

Emotion and Modifier in Henry Rider Haggard’s Novels

Salim Sazed

Department of Computer Science
Old Dominion University
Norfolk, VA, 23529
salim.sazed@gmail.com

Abstract

In recent years, there has been a growing scholarly interest in employing quantitative methods to analyze literary texts, as they offer unique insights, theories, and interpretations. In light of this, the current study employs quantitative analysis to examine the fiction written by the renowned British adventure novelist Sir Henry Rider Haggard. Specifically, the study aims to investigate the affective content and prevalence of distinctive linguistic features in six of Haggard’s most distinguished works. We evaluate dominant emotional states at the sentence level as well as investigate the deployment of specific linguistic features such as modifiers and deontic modals, and collocated terms. Through sentence-level emotion analysis, the findings reveal a notable prevalence of *joy*-related emotions across the novels. Furthermore, the study observes that intensifiers are employed more commonly than the mitigators as modifiers and the collocated terms of modifiers exhibit high similarity across the novels. By integrating quantitative analyses with qualitative assessments, this study presents a novel perspective on the patterns of emotion and specialized grammatical features in some of Haggard’s most celebrated literary works.

1 Introduction

Henry Rider Haggard (1856-1925) was a prominent British novelist and acclaimed adventure fiction writer known for his captivating tales set in exotic locations, particularly Africa. He is considered a pioneer in the lost world genre, characterized by thrilling narratives of exploration and discovery in remote and enigmatic places. One of his most famous series of novels features the adventures of Allan Quatermain, a white hunter. Haggard’s works are notable for their vivid descriptions of African landscapes, depictions of African culture, and imaginative portrayals of ancient civilizations. In addition to the Allan Quatermain series, Haggard’s novels *She* and its sequel, *The Return of*

She, have gained widespread recognition as seminal examples of imperialistic fiction, showcasing a fusion of adventure, romanticism, and supernatural elements.

Digital humanities is an interdisciplinary field that combines humanities disciplines such as history, literature, and philosophy with computer science and technology to study and create new forms of digital culture (Burdick et al., 2016). In recent years, interest in quantitative analysis of digital humanities has been on the rise with the help of accessible tools and methodologies that encompass a range of approaches, including natural language processing (NLP) and text mining, network analysis, data visualization, statistical analysis, and machine learning (Sazed, 2022; Levine, 2022). More researchers and professionals in this field are becoming interested in using numerical and statistical methods to study various cultural and humanistic phenomena. Researchers have been using NLP and text mining techniques to analyze text depicting literary works, historical documents, or online archives for diverse purposes (Samothrakis and Fasli, 2015; Simonton, 1990; Dinu and Uban, 2017; San Segundo, 2017; Stockwell and Mahlberg, 2015).

In this study, we focus on analyzing the emotional and specific linguistic features of six of Henry Rider Haggard’s most celebrated novels, employing a variety of natural language processing (NLP) techniques. In particular, we aim to explore the following research aspects-

RQ1: How emotional tones are illustrated in Henry Rider Haggard’s most popular classics?

RQ2: Whether the usage of two linguist features: modifiers and deontic modals vary across Haggard’s popular novels?

We first analyze the presence of various forms of emotions across the six novels by scrutinizing the distribution of emotions at the sentence level. We find similar patterns of emotions at the sentence

level in all six novels, with *joy* being the most frequently occurring emotion. In addition, we conduct a linguistic analysis to identify the occurrence of specific linguistic phenomena, such as the usage of mitigators, intensifiers, and deontic modals. Our results indicate that although the comparative presence of intensifiers and modifiers varies across the novels a bit, in general, the percentages are similar, within the range of 0.2%- 0.3% for intensifiers and 0.19%-0.25% for mitigators. Overall, our findings indicate substantial degrees of consistency in all the attributes studied across all six novels.

2 Dataset

The six literary works, namely *King Solomon’s Mines* (KSM), *Allan Quatermain* (AQ), *The Holy Flower* (HF), *The Ivory Child* (IC), *She* (SHE), and *Ayesha, the Return of She* (ARS), are obtained from the Project Gutenberg¹ library as UTF-8 formatted text files. To ensure only literary content is analyzed, we manually remove the metadata present in the text file of each novel. The NLTK tokenizer (Bird et al., 2009) is employed to segment the text of each novel into sentences. Very short sentences containing fewer than three words are excluded from the analysis. The resulting dataset is summarized in Table 1.

Novel	#Sentence	#Words	#Words/Sent.
KSM	3251	81078	24.93
AQ	3801	104942	27.61
HF	4995	119918	24.00
IC	4190	111884	26.70
SHE	3977	111192	27.95
ARS	4504	116175	25.79

Table 1: Length related statistics of six novels

3 Emotion Analysis

Emotion analysis in literature is the study of emotions and sentiments expressed in written works, such as novels, poems, and short stories employing computational and linguistic methods. Emotion analysis can recognize emotional words and phrases, identify patterns of emotion over time, and categorize emotions into broad categories, such as joy, anger, or sadness. Here, we explore the distributions of prevalent emotions at the sentence-level.

¹<https://www.gutenberg.org>

We utilize the EmoNet emotion recognition framework (Abdul-Mageed and Ungar, 2017) to ascertain the prevailing emotions at the sentence level. The EmoNet framework can identify eight primary categories of emotions (Plutchik, 1980), namely *joy* (JOY), *anticipation* (ANT), *surprise* (SUR), *trust* (TRU), *anger* (ANG), *disgust* (DIS), *fear* (FEA), and *sadness* (SAD) in text. We compute the relative frequencies of each primary emotion category in each of the novels and report their respective distributions. It should be noted that according to the authors of EmoNet, each primary emotion category in EmoNet encompasses three related types (i.e., subcategories) of emotions, as defined by Plutchik (1980). For example, *joy* encompasses the following three types of emotions- *ecstasy*, *joy*, and *serenity*. Therefore, overall, 24 types of emotions are considered in this study.

4 Specialized Modifiers

We analyze the presence of three specific types of linguistic feature, intensifier, mitigator and deontic modal, which can be grouped under a broader category of modifiers. Intensifiers and mitigators allow the precise representation of attitudes and opinions by adapting the strength or weakness of the language to correspond to the circumstance. On the other hand, the deontic modal expresses obligations, permissions, or requirements in relation to actions or events.

4.1 Intensifier

An intensifier is a word or phrase employed to strengthen or increase the impact of an adjective, adverb, or verb in a sentence. Intensifiers are used to express degree or emphasis and can help to convey the speaker’s attitude or level of certainty about the information being communicated. Some common intensifiers include- *very*, *quite*, *absolutely*, *totally*, *completely*, and *utterly*. In addition, we examine which words are collocated with the top intensifiers.

4.2 Mitigator

A mitigator is a word or phrase used to soften or lessen the impact of an adjective, adverb, or verb in a sentence. Similar to intensifiers, mitigators are used to articulate degree or emphasis; however, they have the opposite effect of the intensifier. Instead of strengthening the impact of a word, mitigators weaken it. Some common mitigators include

Novel	Emotion type							
	ANG (%)	ANT (%)	DIS (%)	FEA (%)	JOY (%)	SAD (%)	SUR (%)	TRU (%)
KSM	10.24	3.51	7.35	13.32	46.94	11.17	5.97	1.51
AQ	11.52	1.74	7.71	12.71	47.75	11.89	5.34	1.34
HF	11.93	2.36	10.01	13.23	41.70	13.85	5.51	1.4
IC	12.67	1.62	8.57	12.96	42.89	13.6	6.35	1.34
SHE	12.79	1.77	7.25	11.29	48.45	11.49	5.55	1.4
ARS	10.52	1.87	7.17	12.21	48.51	12.46	5.77	1.49

Table 2: Distributions of dominant emotions (%) at sentence level in six novels

Novel	Intensifier (%)	Top intensifiers (with frequency)
KSM	0.202 (%)	very: 122, really: 14, utterly: 12, absolutely: 4
AQ	0.322 (%)	very: 248, really: 36, utterly: 16, absolutely: 10, particularly: 6
HF	0.299 (%)	very: 275, really: 53, extremely: 9, particularly: 4, absolutely: 4
IC	0.302 (%)	very: 280, really: 29, absolutely: 6, utterly: 6, extremely: 4
SHE	0.299 (%)	very: 244, absolutely: 29, utterly: 20, really: 14, particularly: 7
ARS	0.182 (%)	very: 187, really: 7, utterly: 7, absolutely: 3, extraordinarily: 3

Novel	Mitigator (%)	Top mitigators (with frequency)
KSM	0.192 (%)	quite: 52, rather: 45, almost: 32, pretty: 10, somewhat: 9
AQ	0.215 (%)	quite: 75, almost: 58, rather: 52, pretty: 17, somewhat: 8
IC	0.278 (%)	quite: 151, rather: 81, almost: 51, somewhat: 17, pretty: 6
HF	0.248 (%)	quite: 118, rather: 95, almost: 45, somewhat: 14, pretty: 13
SHE	0.207 (%)	quite: 61, rather: 61, almost: 57, fairly: 13, somewhat: 12
ARS	0.139 (%)	quite: 51, rather: 46, almost: 28, somewhat: 27, faintly: 5

Table 3: Percentage of intensifiers and mitigators in six novels along with the frequency of top intensifiers and mitigators

Novel	Deontic modal (%)
KSM	0.61% (could:161, should:108, must:95)
AQ	0.56% (could:241, should:123, may:68)
HF	0.77% (could:259, should:223, must:160)
IC	0.78% (could:276, should:184, might:132)
SHE	0.65% (could:244, should:144, must:129)
ARS	0.80% (could:293, must:188, should:170)

Table 4: Percentages and occurrences of deontic modals in all six novels

As Table 4 depicts, the prevalent deontic modals exhibit a similar distribution across all novels. *Could* is the most frequently occurring deontic modal in all six novels, followed by *should* in all cases, except the SHE. We observe a consistent presence of the deontic modals in all novels, ranging from 0.61% to 0.80%.

6 Summary and Future Work

As a preliminary study, here, we scrutinize the emotional and specific linguistic aspects of six celebrated works of Henry Rider Haggard leveraging

various NLP techniques. The emotion recognition framework reveals similar patterns of emotions in all six novels, with *joy* being the most dominant. The linguistic analysis uncovers the frequency and presence of modifiers, intensifiers, and deontic modals and the collocated words and phrases. Overall, this research observes uniformity in the examined features across all six novels. The findings of this preliminary study reveal emotional and specific linguistic aspects of some of Haggard’s most celebrated works.

Some possible future works will focus on a fine-grained analysis of emotion, such as identifying sub-categories of primary emotions and understanding the changes of emotions throughout the story, and finding how it is related to plot twists and other narrative elements. Besides, we will encompass an augmented set of linguistic features to conduct a more exhaustive analysis. Furthermore, additional novels authored by Henry Rider Haggard from multiple genres will be explored to find the consistency and divergence of linguistic and psychological patterns.

References

- Muhammad Abdul-Mageed and Lyle Ungar. 2017. [EmoNet: Fine-grained emotion detection with gated recurrent neural networks](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 718–728, Vancouver, Canada. Association for Computational Linguistics.
- Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc."
- Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. 2016. *Digital Humanities*. Mit Press.
- Liviu P Dinu and Ana Sabina Uban. 2017. Finding a character's voice: Stylome classification on literary characters. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 78–82.
- Lauren Levine. 2022. The distribution of deontic modals in jane austen's mature novels. In *Proceedings of the 6th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 70–74.
- Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier.
- Spyridon Samothrakis and Maria Fasli. 2015. Emotional sentence annotation helps predict fiction genre. *PloS one*, 10(11):e0141922.
- Pablo Ruano San Segundo. 2017. Reporting verbs as a stylistic device in the creation of fictional personalities in literary texts. *Atlantis*, pages 105–124.
- Salim Sazed. 2022. An annotated dataset and automatic approaches for discourse mode identification in low-resource bengali language. In *Proceedings of the Workshop on Multilingual Information Access (MIA)*, pages 9–15.
- Dean Keith Simonton. 1990. Lexical choices and aesthetic success: A computer content analysis of 154 shakespeare sonnets. *Computers and the Humanities*, 24:251–264.
- Peter Stockwell and Michaela Mahlberg. 2015. Mind-modelling with corpus stylistics in david copperfield. *Language and Literature*, 24(2):129–147.