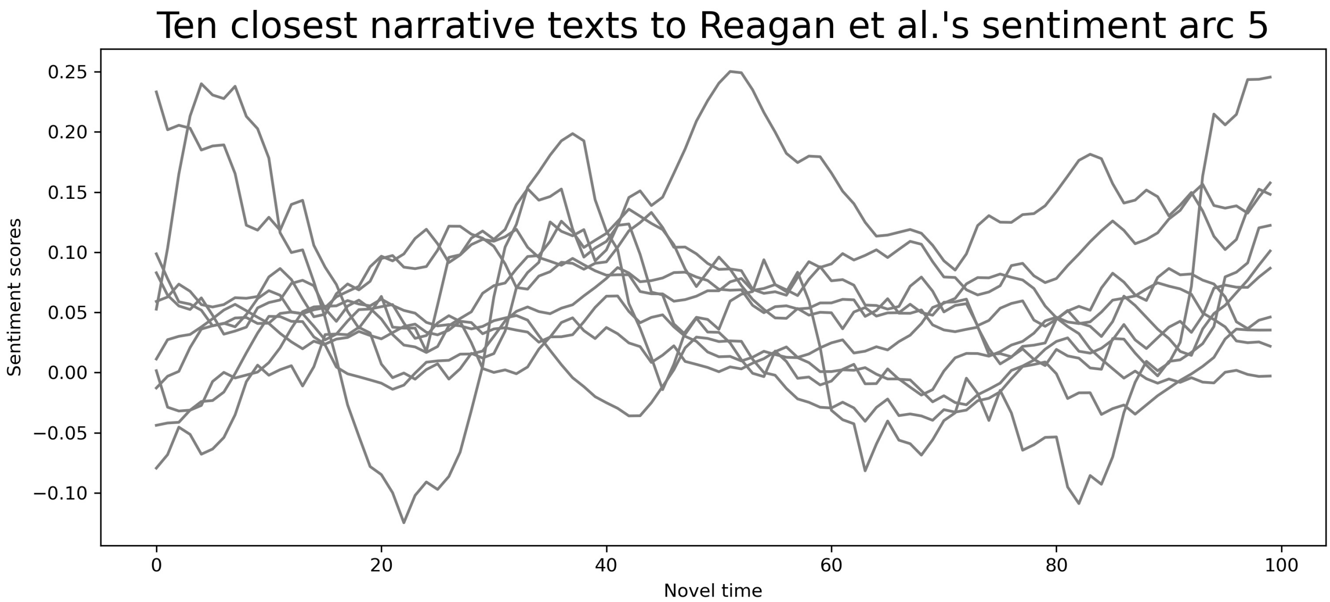

Figure 4.22: The ten closest narrative texts to Reagan et al.’s arc shape 5, named ‘Cinderella’ (rise-fall-rise), analysed by VADER

Lastly follows the results for the ‘Cinderella’ (rise-fall-rise) and ‘Oedipus’ (fall-rise-fall) shapes. Looking at the results of Reagan et al.’s analysis (figure 4.23), it appears that, the more complicated the shapes become, the more noise appears around the ideal shape line. While there will likely be some additional noise in VADER's graph as well, it does appear difficult to discern a 'rise-fall-rise' shape from this graph (figure 4.22).

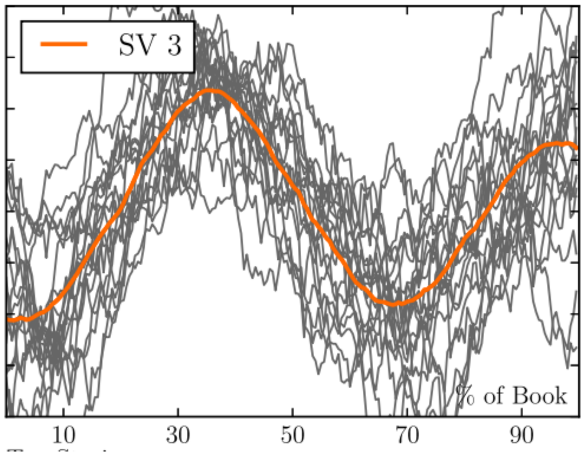

Figure 4.23: Reagan et al.'s 'Cinderella' arc shape and the 20 closest texts

For the 'Cinderella' and 'Oedipus' shapes, a cubic function (y = ax^3 + b2 + cx + d) or third-degree polynomial was chosen. Again, 'x' represents the novel time, and 'a', 'b', 'c', and 'd' are parameters that require optimization. The cubic function allows us to determine if the stories classified by Reagan et al. as 'Cinderella' or 'Oedipus' exhibit the expected trajectory, as it can help to depict a rise-fall-rise or fall-rise-fall shape. The cubic term (ax^3) makes the curve go up or down, the quadratic term (bx^2) affects how the curve bends, and the linear term (cx) changes the direction of the curve. By adjusting the values of 'a', 'b', 'c', and 'd', we can create curves that rise, fall, and rise again, or fall, rise, and then fall again.

Similar to the first four shapes which have been analysed, the 'curve_fit' function in Python is used to fit the data points to the cubic function, optimizing the parameters (a, b, c, and d) of the cubic polynomial equation to best match the novel time and mean sentiment scores. The resulting curve-fitted line, representing the sentiment progression, is again plotted alongside the original scores for each individual story, enabling us to visually assess the emotional arc and its alignment with the expected 'Cinderella' or 'Oedipus' shape.

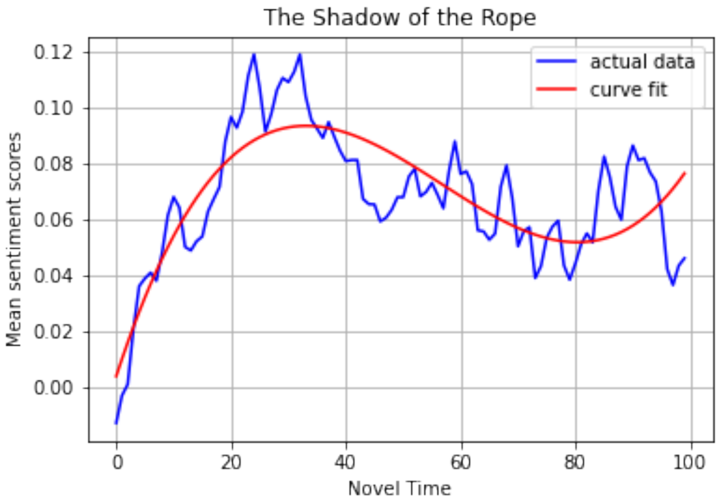

Figure 4.24: Sentiment shape and curve fit for The Shadow of the Rope by E. W. Hornung

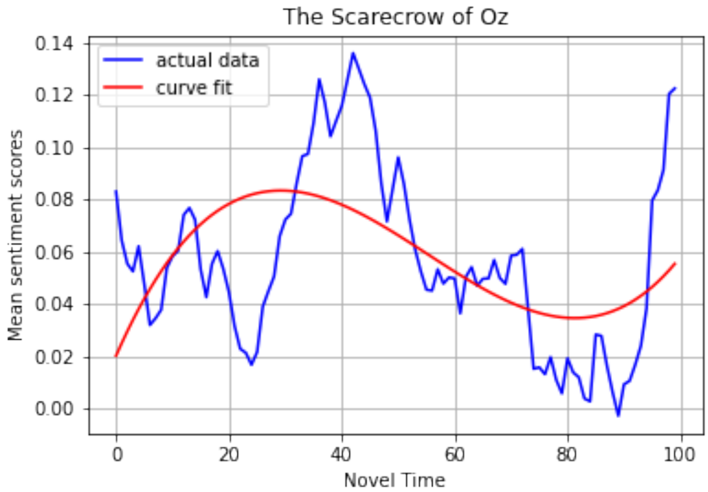

Figure 4.25: Sentiment shape and curve fit for The Scarecrow of Oz by L. Frank Baum

Of the ten stories which Reagan et al. found to be closest to a ‘rise-fall-rise’ shape, only five stories fit this shape according to the VADER analysis. The other five stories follow a ‘rise-fall’ shape (*The Mystery of the Hasty Arrow* by Anna Katharine Green and *Dave Dawson at Dunkirk* by Robert Sidney Bowen), a ‘fall-rise’ shape (*The Affair at Elizabeth* by Burton Egbert Stevenson and *The Haunted Man and the Ghost's Bargain* by Charles Dickens), and a ‘fall-rise-fall’ shape (*Through the Magic Door* by Arthur Conan Doyle). The story that scored the lowest normalized RMSE score (0.1101), and follows this shape the closest, is *The Shadow of the Rope* by E. W. Hornung (figure 4.24). Regarding Reagan et al.'s findings, The *Mystery of the Hasty Arrow* by Anna Katherine Green emerged as the story that demonstrated the closest fit to this arc shape. The story that follows this shape the least close is *The Scarecrow of Oz* by L. Frank Baum (figure 4.25).

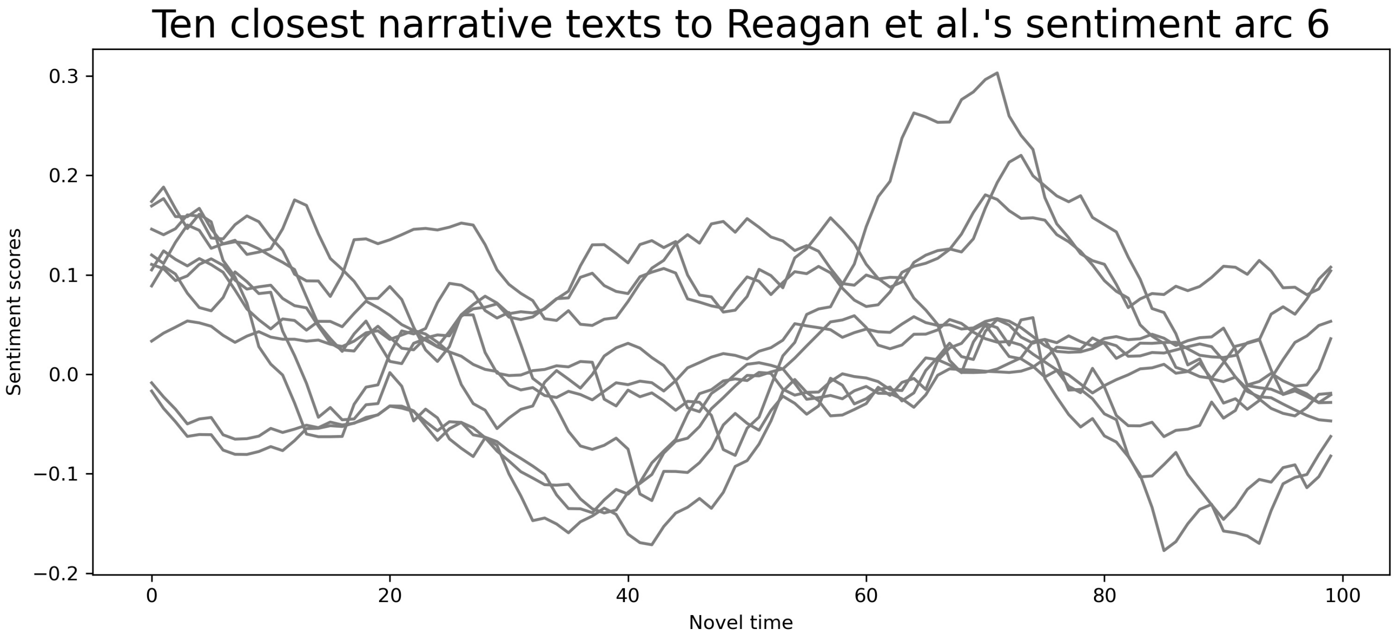

Figure 4.26: The ten closest narrative texts to Reagan et al.'s arc shape 6, named 'Oedipus (fall-rise-fall)

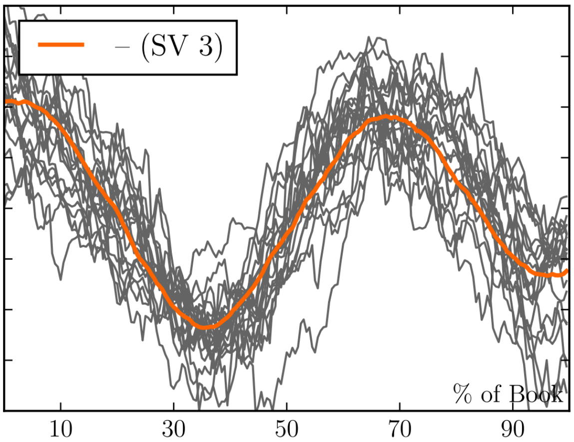

Figure 4.27: Reagan et al.'s 'Oedipus' arc shape and the 20 closest texts

The graphs that show the VADER analysis for the 'fall-rise-fall' shapes (figure 4.26) shows a slightly more uniform shape than the 'rise-fall-rise' graph. Of the ten stories analysed, nine showed to have a shape similar to Reagan et al.'s results (figure 4.27), with one story portraying a 'fall' shape (*The Blue Bird for Children* by Georgette Leblanc and Maurice Maeterlinck). Of the other nine stories, the one closest to the 'fall-rise-fall' shape is *Pariah Planet* by Murray Leinster with a normalised RMSE score of 0.1406 (figure 4.28). It must be stated that *This World is Taboo* by Murray Leinster is a later edition of this same story, and therefor shows an almost identical sentiment arc. Interestingly enough, *This World Is Taboo* was the story that closest fit arc shape 6 in Reagan et al.'s results. This makes this arc shape the only instance that showed the closest similarity between VADER's and Reagan et al.'s analysis. The story that showed data the least close to this shape is *Old Indian Days* by Charles A. Eastman with an RMSE score of 0.1755 (figure 4.29).

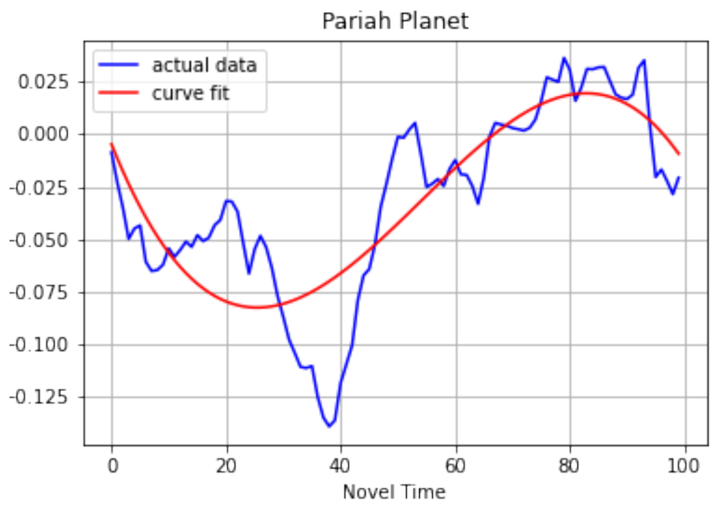

Figure 4.28: Sentiment shape and curve fit for This World is Taboo by Murray Leinster

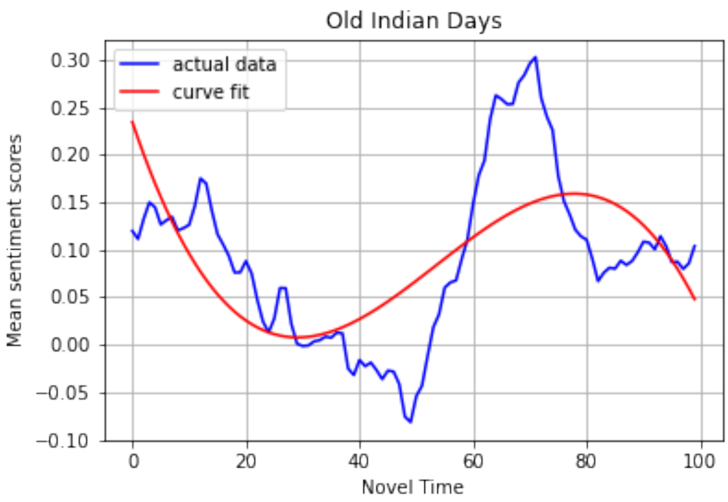

Figure 4.29: Sentiment shape and curve fit for Old Indian Days by Charles A. Eastman

In summary, when attempting to replicate Reagan et al.'s sentiment arc shapes using the VADER analysis method, both similarities and differences between the two approaches were found. While some stories did exhibit sentiment trajectories that matched the identified shapes, the overall patterns and consistency differed from Reagan et al.'s results. Additionally, in the majority of the analysed story sections per shape, there were multiple stories that did not align with the shapes attributed to them by Reagan et al. Notably, it is worth mentioning that the stories identified by Reagan et al. as best fitting to specific arc shapes differed significantly from the results obtained through VADER's analysis, except for an interesting case in shape 6. These results suggest that the choice of analysis method can greatly impact the identification and interpretation of sentiment progressions in literary works. These findings underscore the importance of acknowledging the limitations and intricacies of different sentiment analysis methods when studying emotional arcs in narratives. Although the VADER analysis provided valuable insights into the sentiment trajectories of the analysed stories, the disparities observed in comparison to Reagan et al.’s conclusions emphasize the need for caution when interpreting and generalizing sentiment arc shapes across diverse analytical approaches.

5. A critical evaluation of sentiment analysis and its academic and practical use

This chapter will function as a critical evaluation of sentiment analysis as a tool for plot extraction, and will consist of three main topics. The first section will go into why narratology remains a field that deserves to be explored. The digital humanities provides new ways for investigating narrative advancement, but why is it worth to spend our focus on this specific subfield? The second section will go more into what exactly is being measured by using sentiment analysis. Many researches (Jockers, Reagan et al., etc.) mention the word 'plot' in one way or the other when discussing the extraction of sentiment from a story. However, stating that sentiment analysis extracts 'plot' requires a bit more nuance, as plot is a multifaceted concept. Continuing on this, this section will also provide suggestions for improving the sentiment analysis method used in this thesis. These suggestions are partially motivated by the perceived flaws in this thesis' analytical methodology, but will also go into methods on how to get a better representation of plot using sentiment analysis. Lastly, this chapter will go more into the use of sentiment analysis as a tool for extracting sentiment data of a story, focusing on future research suggestions and practical applications.

5.1 Why narratology matters

Before beginning to take a critical look at what the use of sentiment analysis can be for the field of narratology, it is important to understand what narratology as a scientific field brings to the humanities. Narratology refers to the structuralist study of narrative. In addition to attempting to define the narratively relevant set of rules and norms controlling narrative interpretation, narratology studies what all conceivable stories have in common, as well as what distinguishes them from one another as narratives.[80]

Narratology focused much of its early work on describing the elements of the narrated (the "what" that is represented), the narrating (the manner in which the "what" is represented), and the rules governing how these elements interact.[81] Apart from understanding the "hidden" design or arrangements of stories, narratology also focuses on how narrative affects human perception. Stories are believed to tell much about humans on a psychological or sociological level. Psychologist George Kelly has explained how our personalities develop from the

80 G. Pradl, 'Narratology: The Study of Story Structure', *ERIC Digest* (1984).

81 G. Prince, 'Classical and/or postclassical narratology', *L'Esprit créateur, 48* (2008), pp. 115-123.

narratives we decide to create based on our interpretations of our past experiences, and how these narratives affect our expectations for the future.[82] Similar to this, sociologist Peter Berger has underlined the role that tales play in forming social realities by demonstrating how individuals' defining stories shift as they go from one phase of life to another.[83]

As Reagan et al. argue: '[t]hrough the identifications of motifs, narrative theories allow us to analyse, interpret, describe, and compare stories across cultures and regions of the world.'[84] Reagan et al. state here that important differences between cultural groups are reflected in the stories they tell. The narrative arts are created in a sociocultural context, much like any other kind of art. Every human filters and interprets information depending on their personal experiences and cultural background. For example, in 1980, Wallace Chafe presented a group of participants from various ethnicities a short film in which three young people steal pears from a man who is collecting them. [85] The participants gave many different interpretations of the story they saw. The Greek audience sought a greater narrative context and assigned societal motivations to the characters, whereas Americans responded in a manner that concentrated on specifics and temporal sequencing. This shows us that, while some narrative aspect may be shared worldwide, the differences can tell us much about their cultural origin.

Hayden White, an American Historian and literary critic, argued for the importance of understanding narratology as well. He argued that the study of narrative is particularly crucial since the organization of time and space in narrative forms is one of the main ways humans create meaning in general.[86] This influence can be seen in other scientific fields; White believes that literary writing and historical writing both share a heavy emphasis on story to convey meaning.[87] A narrative and linguistic discourse of some sort always exists in historiography. White contends that history is a form of "protoscience" that derives all of its theoretical legitimacy from narrativism and is therefor at least as much art as science. Understanding narratology is not only helpful in understanding the structure of stories, but in

---

82 G. Kelly, *The psychology of personal constructs: Clinical diagnosis and psychotherapy.* Abingdon: Routledge, 1955).
83 P. Berger, *The Social Construction of Reality*, (New York City: Anchor Books, 1966).
84 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', *EPJ Data Science*, 5 (2016), p. 1.
85 W. Chafe, *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (New York City: Ablex Publishing,1980).
86 H. White,*The content of the form: Narrative discourse and historical representation* (Baltimore: John Hopkins University Press, 1990), p. 2.
87 H. White, 'Interpretation in history', *New Literary History*, *4* (1973), pp. 281-314.

helping us understand how and why we structure information from other fields the way we do.

Contemporary academics still support White's beliefs. Szurmak and Thuna acknowledged in 2014 how narrative can be a powerful tool for teaching and learning.[88] Storytelling has great power since it exploits the methods the brain already employs for learning. For one, narrative allows for abstract information to become more immediate and tangible. By providing the frame for pupils to place their knowledge into, narrative contextualizes information. Second, a narrative framework can make pupils experience more direct emotional experiences, and therefor remember them better. Teachers can encourage the brain's natural learning heuristics by framing the curriculum in a narrative structure.

Based on the sociological and psychological theories behind narratology, understanding narratives ultimately allows us to understand others through their languages, histories and cultures. Moreover, it helps us to understand why we structure certain information the way we do, and how doing so improves our ability to perceive it. The digital humanities now provides techniques for extracting narratological data that were unimaginable 20 to 30 years ago. Continuing to enhance and apply these methodologies will eventually allow us to fully exploit the benefits that narratology offers to its own and other scientific domains.

5.2 To what degree can sentiment analysis extract information on plot?

The studies by Reagan et al. and Jockers frequently use the word 'plot', creating the impression that the findings of a sentiment analysis indeed reveal information about a story's plotline. However, as described in the literary framework, what exactly entails as the plot of a story is a matter that narratologists have not completely agreed on. There are many ways to conceptualize the term plot, and many factors that play a part in constructing a plot. Jockers and Reagan et al. acknowledge this and include a disclaimer in their papers that denies that their results are a direct representation of the plot of a story. In his first demonstration of the R.Syuzhet package, Jockers wrote that the main function of the package was to 'extract sentiment and plot information from prose'. Later on, Jockers also mentioned the extractions

---

88 J. Szurmak, J. and M. Thuna, 'Tell me a story: The use of narrative as tool for instruction', *ACRL 2013 Conference, American Library Association,* (2014).

of a 'sentiment-based plot'.[89] These intentions make clear that the analysis does not extract the literal plot of a story, but merely components of the plot. Reagan et al. too point out that the emotional arcs visualized through sentiment analysis are not a direct representation of the plot structure of a story or corpus; the emotional arcs merely allow for a partial analysis of plot:

> While the plot captures the mechanics of a narrative and the structure encodes their delivery, in the present work we examine the emotional arc that is invoked through the words used. The emotional arc of a story does not give us direct information about the plot or the intended meaning of the story, but rather exists as part of the whole narrative.[90]

In fact, Rebora emphasizes that when focusing on the most established theorizations of narrative form and plot, the mention of emotions can only rarely be found.[91] The manipulation and interaction of components like space, time, and characters create a plot, whereas emotions are merely a tool used to stimulate the reader's interest with only secondary and indirect effects on the narrative's overall structure. This begs the question to what degree sentiment analyses can truly create a formal model of what is known as the 'plot' of a story.

As described in the literary framework, Kukkonen's definition of plot, which functions as one of the most established theorizations on plot today, does not include emotions or valence as one of the main aspects that makes up the plot of a story.[92] Kukkonen based her broad definition on a collection of definitions by established narratology scholars, such as Vladimir Propp and Tzvetan Todorov. Nevertheless, some scholars have recently advocated for re-evaluating the impact of emotions on narrative advancement.

A 2011 study by Hogan theorizes in favour of sentiment analysis being a useful tool in visualizing the plot of a story.[93] Hogan argues that emotion systems define the standard features of all stories, as well as cross-culturally recurring clusters of features in universal

---

89 M. Jockers, 'Revealing Sentiment and Plot Arcs with the Syuzhet Package', *matthewjockers.net*, 2 February, 2015. <https://www.matthewjockers.net/2015/02/02/syuzhet/> (12 September, 2022).

90 A.J. Reagan, L. Mitchell, D. Kiley, C.M. Danforth and P.S. Dodds, 'The emotional arcs of stories aredominated by six basic shapes', *EPJ Data Science*, 5 (2016), pp.1-12.

91 S. Rebora, 'Sentiment Analysis in Literary Studies. A critical Survey', *Digital Humanities Quarterly*.

92 K. Kukkonen, 'Plot', in P. Hühn, J. Meister, J. Pier and W. Schmid, *Handbook of narratology*, (Berlin: de Gruyter, 2014), pp. 706-719.

93 P.C. Hogan, *Affective narratology: The emotional structure of stories* (Lincoln: University of Nebraska Press, 2011).

genres.[94] Without contradicting Hogan's theory, Rebora does believe that his reasoning nonetheless renders it a problematic analysis for computers. Hogan's reasoning is based on the idea that little stories make up one larger narrative. Those are composed of 'episodes', and episodes are composed of 'incidents'. The notion of 'normalcy' in this context is essential: an episode starts and ends with normalcy. With sentiment analysis, normalcy is difficult to quantify, because it is based on a more nuanced emotional system than just a binary system of positive or negative emotional states. Nonetheless, Hogan's argumentation contributes to the argument made by Jockers that emotions do, in fact, play a large part in identifying a narrative pattern.

In 2020, Carmen Tu and Steven Brown published a paper that also supports the claim that emotions play a larger role in narrative structure than often acknowledged.[95] They argue that emotional structure is generally seen as more important than characters in establishing the essence of stories. According to this theory, the protagonist's actions are dictated by the thematic objectives of the narrative, which is an abstract framework that exists outside of the protagonist. Their analysis calls for a shift in the order of importance for these components. The main idea behind this concept is that the dramatic arc of a tale corresponds to the ups and downs in the protagonist's emotions. In other words, the plot's structure is closely related with the protagonist's psychological experience inside the story, particularly the dynamics of the character's problem-solving capabilities.

Based on these assessments by Tu and Brown, it becomes clear that the characters of a story also play a substantial role in the construction of a plot. In general, researchers who have attempted to extract plot data using computational technologies share this notion. Reagan et al. have stated that they believe their analysis would become more detailed with, for instance, the addition of a character network analysis.[96] Nalisnick and Baird, who combined the methods of sentiment analysis and character network analysis, also emphasized the importance of researching the relations between characters to get a greater understanding of the plot of a story.[97] Similarly, Min and Park proposed a network-based framework for

---

94 S. Rebora, 'Sentiment Analysis in Literary Studies. A critical Survey', *Digital Humanities Quarterly*.

95 C. Tu, and S. Brown, 'Character mediation of plot structure: Toward an embodied model of narrative', *Frontiers of narrative studies*, *6* (2020), pp. 77-112.

96 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', *EPJ Data Science*, 5 (2016), p.11.

97 E.T. Nalisnick and H.S. Baird, 'Character-to-character sentiment analysis in Shakespeare's play', *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2 (2013), p. 479.

modelling a narrative by emphasising the character and their interactions.[98] This study resulted in a series of graphs showing the interconnect character trajectories. A more character centred approach to researching plot structure was also performed by Micha Elsner in 2012.[99] He argued that current approaches of analysing and producing plot structures are excessively focused on events and lack abstraction. Computational approaches might benefits substantially from a better plot structure representation. By providing a method for comparing novelistic plots in terms of the cast of characters and their social ties, this improved structure might be achieved.

Based on many established definitions and theorizations of plot, such as those by Kukkonen, Propp and Todorov, emotions are merely a small fragment of what a plot consist of. In accordance with this definition, it is false to assert that a sentiment analysis graph depicts the development of a story's narrative. Nonetheless, scholars continue to debate whether or not emotion should be seen as a more crucial aspect of plot, if not the driving force. The emotions of the protagonist are, for example, mentioned as a useful criterion for the progression of plot. On top of that, it is apparent that many academics highlight the importance of characters in the development of a narrative; they are generally cited as one of the most important elements of a plot.

5.3 Can sentiment analysis be adapted to improve the extraction of plot data?

As discussed in §5.2, the emotional arc that results from a sentiment analysis is merely a component of the entire narrative of a story, and does not directly reveal the plot or the story's intended purpose. The challenge that derives from this is to modify VADER, Syuzhet, or any other sentiment analysis tool in order to improve the depiction of a story's plot. In this chapter, this thesis proposes some alterations to the performance of sentiment analysis that would reinforce the notion that such an analysis could be of great value in analysing and visualizing the plot of a story.

As discussed in §2.3 of the literature review and §5.2, there are many ways to conceptualize the term plot, and many factors that play a part in constructing a plot. Scholars do not unanimously agree on one definition of the word plot, or the importance that emotions

[98] S. Min and J. Park, 'Narrative as a Complex Network: A Study of Victor Hugo's Les Misérables', *Proceedings of HCI Korea*, (2016), pp. 100-107.
[99] M. Elsner, 'Character-based kernels for novelistic plot structure', *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics,* 1 (2012), pp. 634-644.

play in it. As Jockers mentions in his introductory blog post on Syuzhet in 2015, the numerous explanations of plot suffer from a common problem when it comes to computational analysis: they all lack data.[100] He states that all of the proposed taxonomies suffer from anecdotalism. If there is no data to measure, and we rely on our imagination to recollect familiar stories that fit a certain structure (e.g. Vonnegut's Cinderella or Joseph Campbell's Monomyth), it becomes a difficult task to compare stories mathematically and computationally. According to Jockers, fluctuations in sentiment can serve as a fairly natural proxy for variation in story progression. By converting the data of stories of different lengths into a normalized space, Jockers made it possible to quickly compare the sentiment-based plot of multiple stories using only one graph. Elaborating on Jockers' work, the objective is now to improve upon this analysis without losing one of the tool's main advantages: its capacity to quickly and clearly visualize the evolution of valence in a single graph. In an effort to maintain the overall result's clarity and simplicity, this thesis aims to incorporate as little extra analysis as possible while still enhancing overall extraction of plot data.

One method for doing so is pairing the results of the sentiment analysis with a character network analysis. A character network is a graph that is taken directly from a narrative, where the nodes stand in for individual characters and the edges for their interactions. According to Kukkonen and Rebora's concept of plot, including an examination of a story's key characters allows for a more thorough grasp of the whole narrative. Through the study of character networks, a variety of narrative-related issues, such summarization, categorization, or role recognition, may be automatically handled. An example of such a character network is displayed in figure 5.1.

---

100 Jockers, M., 'Revealing Sentiment and Plot Arcs with the Syuzhet Package', *matthewjockers.net*, 2 February, 2015. <https://www.matthewjockers.net/2015/02/02/syuzhet/> (12 September, 2022).

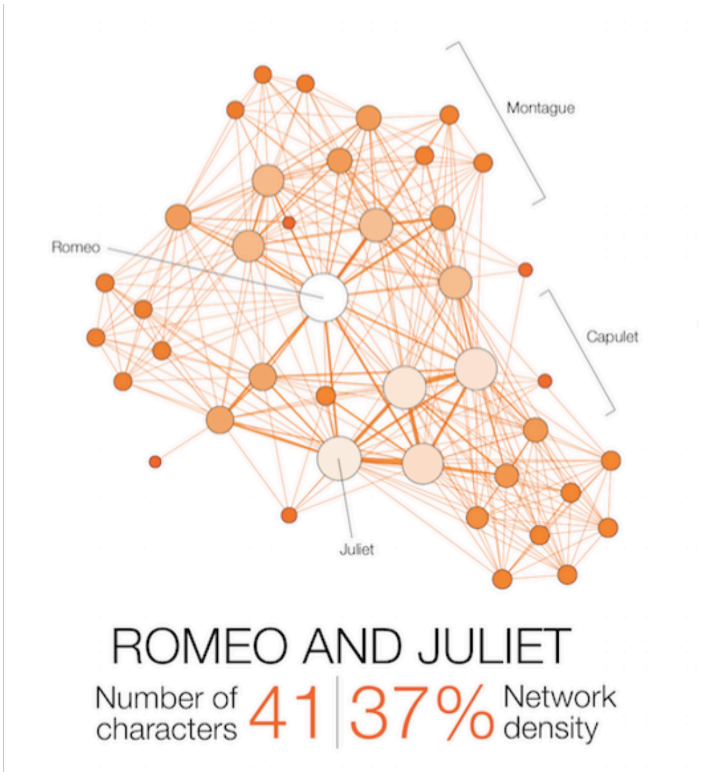

Figure 5.1: a character network analysis of Shakespeare's Romeo and Juliet

Figure 5.1 is taken from the online blog by Martin Grandjean, who is a digital humanities researcher. Grandjean's aim with creating these network analysis visualizations was to understand whether all Shakespeare's plays are structured in a similar way. For example, based on these character network analyses, Grandjean argued that *Hamlet* (despite being Shakespeare's longest tragedy) is less structurally dense than *King Lear*, *Titus Andronicus* and *Othello*. He also found that some graphs make the social groupings that impact the drama very evident, such as the Capulets and Montagues in *Romeo and Juliet*. Since the graph itself gives a quite complicated visualization of the characters of *Romeo and Juliet*, this analysis often benefits from being visualized in an interactive manner, in which the connection between each node can be clearly understood. When being provided in such a way, this character network analysis can be a clear and valuable asset in understanding the plot of a story, in combination with sentiment analysis.

Another way of incorporating a story's characters into the analysis of a plot is by plotting the collected data in numerous graph lines, rather than just one. Figure 5.2 illustrates the possible format of such an analysis, again using *Romeo and Juliet* as an example.

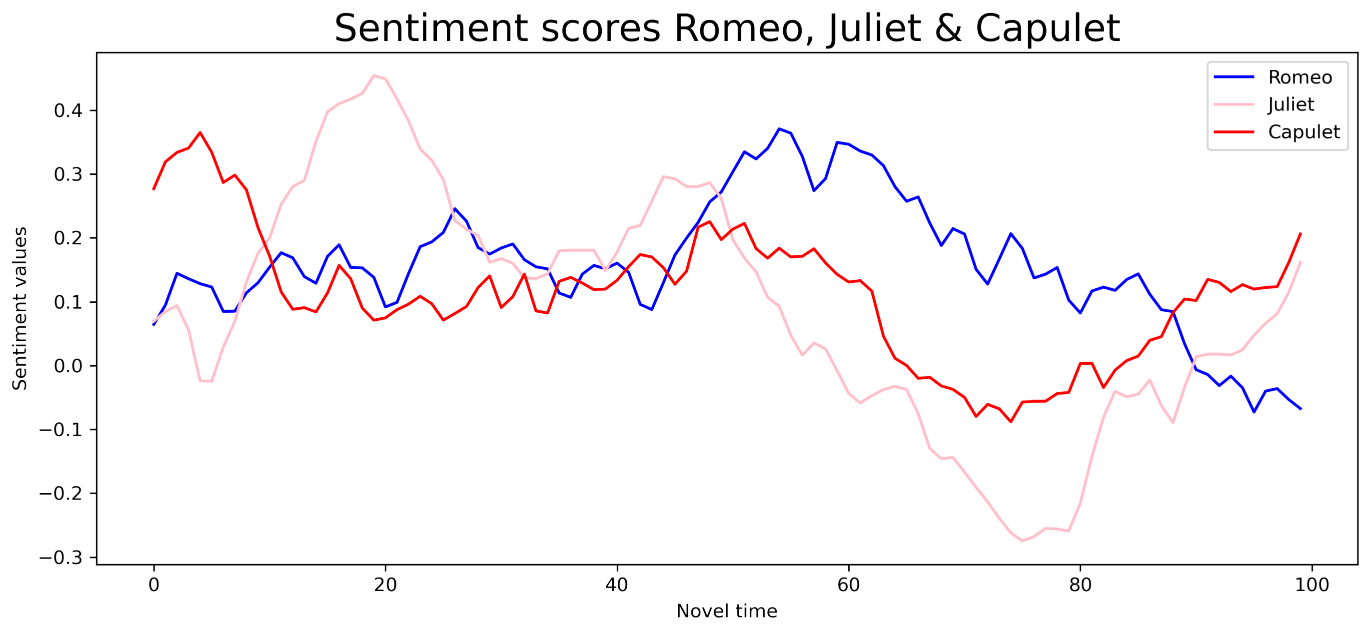

Figure 5.2: sentiment analysis of Romeo's, Juliet's and Capulet's dialogue in Shakespeare's Romeo and Juliet

This graph shows the sentiment of three characters throughout Shakespeare's play: Romeo, Juliet and Lord Capulet. As discussed in §4.2.3, the different levels of focalisation can often bring rapid changes in the course of a sentiment analysis graph. It is important to understand the shifts in focalisation that occur in a story when interpreting these results. By plotting the sentiment of each character separately, these shift in perception become much clearer straight away. In contrast to one line that represents the sentiment of the text throughout the entire play, this graph shows the progression of valence for three different characters, and portrays how each experiences the story. In theory, this serves as a technique for visualizing these characters' hardships and triumphs in the form of the emotional highs and lows that they express in their dialogue. Nevertheless, it must be acknowledged that figure 5.2 already displays some unexpected results. It is to be expected that Romeo and Juliet's line follow a relatively similar trajectory, showing their emotional highs when they fall in love in the beginning of the play, and an emotional low point at the tragic ending of the play. However, their emotional trajectories seem to go rather different, especially from the halfway point of the play onwards. As a result, this graph brings up similar questions as the results from chapter 4. The takeaway from this graph is that context—both narrative and linguistic—is again necessary in order for it to be interpreted correctly. Shakespeare's poetic writing style, as well as the lack of descriptive language in a play, is likely a factor in these unexpected results.

While figure 5.2 can show us the sentiment of each character individually, it tells us little about the interactions between these character like the character network analysis from figure 5.1 does. In order to adjust for this, these two methods can be combined to some extent in the form of a character-to-character sentiment analysis. This method is based on the study by Nalisnick and Baird, who conducted an analysis on characters from Shakespeare's plays.[101] Their goal was to track the emotional trajectories of interpersonal relationships rather than of a whole text or an isolated character. Whereas Nalisnick and Baird measured the valence of sentences using the AFINN sentiment lexicon, figure 5.3 shows this method adapted to the analysis approach used for this thesis; the VADER lexicon was used to measure the valence per sentence, after which rolling averages were created so that multiple characters can be compared to each other. When extracting lines using NLP from a play like *Romeo and Juliet*, it is challenging to identify precisely to whom the dialogue is directed towards. In order to arrive at the results in figure 5.3, the dialogue valence for each occurrence of continuous speech was measured, and it was assumed that the measured sentiment was directed at the character that spoke immediately following the current speaker.

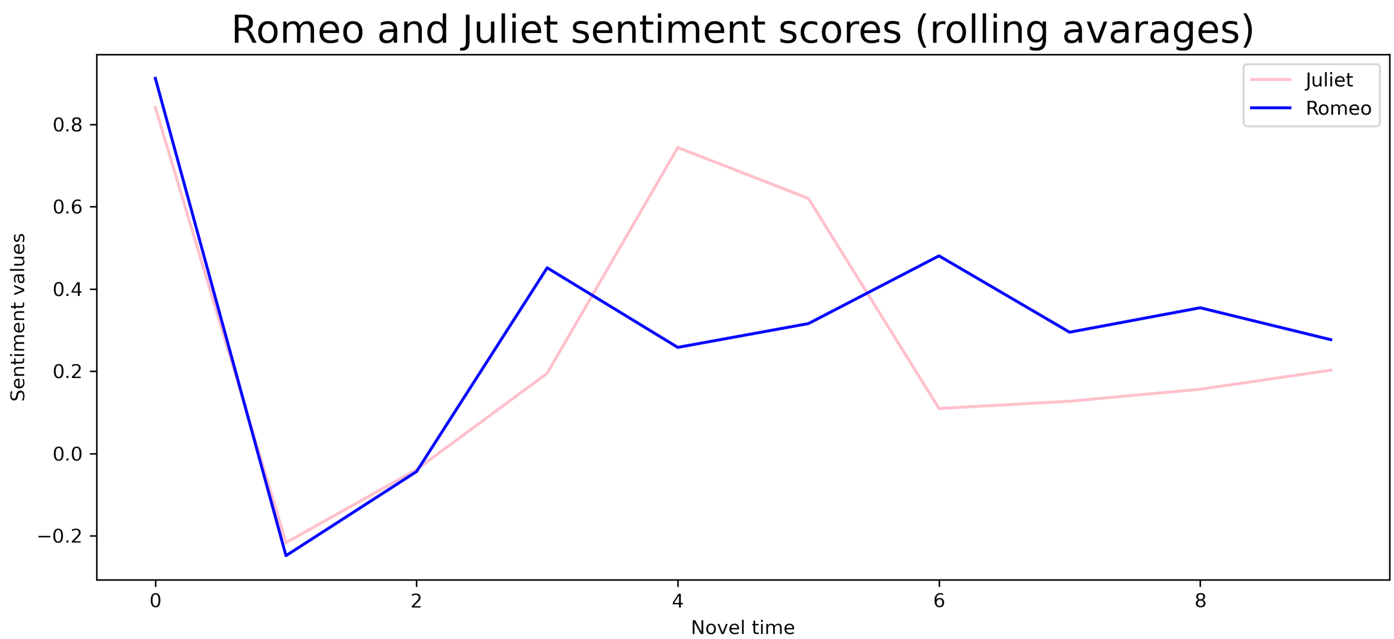

Figure 5.3: character-to-character sentiment analysis for Romeo and Juliet

[101] E.T. Nalisnick and H.S. Baird, 'Character-to-character sentiment analysis in Shakespeare's play', *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2 (2013), pp. 479-483.

Again, figure 5.2 shows some surprising results. When familiar with the story of Romeo and Juliet, one might expect that the two show a steadily growing affection for one another, which only dwindles as the drama takes a tragic turn in the second half. Whereas figure 5.3 shows a mutual decline in affection to each other in the beginning of the play, after which the affection grows. However, the graph does display varying levels of sentiment between the two: Juliet expresses more favourable feelings about Romeo midway through the play than she does towards the conclusion. A valid explanation for this can be found by comparing the graph to the sentiment scores for each sentence, which is not dissimilar from the findings in §4.2 from the results section. The first few lines of dialogue between Romeo and Juliet shows great signs of affection: through three illustrations of intimacy—eyes, palms, and lips touching—their first meeting is a dramatization of love at first touch. Though as a result of the mentions of 'enemy', 'envy' and the many references to 'sin', the sentiment quickly decreases as Romeo and Juliet learn the truth about which families they originate from. Especially Romeo's upset confession about his family provides low sentiment scores towards the middle of the play, whereas Juliet's dialogue remains more positive. It is here that shortcomings can be seen in this method; the negative dialogue of characters is not always directed towards one another. Whereas Nalisnick and Baird's research showed promising results in establishing a character-to-character network analysis, applying their concept to the method utilised in this thesis produces questionable results. Creating a more enhanced form of this method, which can more effectively identify to whom a line is directed towards, can account for these inaccuracies. Accounting for these early limitations, a customized version of this method can be an effective tool for extracting and visualising data from a plot in a clear manner.

5.4 Research practices

A large portion of this thesis is aimed at evaluating a sentiment analysis programme based on VADER, in order to explore whether it is possible to successfully extract plot data from stories. Regardless of whether or not this tool is successful, the applicability of such a computational method must also be considered. After all, computational approaches to literature are often criticized for offering no more than confirmation of the obvious.[102] It may be argued that Reagan et al. just confirm what we already know; that the great majority of

---

[102] M. Jockers, *Macroanalysis: Digital methods and literary history* (Champaign: University of Illinois Press, 2013).

stories are really multiple variations of a small number of storylines. However, the method of sentiment analysis, as well as the data resulting from such analyses, could provide more than just confirmation. This section will focus on how sentiment analysis tools can be used for future research, and what the data received from the sentiment analysis can mean for research in the digital humanities field and beyond.

Firstly, as discussed in §5.1, using literature to examine society and culture is a significant aspect of folklore studies. One of the future applications of sentiment analysis mentioned by Reagan et al. focuses on the comparison of cultural differences.[103] By using a 'big data' lens, sentiment analysis can be used to study a culture's evolution through the stories they tell. The identification of emotional progression allows us to not only analyse the evolution of story telling of one specific culture, but also compare stories from across regions of the world. Analysing and interpreting a large corpus could provide us with new insights into these practices, without being confined by the time-consuming practice of close reading.

Building on this, sentiment analysis could be used to interpret the differences and similarities of translated literature. Exploratory research on this matter has already raised some interesting questions. In 2019, Erin Shaheen utilized sentiment analysis to explore different English translations of *The Odyssey* (from 1725, 1900 and 2017).[104] The purpose of this study was to investigate how the translators' experiences and time periods of working can affect how they interpret a text. While the overall arcs remained relatively similar, all three of the translations some outliers when it came high and low points. Using Syuzhet allowed Shasheen to compare these translations by means of computational analysis, in contrast to close reading. She claimed the tool to be extremely useful in tracing the differences between translations, comparing the noise of valence of the graphs to interpret differences between translators. For future research, Shasheen argued that pairing this method with a human perspective could identify the reason for the different crux points in the graphs. Jennifer Isasi performed a similar study in 2019.[105] She studied the effects of translating sentiment lexicons by comparing the findings to the original lexicon. She also compared the sentiment analysis results from the original works to those from the translations. Apart from these two

---

103 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', *EPJ Data Science*, 5 (2016), p. 1.
104 E. Shaheen, 'Lost in Translation: Using Sentiment Analysis to Analyze Translations of Homer's Odyssey', (2019).
105 Isasi, J., 'Sentiment analysis methods in translation', *Conference presentation for Digital Frontiers Annual Conference, Austin, TX, United States*, (2019).

exploratory studies, research on the emotional transformations of translated texts is scarce, and sentiment analysis might play an important role in learning more about this area.

Lastly, data extracted by means of sentiment analysis can be used to improve artificial intelligence. In 2016, Riedl and Harrison argued that stories could be used to teach human values to artificial agents, with the goal of pursuing successful ‘value alignment’.[106] Value alignment means that an intelligent agent can only pursue objectives that are advantageous to humans. A universal AI should be prevented from performing actions that harm humans, whether on purpose or accidentally, by having its values aligned. Riedl and Harrison hypothesize that an artificial intelligence that can read and comprehend stories would be able to pick up on the values that cultures hold subliminally. Apart from teaching AI common sense, the data extracted by means of artificial intelligence can also teach AI to write novels. AI generated novels are no longer limited to since-fiction. Natural language generation (NLG) algorithms can generate new texts based on textual resources that are provided as training data. GPT-3, a human language generator created by Open AI, is an example of such an NLG algorithm. As of 2020, dozens of novels utilizing language models like GPT-3 have been co-authored on various platforms, from science-fiction to children’s books.[107] Teaching such AI’s more about the emotional trajectory of stories could improve their overall storytelling abilities. Of course, this brings on a whole lot of new questions: is it still possible for readers to connect with a text that has been created by a computer, and who should be regarded as the author of a text created by an algorithm? Sentiment analysis can be a key component in enhancing these tools if the development and enhancement of NLG algorithms gain increased support.

5.5 Practical use

As discussed in the literature review of this thesis (chapter 2), sentiment analysis has already found practical use outside of the field of research as well. Businesses use it to measure consumer interest in a particular product or to measure public opinion on a political figure or policy choice. Nevertheless, since sentiment analysis tools such as Syuzhet are relatively new compared to programmes measuring sentiment from social media platforms, such practical

[106] M.O. Riedl and B. Harrison, ‘Using stories to teach human values to artificial agents’, *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence* (2016).

[107] L. Floridi and M. Chiriatti, ‘GPT-3: Its nature, scope, limits, and consequences’, *Minds and Machines*, *30* (2020), pp. 681-694.

uses have yet to be found for literary applications. However, it is feasible to use this research tool in a more practical manner as well. This section will offer suggestion for such practical options.

For one, similar to how companies use sentiment analysis on user reviews to dictate the popularity of a certain product, the same measure could be applied to literature. Reagan et al. and Woods suggest using these tools to find patterns that dictate the success of a story. Apart from recognizing and illustrating the six basic shapes which Reagan et al. found, they also determined that some emotional arcs produce more successful stories, as measured by downloads.[108] They found that the first four shapes, which have the most amount of texts overall, are not the most popular. These emotional arcs' effectiveness as stories implies that readers' emotional experiences have a significant impact on how stories are presented. The three most effective emotional arcs, according to their research, are 'Icarus', 'Oedipus', and 'Man in a Hole'. While Woods did not apply this same principal to her analysis of Stephen King's novels, she too acknowledges that such an additional analysis would be a useful addition to her research.[109] These results could be of practical use for both publishers and authors. Reagan et al. note how generating sentiment graphs can be applied in the opposite direction as well: starting with the emotional arc when structuring or writing a compelling story. Publisher could use this information in producing a more universally liked or better selling work, or support the publication of a book which they see more fit to a 'successful' emotional arc.

Second, the graphs produced by sentiment analysis graphs could be of use in education as well. As discussed earlier, the progression of sentiment is not the same as the progression of plot. Nonetheless, depending on which narratology scholars we base our perception of plot on, emotions do in fact give information about plot to a certain degree. Extracting these emotions and presenting them in the form of a graph provides us with a clear visualization of a large dataset; the progression of valence of an entire novel or play. One way in which these graphs can be used is as a visual tool in teaching narratology and plot structure. In 2007, Dymoch stated that pupils who know the structure of narrative texts could more easily understand stories.[110] He also acknowledges that many students need to be

---

108 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', *EPJ Data Science*, 5 (2016), pp.1-12.

109 D. Woods, 'The emotional arcs of horror: a distant reading of Stephen King's novels', *UVM Honors College Senior Theses*, (2022), https://scholarworks.uvm.edu/hcoltheses/509.

110 S. Dymock, 'Comprehension strategy instruction: Teaching narrative text structure awareness', *The Reading Teacher*, 61 (2007), pp.161-167.

explicitly taught how to understand this form of literature. Teachers are crucial in helping students gain a solid knowledge of the awareness of narrative text structure. Understanding the sentimental trajectory of stories can give students a greater expectation of how similar stories develop, which will help them comprehend the story as a whole more successfully. Williams and Pao made the same reasoning in *Handbook of Reading Interventions*.[111] They too argue that a better understanding of the underlying structures of stories will greatly improve the reading comprehension of children who struggle with narrative apprehension.

Apart from the fact understanding the structure of stories helps in the overall comprehension of a story, researchers have argued as well that it might help in constructing arguments. In 2010, Bex and Bench-Capon argued that, in the context of persuasion, people would often tell a story.[112] As an example, they stated that telling the story of *The Boy Who Cried Wolf* might be a more effective way of persuading someone not to lie, rather than by making the case why one should not fabricate lies when there would be no reason to do so. They suggest more study on the composition of tales and the conditions under which a narrative might persuade an agent in order to reach a decision. Again, the use of data on the progression of sentiment throughout a story could prove to be helpful in this regard as well.

Based off of the suggestions of scholars on how to use sentiment analysis programmes for literature in a more practical manner, it must be said that these applications are not as numerous as the applications for social media texts. Nevertheless, there is definitely room for applying this computational method outside of the field of research as well. Publishers could use this strategy in a variety of ways to increase the possibilities of a novel succeeding. Moreover, researchers generally agree on the importance of teaching underlying narrative structures and narrative theories for people to improve upon comprehensive reading. While the sentiment analysis graphs are not interchangeable with the theoretical approaches of scholars who aim at structuralizing the plot of stories, they could aid as a visual tool in understanding the emotional progression of a story. To what degree the visualization of valence can improve the overall reading comprehension of children must be further researched, but its role in education could be useful in the near future.

---

[111] J.P. Williams, and L.S. Pao, 'Teaching narrative and expository text structure to improve comprehension', in R. O'Connor and P. Vadasy (eds.), *Handbook of reading interventions*, (New York City: The Guilford Press, 2013), pp. 254-278.

[112] F.J. Bex and T. Bench-Capon, 'Persuasive stories for multi-agent argumentation', *2010 AAAI Fall Symposium Series* (2010).

6. Conclusion

This thesis aimed to contribute to the existing body of knowledge on using sentiment analysis to extract plot information from literary texts. To achieve this, the thesis was divided into three main objectives. The first objective involved experimenting with and evaluating a Python-based method for sentiment analysis on literary texts. The second objective was to replicate a study conducted by Reagan et al. in 2016, which identified six common emotional arcs in stories. Lastly, based on the analysis's findings and a review of the literature, a critical evaluation of the use of sentiment analysis on literary texts was undertaken.

For the sentiment analysis in this thesis, the VADER method and its lexicon were utilized as an alternative to the Syuzhet package in R. The aim was to determine if a similarly effective tool could be created. The findings revealed that the VADER lexicon and analysis method produced results similar to those of the Syuzhet package, indicating that a lexicon based on social media texts can perform sentiment analysis almost identically. However, some differences were observed, particularly in the evaluation of individual sentences. The VADER lexicon, being smaller than Syuzhet's, assigned more sentences a neutral sentiment score, while Syuzhet classified these sentences as positive or negative. Additionally, VADER excelled at recognizing and analysing sentences containing negation due to its use of syntactic structure analysis, while Syuzhet relied on word count. These differences offered advantages to each approach, and the overall sentiment graphs appeared almost identical. However, upon comparing the graphs to the plot of the stories, it was discovered that the sentiment progression only partially aligned. Some narrative low points in the tragedies were mistakenly classified as positive, possibly due to Shakespeare's style of writing. Shakespeare’s poetic writing style can often conceal the literal connotation of a sentence, making it difficult for a sentiment analysis tool to extract the emotion expressed by a character. The analysis also failed to recognize the happy endings of the comedies, as the context and overall mood of those scenes could not be accurately measured beyond the words used, which were generally neutral or negative. Furthermore, it's important to note the structure of plays in general, which prioritize dialogue over detailed descriptions and characterizations. This aspect might play a role in the divergent findings. If significant moments, like character deaths, were portrayed with greater details and more exposition, it is possible that the measured sentiment scores could vary. Further research is necessary to fully understand this hypothesis.

Further limitations were identified when analysing the texts on a sentence-to-sentence scale. The sentiment analysis was unable to recognize certain linguistic connotations such as sarcasm or irony, resulting in inaccurately assigned sentiment scores. Additionally, due to the creation of the VADER and Syuzhet lexicons, they were less effective in analyzing the sentiment of Shakespeare's plays, particularly when encountering archaic words. However, these occurrences were infrequent enough that they did not significantly alter the overall shapes of the sentiment graphs.

The VADER sentiment analysis tool was also used to replicate the 'emotional arc' study conducted by Reagan et al. Out of the 1,327 texts analysed by Reagan et al., 60 were replicated in this study. The aim was to analyse the ten stories that came closest to each emotional arc shape as determined by Reagan et al. and assess how closely they adhered to those plot shapes when using a different sentiment analysis method. The findings revealed that while most of these stories followed the predetermined plot shape to some extent, they were not as consistent as indicated by Reagan et al.'s study. The presence of numerous outliers prevented the identification of a recognizable common arc shape. Moreover, when determining the shape of each story through curve fitting, it was discovered that some stories exhibited entirely different arc shapes from what Reagan et al. initially found. These results indicate that replicating Reagan's study for all 1,327 texts would likely result in a different distribution of stories by arc shape overall, potentially contradicting their claim that certain shapes correlate with a story's popularity as indicated by the number of downloads.

Based on the results of this thesis, a comprehensive assessment of sentiment analysis and its application to literary texts was conducted. While sentiment analysis is a valuable tool for extracting plot-related information, it has limitations that must be acknowledged and addressed. Sentiment analysis is unable to fully capture the complexity and nuances of a narrative, and relying solely on sentiment analysis for plot extraction may lead to an incomplete understanding of the story. However, sentiment analysis can provide valuable insights into the emotional progression within a story and serve as a visual aid for understanding the overall sentiment. By integrating sentiment analysis with traditional approaches employed by literary scholars, a more comprehensive understanding of the emotional arcs and narrative structures can be achieved.

The potential applications of sentiment analysis in various fields were also explored in this thesis. In addition to its use in literary analysis, sentiment analysis holds promise in education for enhancing reading comprehension. The visualization of valence and emotional trajectories can assist children in grasping the emotional aspects of a story and better connecting with the characters and events. Furthermore, sentiment analysis can be employed in the field of narratology to examine the evolution of storytelling within specific cultures, enabling comparative analyses of stories across different regions of the world. It also has the potential to uncover variations and similarities in translated literature, opening possibilities for intriguing research questions. The modification of sentiment analysis tools to provide more comprehensive plot information, such as character-to-character sentiment analysis, was demonstrated through the analysis of Shakespeare's play 'Romeo and Juliet.' By tracking the emotional trajectories of interpersonal relationships, a deeper understanding of the story's dynamics can be obtained, surpassing the limitations of traditional sentiment analysis.

In summary, this thesis has contributed to our understanding of sentiment analysis as a technique for extracting plot information from literary texts. The choice of sentiment analysis method significantly influences the identification and interpretation of sentiment progressions in stories, emphasizing the importance of considering the limitations and complexities of different sentiment analysis methods. While sentiment analysis has its limitations, it offers valuable insights into the emotional aspects of a story, complementing traditional approaches employed by literary scholars. The research conducted in this thesis has paved the way for further development and refinement of sentiment analysis techniques tailored to the analysis of narrative structures. By recognizing these complexities and proposing modifications to existing methods, this thesis contributes to a more nuanced understanding of how sentiment analysis can be effectively utilized to extract plot-related information from literary works. Additionally, it creates opportunities for future research to explore the role of sentiment analysis in education and delve into the cultural implications of emotional experiences within stories. This thesis highlights the promising applications of sentiment analysis in narratology and emphasizes the integration of narrative structures into reading comprehension. Further research is needed to fully explore the potential of sentiment analysis in education and address the encountered limitations. Overall, this thesis demonstrates the potential of sentiment analysis in deepening our comprehension of emotions within stories.

7. Bibliography

Al-Garaady, J. and Mahyoob, M., 'Public sentiment analysis in social media on the SARS-CoV-2 vaccination using VADER lexicon polarity', *Humanities and Educational Sciences Journal* (2022), pp.591-609.

Al-Shabi, M.A., 'Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining' *IJCSNS*, *20* (2020), p.1-7.

Blythe, M., Research fiction: storytelling, plot and design. '*Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems'*, pp. 5400-5411.

Bonta, V. and Janardhan, N., 'A comprehensive study on lexicon based approaches for sentiment analysis', *Asian Journal of Computer Science and Technology*, 8 (2019), pp.1-6.

Boyd, R.L., Blackburn, K.G. and Pennebaker, J.W., 'The narrative arc: Revealing core narrative structures through text analysis', *Science advances*, *6* (2022), n.pag.

Campbell, J., *The hero with a thousand faces*, (New York City: Pantheon Books, 1949).

Chai, T. and Draxler, R.R., 'Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature', *Geoscientific model development*, *7* (2014), pp.1247-1250.

Dale, K.R., Fisher, J.T., Liao, J. and Grinberg, E., 'The Shape of Inspiration: Exploring the Emotional Arcs and Self-Transcendent Elicitors Within Inspirational Movies', *Media Psychology*, 2023, pp. 1-23.

Dymock, S., 'Comprehension strategy instruction: Teaching narrative text structure awareness', *The Reading Teacher*, 61 (2007), pp.161-167.

Elkins, K. and Chun, J., 'Can Sentiment Analysis Reveal Structure in a Plotless Novel?', *arXiv preprint*, (2019).

Elsner, M., 'Character-based kernels for novelistic plot structure', *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics,* 1 (2012), pp. 634-644.

Finkelberg, M., 'Aristotle and episodic tragedy', *Greece & Rome*, *53* (2006), pp. 60-72.

French, K., 'Kurt Vonnegut Graphs the Shapes of Stories', *Visage*, 13 August, 2014. < https://visage.co/kurt-vonnegut-shows-us-shapes-stories/> (14 September, 2022).

Hutto, C.J., 'vaderSentiment', *github*, 1 april, 2022. < https://github.com/cjhutto/vaderSentiment> (1 June, 2023).

Garcia-Rudolph, A., Laxe, S., Saurí, J. and Bernabeu, M.G., 'Stroke survivors on twitter: sentiment and topic analysis from a gender perspective', *Journal of medical Internet research*, *21* (2019), n.pag.

Goodman, S.N., Fanelli, D. and Ioannidis, J.P., 'What does research reproducibility mean?', *Science translational medicine*, 8 (2016), pp. 1–6.

Grover, N., 'Just say no: negativity is secret of political tweet success, study finds', *The Guardian*, 13 April, 2021. <https://www.theguardian.com/technology/2021/apr/13/just-say-no-negativity-is-secret-of-political-tweet-success-study-finds>

Holbrook, B., Penders, B., de Rijcke, S., 'The humanities do not need a replication drive', Centre for Science and Technology Studies, 21 January, 2019. < https://www.cwts.nl/blog?article=n-r2v2a4&title=the-humanities-do-not-need-a-replication-drive> (7 July, 2022).

Hogan, P.C., *Affective narratology: The emotional structure of stories* (Lincoln: University of Nebraska Press, 2011).

Hu, Q., Liu, B., Thomsen, M.R., Gao, J. and Nielbo, K.L., 'Dynamic evolution of sentiments in Never Let Me Go: Insights from quantitative analysis and implication, *6th International Conference on Behavioral, Economic and Socio-cultural Computing (BESC),* (2019), (pp. 1-6).

Hutto, C. and Gilbert, E., '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*, 8, (2014), pp. 216-225.

Isasi, J., 'Sentiment analysis methods in translation', *Conference presentation for Digital Frontiers Annual Conference, Austin, TX, United States*, (2019).

Jain, V., 'Prediction of movie success using sentiment analysis of tweets', *The International Journal of Soft Computing and Software Engineering*, 3 (2013), pp. 308-313.

Jockers, M., *Macroanalysis: Digital methods and literary history* (Champaign: University of Illinois Press, 2013).

Jockers, M., 'Revealing Sentiment and Plot Arcs with the Syuzhet Package', *matthewjockers.net*, 2 February, 2015. <https://www.matthewjockers.net/2015/02/02/syuzhet/> (12 September, 2022).

Jockers, M. and, Thalken R., 'Sentiment analysis', *Text Analysis with R: For Students of Literature*, (2020), pp. 159-174.

Jockers, M., 'Syuzhet Package in R', RDocumentation, December 14, 2017. < https://www.rdocumentation.org/packages/syuzhet/versions/1.0.4> (2 September, 2022).

Kennedy, H., ‘Perspectives on sentiment analysis’, *Journal of Broadcasting & Electronic Media*, *56* (2012), pp. 435-450.

Kim, E. and Klinger, R., ‘A survey on sentiment and emotion analysis for computational literary studies’, *arXiv preprint arXiv:1808.03137*, (2018).

Kim, H., ‘Sentiment Analysis: Limits and Progress of the Syuzhet Package and Its Lexicons.’ *DHQ*, 16 (2022), n.pag.

Kukkonen, K., ‘Plot’, in Hühn, P., Meister, J., Pier, J. And Schmid, W., *Handbook of narratology*, (Berlin: de Gruyter, 2014), pp. 706-719.

Larose, C.D. and Larose, D.T., *Data science using Python and R* (New Jersey: John Wiley & Sons, 2019).

Liu, B., *Sentiment analysis: Mining opinions, sentiments, and emotions* (Cambridge: Cambridge university press, 2020).

Luhmann, J. and Burghardt, M., ‘Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape’, *Journal of the Association for Information Science and Technology*, *73* (2022), pp.148-171.

Ma, L.D., ‘Evaluation of product conceptual design based on Pythagorean fuzzy set under big data environment’, *Scientific Reports*, *12* (2022), pp.1-20.

Matalon, Y., Magdaci, O., Almozlino, A. and Yamin, D., ‘Using sentiment analysis to predict opinion inversion in Tweets of political communication’, *Scientific reports*, 11 (2021), pp. 1-9.

Min, S. and Park, J., ‘Narrative as a Complex Network: A Study of Victor Hugo's Les Misérables’, *Proceedings of HCI Korea*, (2016), pp. 100-107.

Moretti, F., ‘Network theory, plot analysis’, *New Left Review*, 68 (2011), pp. 80-102.

Nalisnick, E.T. & Baird, H.S., ‘Character-to-character sentiment analysis in Shakespeare’s play’, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2 (2013), pp. 479-483.

Onega, S. and Landa, J.A.G., *Narratology: an introduction* (Abingdon, Oxfordshire: Routledge, 2014).

Pang, B., Lee L., and Vaithyanathan, S., ‘Thumbs up? Sentiment Classification using Machine Learning Techniques’, *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (2002), pp. 79–86.

Pang, B., and Lillian L., ‘Opinion mining and sentiment analysis’, *Foundations and Trends in information retrieval*, 1–2 (2008), pp. 1-135.

Peels, R., Replicability and replication in the humanities. *Research Integrity and Peer Review*, 4 (2019), pp.1-12.

Peels, R. and Bouter, L., 'The possibility and desirability of replication in the humanities', *Palgrave Communications*, 4 (2018), pp. 1-4.

Rank, O., Raglan, F. and Dundes, A., *In quest of the hero* (Princeton: Princeton University Press, 1990).

Reagan, A.J., Mitchell, L., Kiley, D., Danforth, C.M. and Dodds, P.S., 'The emotional arcs of stories are dominated by six basic shapes', *EPJ Data Science*, 5 (2016), pp.1-12.

Rebora, S., 'Sentiment Analysis in Literary Studies. A critical Survey', *Digital Humanities Quarterly*, ????

Russell, J.A., 'A circumplex model of affect', *Journal of personality and social psychology*, 39 (1980), pp. 1161-1178.

Shaheen, E., 'Lost in Translation: Using Sentiment Analysis to Analyze Translations of Homer's Odyssey', (2019).

Singh, M., Jakhar, A. and Pandey, S., 'Sentiment analysis on the impact of coronavirus in social life using the BERT model', *Social Network Analysis and Mining*, *11* (2021), pp. 1-11.

Sprugnoli, R., Tonelli, S., Marchetti, A. and Moretti, G., 'Towards sentiment analysis for historical texts', *Digital Scholarship in the Humanities*, 31 (2016), pp. 762-772.

Swafford, A., 'Why Syuzhet Doesn't Work and How We Know', *Anglophile in Academia: Annie Swafford's Blog*, March 30, 2015. <https://annieswafford.wordpress.com/2015/03/30/why-syuzhet-doesnt-work-and-how-we-know/> (5 September, 2022).

Sonntag, J. and Manfred S., 'Sentiment analysis: What's your opinion?', in Biemann, C. and Mehler, A., (eds.), *Text mining*, *From Ontology Learning to Automated Text Processing Applications*, (New York City: Springer, 2014), pp. 177-199.

Tu, C. and Brown, S., 'Character mediation of plot structure: Toward an embodied model of narrative', *Frontiers of narrative studies*, *6* (2020), pp. 77-112.

Turney, P., 'Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews', *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02),* (2002), pp. 417-424.

Vonnegut, K., 'Kurt Vonnegut on the Shapes of Stories', *Youtube*, 30 October, 2010. < https://www.youtube.com/watch?v=oP3c1h8v2ZQ&ab_channel=DavidComberg> (14 September, 2022).

Walsh, J., Vaida, N., Coman, A. and Fiske, S.T., 'Stories in Action', *Psychological Science in the Public Interest*, 23 (2022), pp. 99-141.

Woods, D., ‘The emotional arcs of horror: a distant reading of Stephen King’s novels’, *UVM Honors College Senior Theses*, (2022), https://scholarworks.uvm.edu/hcoltheses/509.

Xie, H., Lin, W., Lin, S., Wang, J. and Yu, L.C., ‘A multi-dimensional relation model for dimensional sentiment analysis’, *Information Sciences*, 579 (2021), pp. 832-844.

Zehe, A., Becker, M., Hettinger, L., Hotho, A., Reger, I. and Jannidis, F., ‘Prediction of happy endings in German novels based on sentiment information’, In *3rd Workshop on Interactions between Data Mining and Natural Language Processing, Riva del Garda, Italy*, (2016), pp. 9-16.

8. Appendix

Appendix A: the sixty analysed texts, organized by arc shape

Title	Author
The Winter's Tale	William Shakespeare
The Terror: A Mystery	Arthur Machen
The Pilgrim’s Progress in Words of One Syllable	Lucy Aikin and John Bunyan
The Road to Oz	L. Frank Baum
The Winter's Tale	William Shakespeare
The Road to Oz (Alternate Version)	L. Frank Baum
As You Like It	William Shakespeare
Eastern Standard Tribe	Cory Doctorow
Sir Thomas More	William Shakespeare
The Spanish Tragedy	Thomas Kyd

Figure 8.1: The ten analysed texts for arc shape 1

Title	Author
The Duchess of Padua	Oscar Wilde
The Island of Doctor Moreau	H. G. Wells
Lady Susan	Jane Austen
Lady Windermere's Fan	Oscar Wilde
The House of the Vampire	George Sylvester Viereck
The Star Lord	Boyd Ellanby
Tom Sawyer, Detective	Mark Twain
Tragedy of Romeo and Juliet	William Shakespeare
Warlord of Kor	Terry Carr
Anna Christie	Eugene O'Neill

Figure 8.2: The ten analysed texts for arc shape 2

Title	Author
Justice	John Galsworth
R. Holmes & Co. Being the Remarkable Adventures of Raffles Holmes, Esq., Detective and Amateur Cracksman by Birth	John Kendrick Bangs
The Life and Adventures of Santa Claus	L. Frank Baum
Tamburlaine the Great – Part 1	Christopher Marlowe
The Door Through Space	Marion Zimmer Bradley
The Great Gray Plague	Raymond F. Jones
The Magic of Oz	L. Frank Baum
The Man of Feeling	Henry Mackenzie
The Sky Is Falling	Lester Del Rey
A Yankee Flier Over Berlin	Rutherford G. Montgomery

Figure 8.3: The ten analysed texts for arc shape 3

Title	Author
Hamlet, Prince of Denmark	William Shakespeare
The Bobbsey Twins	Laura Lee Hope
The Daffodil Mystery	Edgar Wallace
The Food Of The Gods and How It Came to Earth	H. G. Wells
The Moneychangers	Upton Sinclair
The Oakdale Affair	Edgar Rice Burroughs
The Slayer Of Souls	Robert W. Chambers
The Sport of the Gods	Paul Laurence Dunbar
The Way of the World	William Congreve
Twelfth Night; Or, What You Will	William Shakespeare

Figure 8.4: The ten analysed texts for arc shape 4

Title	Author
After London; Or, Wild England	Richard Jefferies
Dave Dawson at Dunkirk	Robert Sidney Bowen
Mystery of the Hasty Arrows	Anna Katharine Green
Once On A Time	A. A. Milne
Ride Proud, Rebel!	Andre Norton
That Affair At Elizabeth	Burton Egbert Stevenson
The Haunted Man and the Ghost's Bargain	Charles Dickens
The Scarecrow Of Oz	L. Frank Baum
The Shadow of the Rope	E. W. Hornung
Trough The Magic Door	Arthur Conan Doyle

Figure 8.5: The ten analysed texts for arc shape 5

Title	Author
Old Indian Days	Charles A. Eastman
Pariah Planet	Murray Leinster
The Blue Bird for Children	Georgette Leblanc and Maurice Maeterlinck
The Cosmic Computer	H. Beam Piper
The Evil Guest	Joseph Sheridan Le Fanu
The Monster Men	Edgar Rice Burroughs
The Revolt of the Star Man	Raymond Z. Gallun
The Wind in the Willows	Kenneth Grahame
The Wind In The Willows (Alternate Version)	Kenneth Grahame
This World Is Taboo	Murray Leinster

Figure 8.6: The ten analysed texts for arc shape 6

Appendix B: a plot-by-plot comparison of VADER and Syuzhet's results

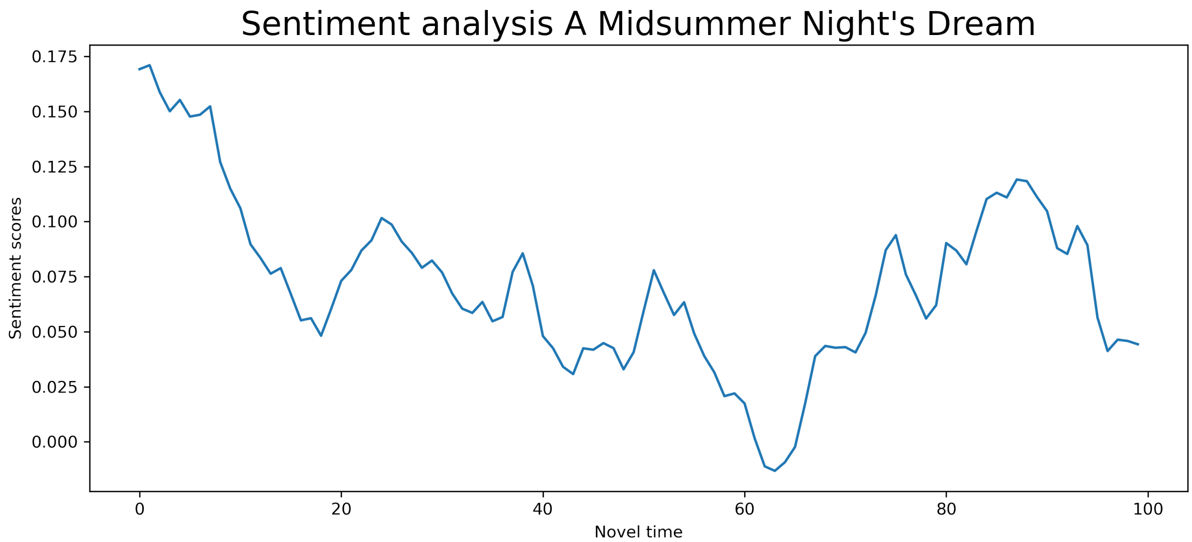

Figure 8.7: VADER sentiment analysis results for *A Midsummer Night's Dream*

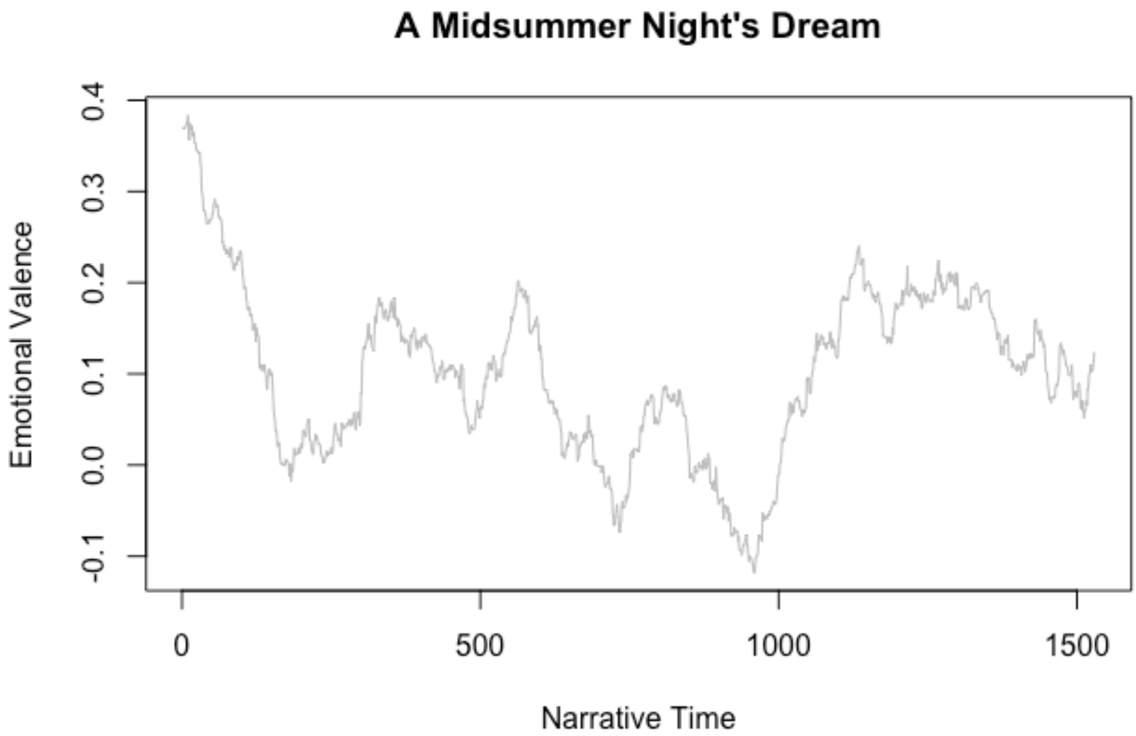

Figure 8.8: Syuzhet sentiment analysis results for *A Midsummer Night's Dream*

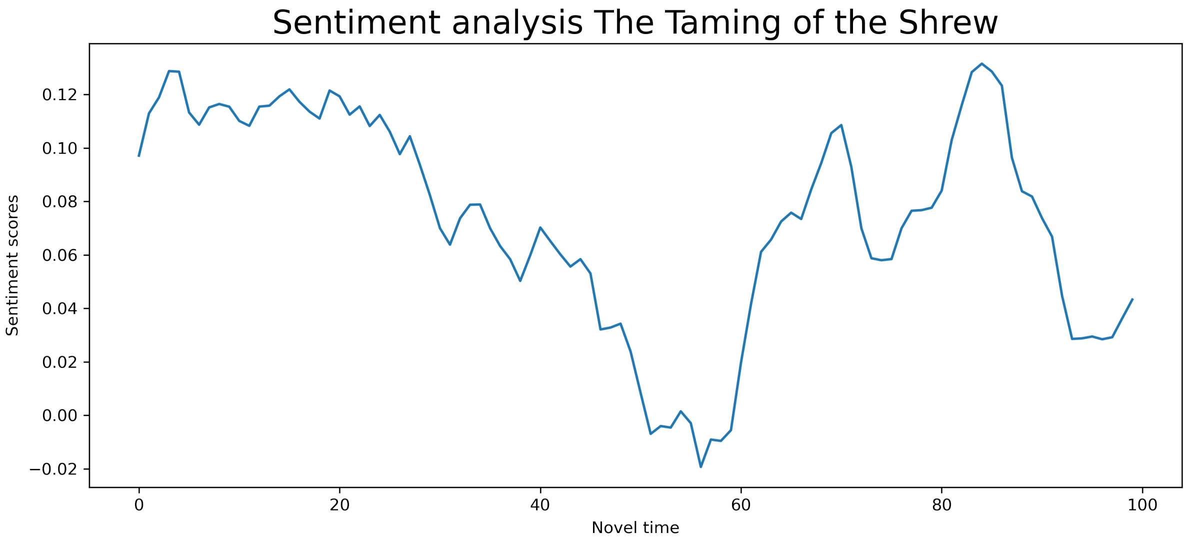

Figure 8.9: VADER sentiment analysis results for *The Taming of the Shrew*

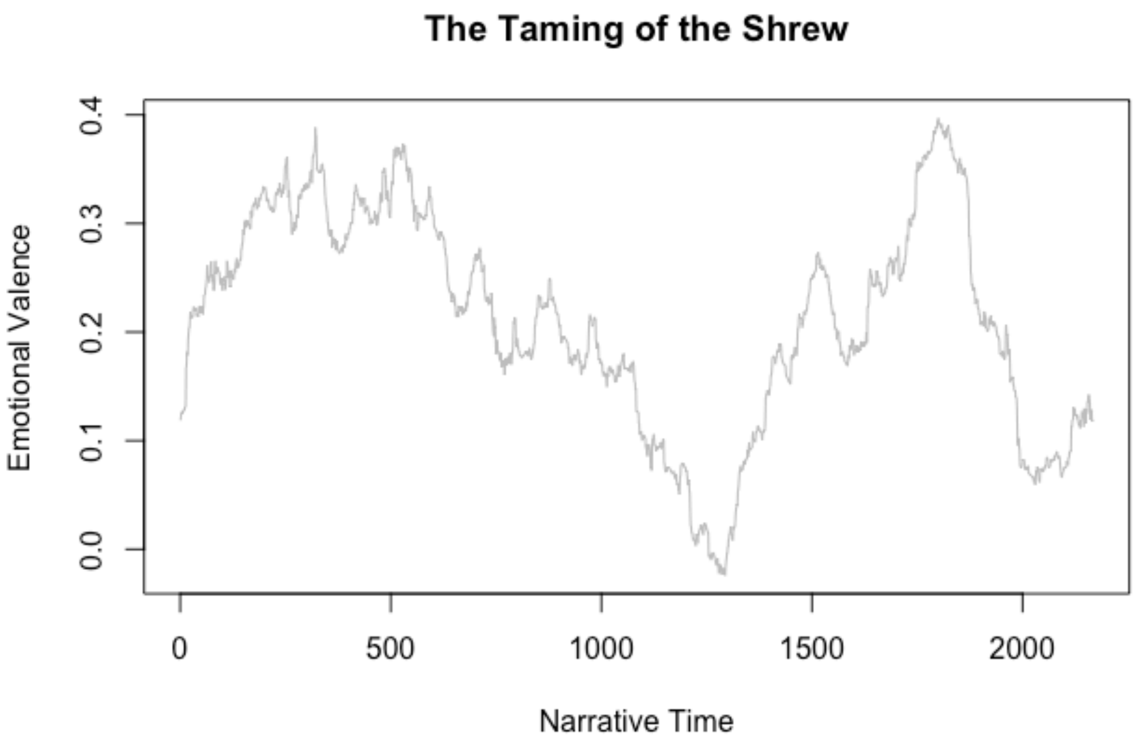

Figure 8.10: Syuzhet sentiment analysis results for *The Taming of the Shrew*

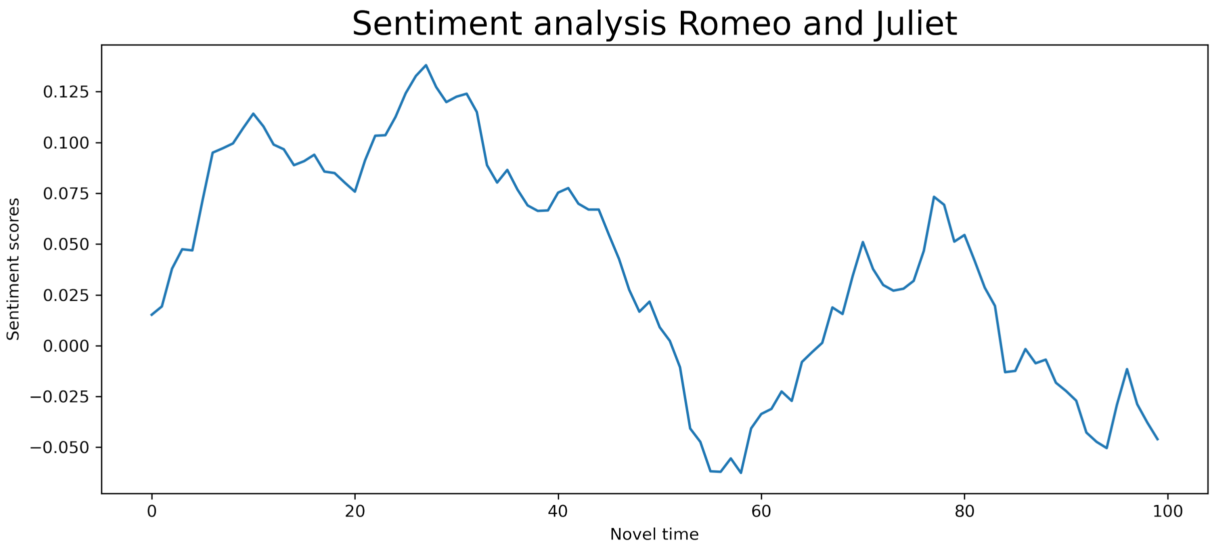

Figure 8.11: VADER sentiment analysis results for *Romeo and Juliet*

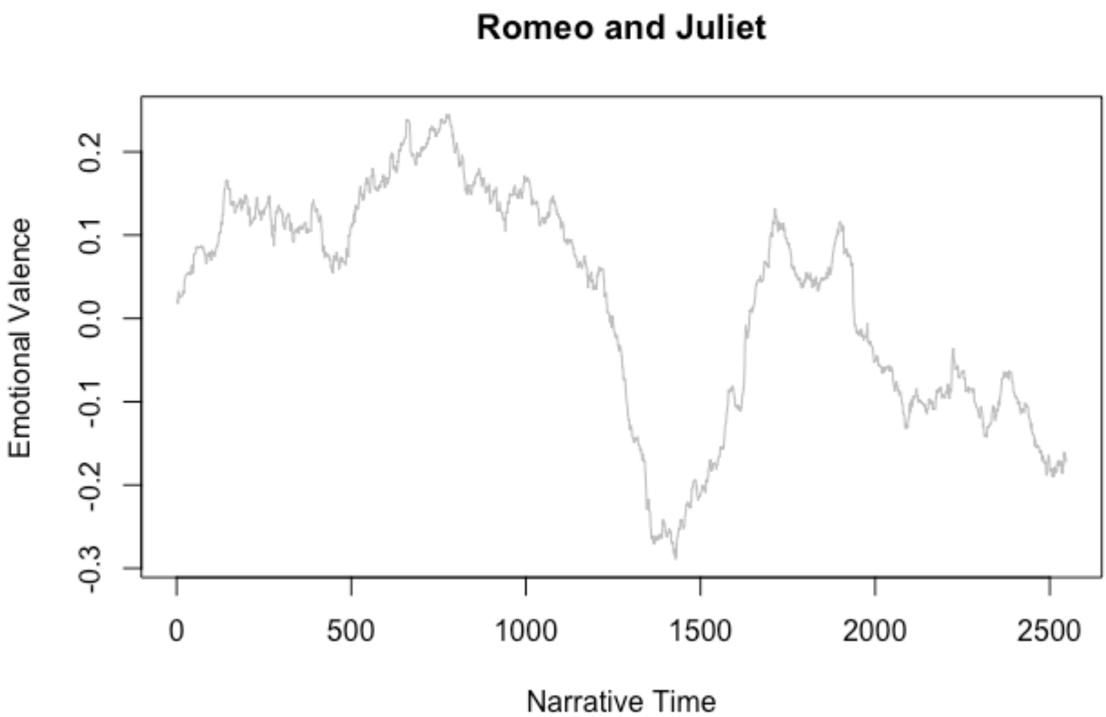

Figure 8.12: Syuzhet sentiment analysis results for *Romeo and Juliet*

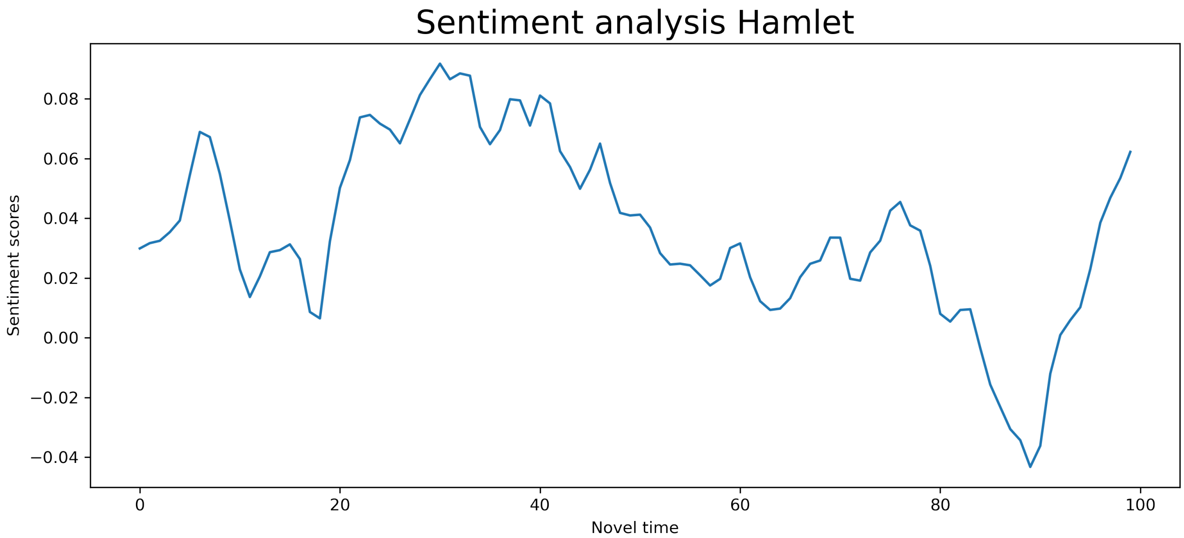

Figure 8.13: VADER sentiment analysis results for *Hamlet*

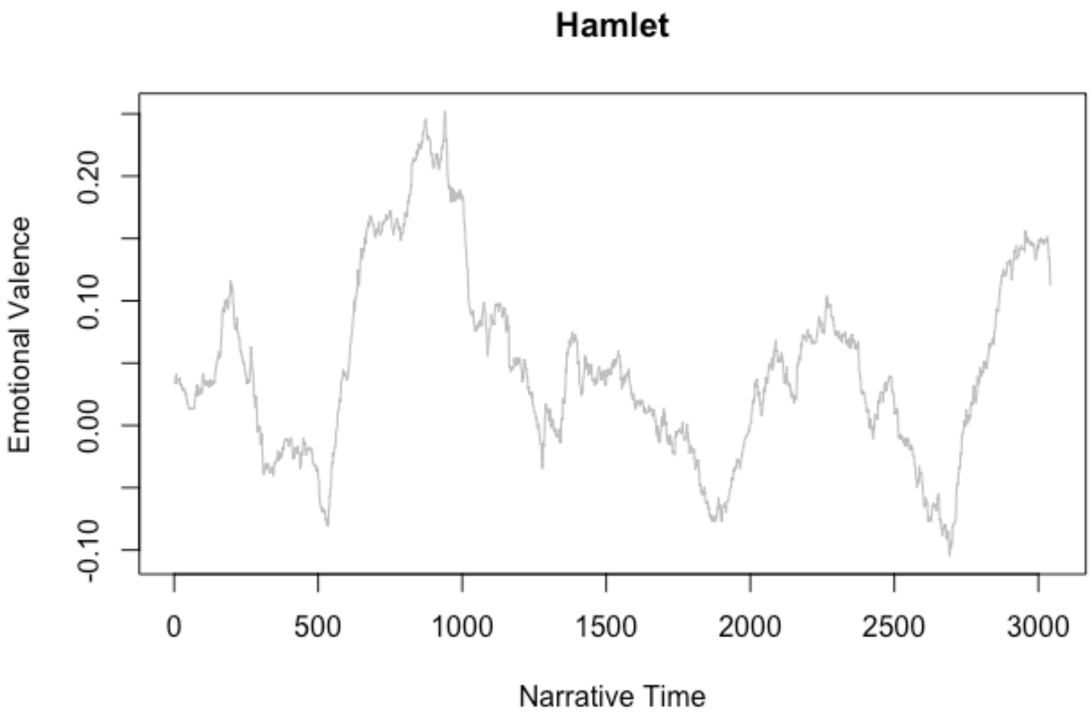

Figure 8.14: Syuzhet sentiment analysis results for *Hamlet*

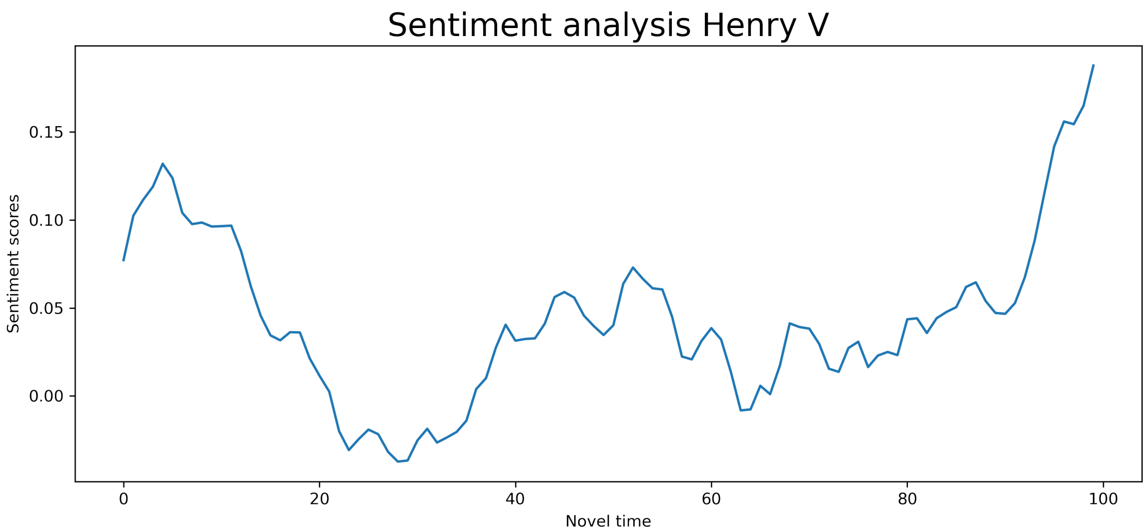

Figure 8.15: VADER sentiment analysis results for *Henry V*

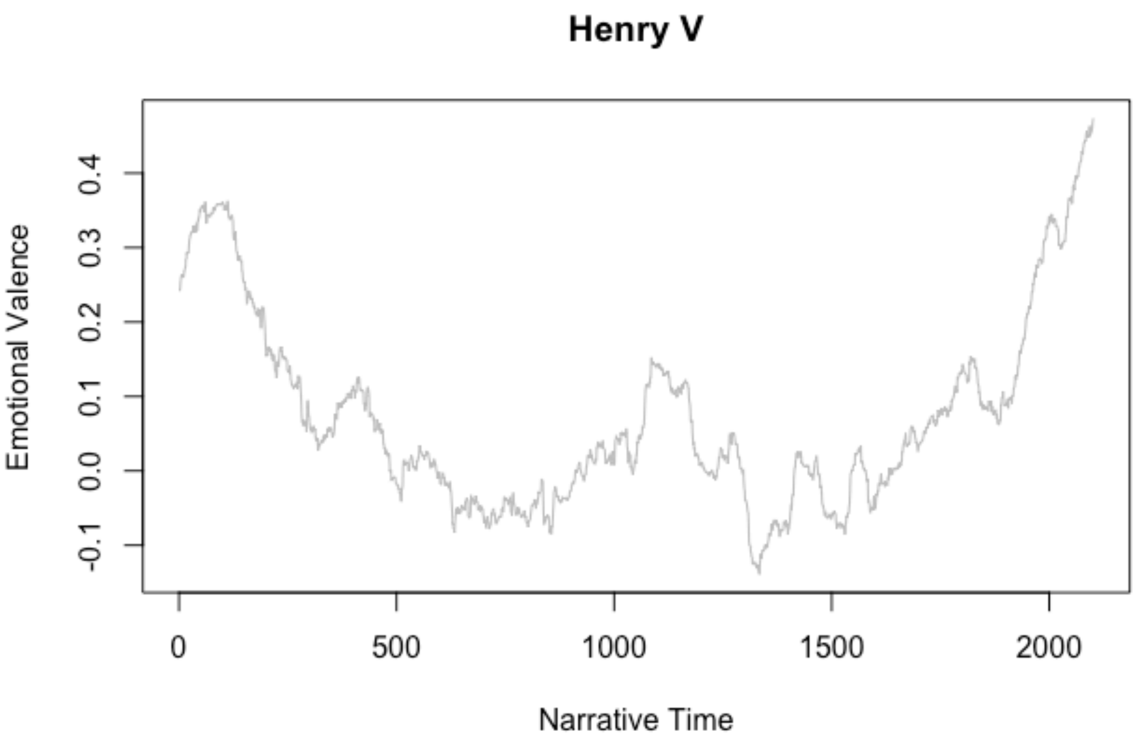

Figure 8.16: Syuzhet sentiment analysis results for *Henry V*