A Corpus-Based Analysis of the Language Features and Plot Representation in *Jane Eyre*

Yinghui Tian[a], *Limin Li

School of Foreign Studies, Northwestern Polytechnical University, Shaanxi, China

[a]1187618777@qq.com; *llm2017@nwpu.edu.cn

Abstract. Corpus-based research methods, widely applied in literary analysis, utilize various tools to examine literary works. This paper employs the corpus retrieval software Wordsmith Tools 7.0 to generate a word list for the literary work *Jane Eyre*. The 50 most frequently occurring words are extracted to further explore the linguistic features of the novel. Subsequently, the Concordance Plot in Wordsmith Tools 7.0 is utilized to predict the plot development of the novel. The objective data obtained not only deepens the reader' s understanding of the linguistic and plot details of *Jane Eyre* but also supplements traditional literary perspectives on the novel.

Keywords: Jane Eyre; Corpus Retrieval; Linguistic Features; Plot

1 Introduction

The use of computational methods in literary research is not a new topic. Literary studies are constantly exploring new avenues and developing new tools through contemporary information technology, represented by big data and artificial intelligence [6]. Corpus linguistics, as an emerging research methodology, has been reasonably applied to studies of various languages. This method relies on a substantial corpus and objective corpus retrieval software to process and organize texts. Through this method, it is possible to statistically analyze the frequency of words appearing in the text, the collocations of words, and the density of keywords across different chapters. Traditional literary studies have typically explored the connotations of a work based on a particular literary theory. However, such analyses inevitably carry a significant subjective element. The rational research methodology of corpus linguistics addresses this subjectivity, offering new interpretive pathways for appreciating literary works.

Currently, many foreign scholars apply corpus linguistics to study language and meaning in literary texts. For example, Stubbs[9] used qualitative corpus methods to analyze the frequency, distribution, and recurring occurrence of individual words and phrases in Joseph Conrad' s Heart of Darkness. He found that corpus methods can not only verify conclusions of previous literary criticism but also reveal hidden detailed features in the textual corpus. Culpeper [3] applied keyword analysis to Shakespeare' s Romeo and Juliet. In addition, scholars have conducted in-depth studies on high-

© The Author(s) 2023

E. Marino et al. (eds.), *Proceedings of the 2023 5th International Conference on Literature, Art and Human Development (ICLAHD 2023)*, Advances in Social Science, Education and Humanities Research 806,
https://doi.org/10.2991/978-2-38476-170-8_80

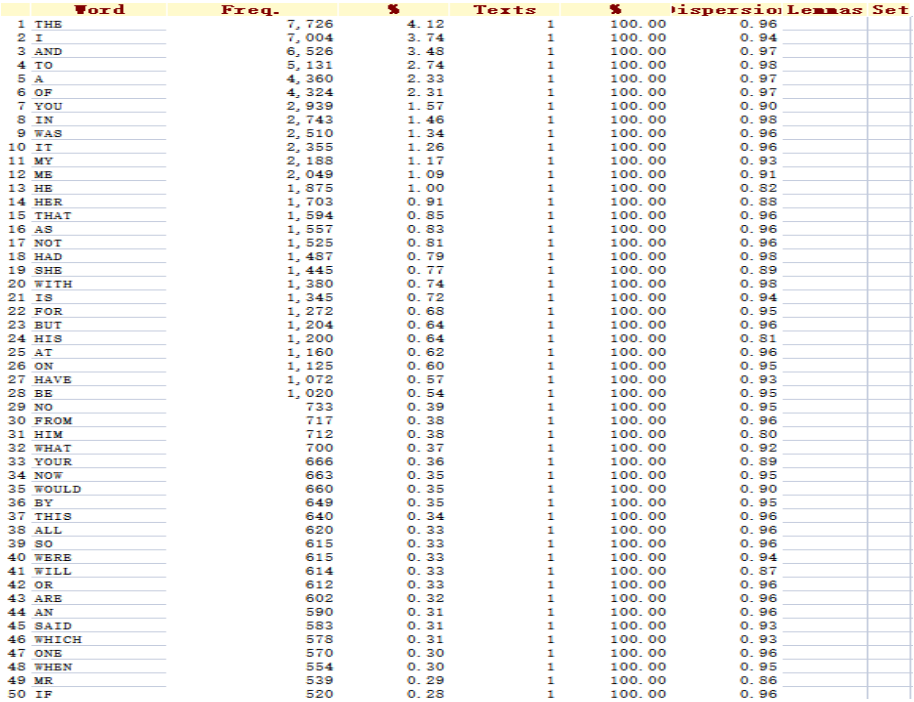
frequency clusters or phrases. Starcke[7] retrieved frequently occurring phrases and their collocations in Persuasion and focused on the collocations of “she could not” and “she had been”, providing a profound analysis of character psychology. Starcke [8] analyzed thematic words and high-frequency clusters in Pride and Prejudice, demonstrating the significant importance of corpus stylistics in the study of literary works. Mahlberg [5], by comparing Pride and Prejudice with works by 19 other writers from the same century in a reference corpus, analyzed thematic words and clusters in Pride and Prejudice. It was discovered that body language plays a certain role in driving the plot.

Compared to the extensive research by foreign scholars on high-frequency word clusters or phrases in literary works, domestic corpus stylistics research mainly focuses on word frequency statistics and thematic word analysis. For instance, Chen Chan and Cheng Le[2] extracted thematic words from Mo Yan’ s novel Frog and conducted a relatively in-depth analysis from aspects such as theme, storyline, and character portrayal. Research on collocations of words, like Wang Qin[10], used corpus retrieval software to conduct word statistics on The Sun Also Rises. The results showed that the frequency and distribution of word occurrences, collocations, and chunks can provide more objective support for previous literary interpretations. Analysis of clusters or frequently occurring phrases in the text is conducted to reveal the text’ s stylistic features. For example, Chen Chan[1] analyzed the characteristics and functions of three to six-word clusters in Alice Munro’ s novels. A comparative study was conducted with clusters in English literary works from the same period. It was pointed out that the use of these clusters reflects the confusion and entanglement of female characters in Munro’ s novels.

Jane Eyre, as a representative work of the famous 19th-century British female writer Charlotte Brontë, is deeply loved by readers worldwide. However, current academic research on Jane Eyre mostly consists of qualitative analysis. This paper attempts to apply corpus research to analyze text word statistics and retrieve high-frequency words. Based on this, it further analyzes the detailed relationships between the novel’ s thematic words and plot, characters, environment, and psychological details.

2 An Analysis of High-Frequency Words in *Jane Eyre*

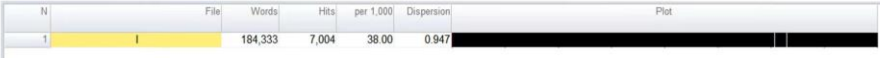
The analysis of high-frequency words is a mainstream method for studying the linguistic features of novels using corpus research. This paper begins by retrieving the high-frequency words in the novel to summarize the overall linguistic characteristics. Subsequently, it further examines high-frequency verbs, nouns, and adjectives to conduct a more in-depth analysis of the novel’ s linguistic features. By utilizing the Word list function in the software Wordsmith Tools 7.0, the author first extracts the frequency table of Jane Eyre, as shown in Table 1.



	Word	Freq.	%	Texts	%	Dispersio	Lemmas	Set
1	THE	7,726	4.12	1	100.00	0.96		
2	I	7,004	3.74	1	100.00	0.94		
3	AND	6,526	3.48	1	100.00	0.97		
4	TO	5,131	2.74	1	100.00	0.98		
5	A	4,360	2.33	1	100.00	0.97		
6	OF	4,324	2.31	1	100.00	0.97		
7	YOU	2,939	1.57	1	100.00	0.90		
8	IN	2,743	1.46	1	100.00	0.98		
9	WAS	2,510	1.34	1	100.00	0.96		
10	IT	2,355	1.26	1	100.00	0.96		
11	MY	2,188	1.17	1	100.00	0.93		
12	ME	2,049	1.09	1	100.00	0.91		
13	HE	1,875	1.00	1	100.00	0.82		
14	HER	1,703	0.91	1	100.00	0.88		
15	THAT	1,594	0.85	1	100.00	0.96		
16	AS	1,557	0.83	1	100.00	0.96		
17	NOT	1,525	0.81	1	100.00	0.96		
18	HAD	1,487	0.79	1	100.00	0.98		
19	SHE	1,445	0.77	1	100.00	0.89		
20	WITH	1,380	0.74	1	100.00	0.98		
21	IS	1,345	0.72	1	100.00	0.94		
22	FOR	1,272	0.68	1	100.00	0.95		
23	BUT	1,204	0.64	1	100.00	0.96		
24	HIS	1,200	0.64	1	100.00	0.81		
25	AT	1,160	0.62	1	100.00	0.96		
26	ON	1,125	0.60	1	100.00	0.95		
27	HAVE	1,072	0.57	1	100.00	0.93		
28	BE	1,020	0.54	1	100.00	0.95		
29	NO	733	0.39	1	100.00	0.95		
30	FROM	717	0.38	1	100.00	0.96		
31	HIM	712	0.38	1	100.00	0.80		
32	WHAT	700	0.37	1	100.00	0.92		
33	YOUR	666	0.36	1	100.00	0.89		
34	NOW	663	0.35	1	100.00	0.95		
35	WOULD	660	0.35	1	100.00	0.90		
36	BY	649	0.35	1	100.00	0.95		
37	THIS	640	0.34	1	100.00	0.96		
38	ALL	620	0.33	1	100.00	0.96		
39	SO	615	0.33	1	100.00	0.96		
40	WERE	615	0.33	1	100.00	0.94		
41	WILL	614	0.33	1	100.00	0.87		
42	OR	612	0.33	1	100.00	0.96		
43	ARE	602	0.32	1	100.00	0.96		
44	AN	590	0.31	1	100.00	0.96		
45	SAID	583	0.31	1	100.00	0.93		
46	WHICH	578	0.31	1	100.00	0.93		
47	ONE	570	0.30	1	100.00	0.96		
48	WHEN	554	0.30	1	100.00	0.95		
49	MR	539	0.29	1	100.00	0.86		
50	IF	520	0.28	1	100.00	0.96		

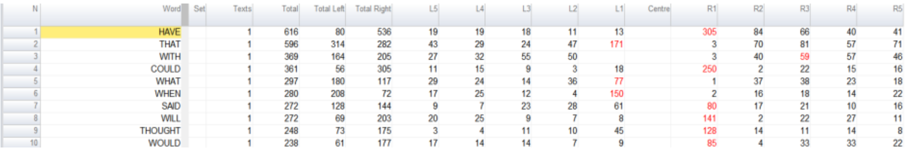
Fig. 1. Top 50 Most Frequently Occurring Words in *Jane Eyre*

By using Wordsmith Tools 7.0 software to analyze Jane Eyre, it was found that the novel contains a total of 187,515 words. From Table 1, it can be observed that pronouns occur most frequently, especially personal pronouns and possessive pronouns, totaling 11. Among them, the first-person singular pronoun “I” has the highest frequency, reaching 7,004 times, while “My” and “Me” rank 11th and 12th respectively. Based on Table 1, it can be assumed that the novel is narrated in the first person narrator. This assumption was validated in the Concordance Plot of Wordsmith Tools 7.0, as shown in Figure 1, where the black portion represents the distribution of the first-person singular pronoun “I” throughout the entire text. “I” is densely distributed in the novel, consistently present. In the first-person narrative, the narrator belongs to a non-omniscient perspective, which facilitates reader engagement and allows them to stand on a similar standpoint as the narrator, making it easier for the narrator to express emotions and conduct psychological descriptions directly. Jane Eyre’ s use of the first person also enhances the authenticity of the novel, enabling readers to empathize with the protagonist, showcasing Charlotte Brontë’ s exceptional writing skills. In addition, the second most frequently occurring personal pronoun is “You,” appearing 2,939 times. The author examined the top ten collocations of the personal pronoun “I” (as shown in Figure 2) and found that among the top ten collocations, there is only one content verb, “Said.” Additionally, “Said” is the only lexical verb in the high-frequency words from Table 1. This indicates that there are many dialogues in the text.



N	File	Words	Hits	per 1,000	Dispersion	Plot
1	I	184,333	7,004	38.00	0.947	

Fig. 2. the co-occurrence context of the personal pronoun “I”

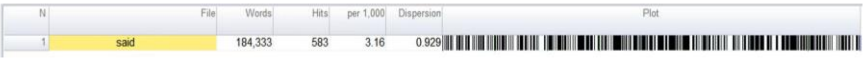


N	Word	Set	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	HAVE		1	616	80	536	19	19	18	11	13		305	84	66	40	41
2	THAT		1	596	314	282	43	29	24	47	171		3	70	81	57	71
3	WITH		1	369	164	205	27	32	55	50			3	40	59	57	46
4	COULD		1	361	56	305	11	15	9	3	18		250	2	22	15	16
5	WHAT		1	297	180	117	29	24	14	36	77		1	37	38	23	18
6	WHEN		1	280	208	72	17	25	12	4	150		2	16	18	14	22
7	SAID		1	272	128	144	9	7	23	28	61		80	17	21	10	16
8	WILL		1	272	69	203	20	25	9	7	8		141	2	22	27	11
9	THOUGHT		1	248	73	175	3	4	11	10	45		128	14	11	14	8
10	WOULD		1	238	61	177	17	14	14	7	9		85	4	33	33	22

Fig. 3. Top Ten Collocates of the First-Person Pronoun “I”

Additionally, it was observed that the total occurrences of “He,” “Him,” and “His” (3783 times) far exceeded the occurrences of “She” and “Her” (3148 times). Furthermore, among the top 50 frequent words, “Mr” appeared 539 times. This indicates that, apart from the first-person pronoun “I,” male characters also hold significant positions in the novel. Moreover, considering that the usage of these pronouns is correlated with the content of the novel, when combined with the fact that the first two names in the frequency list, “Jane,” appeared 330 times (ranking 69th), and “Rochester” appeared 317 times (ranking 70th), it is highly likely that Jane and Rochester are the main characters of the novel.

On the other hand, among the top 50 words listed in Table 1, there are a total of 10 verbs, with 5 of them being in the past tense (Was, Had, Would, Were, Said). The most frequently occurring verb is “Was,” appearing 2510 times. The base form “is” appears 1345 times, ranking 21st; “Were” appears 615 times, ranking 40th, while “are” appears 602 times, ranking 43rd. “Had” occurs 1487 times, ranking 18th, and “Have” occurs 1072 times, ranking 27th. The coexistence of different tenses of the same word in the text indicates temporal shifts in the novel's descriptions. Moreover, the data suggests that the past tense is used more frequently than the present tense. “Said,” being the only action verb in the top 50 high-frequency words list, is also in the past tense. Utilizing the Concordance Plot in Wordsmith Tools 7.0, a search with “Said” as the keyword presents the distribution of “Said” in the text, as shown in Figure 3. The figure illustrates a dense occurrence of “Said” in the text, indicating the prevalent use of the past tense for describing plot development and affirming the earlier conclusion that Jane Eyre contains numerous dialogues.



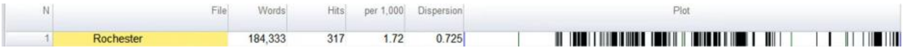
N	File	Words	Hits	per 1,000	Dispersion	Plot
1	said	184,333	583	3.16	0.929	

Fig. 4. the co-occurrence context of “Said” in *Jane Eyre*

The demonstrative pronoun “That” (1594 times), ranks 15th; “What” (700 times), ranks 32nd; “Which” (578 times), ranks 46th, and so on. Their high frequencies within the top 50 words (Table 1) indicate the presence of numerous subordinate clauses in the novel. The sentence structure is complex and leans towards formality, aligning with the written characteristic of the novel.

3 Plot Prediction in *Jane Eyre*

As mentioned earlier, the conclusion that Jane and Rochester are likely the main characters of the novel is validated in the Concordance Plot. This is shown in Figures 4 and 5.



N	File	Words	Hits	per 1,000	Dispersion	Plot
1	Rochester	184,333	317	1.72	0.725	

Fig. 5. the co-occurrence context of “Rochester” in *Jane Eyre*

As mentioned earlier, the novel “Jane Eyre” is narrated in the first person, as shown in Figure 1, where the personal pronoun “I” permeates the entire text. Therefore, in Figure 4, the character “Jane” appears less frequently compared to Figure 1, but it can still be observed that the character “Jane” is distributed throughout the text. Additionally, upon entering “Rochester” in the concordance plot, Figure 5 appears. It is evident that “Rochester” does not appear in the early part of the entire novel. In comparison with Figure 4, which illustrates the co-occurrence context of “Jane”, the absence of “Rochester” in the early part of the text is even more pronounced. From Figure 5, it is evident that Rochester appears in the middle part of the novel. This indicates that Jane and Rochester likely meet in the middle section of the novel. Based on this analysis, it can be inferred that the early part of the novel describes Jane’ s childhood and adolescence at Gateshead Hall and Lowood School. In Figure 5, it can be observed that Rochester’ s presence becomes less dense in the latter half of the novel, with significant blanks and fewer occurrences. This suggests that Rochester and Jane may have been separated for some time due to certain factors. During this period of separation, there might be slight traces of Rochester. However, in the final part, Rochester appears frequently again, indicating that Rochester and Jane ultimately reunite.

4 Conclusion

The introduction of corpus methods into literary research signifies a shift from qualitative to a combined approach of both qualitative and quantitative analysis. The integration of quantitative analysis with qualitative research methods can effectively propel the development of literary studies[4]. This study employs the corpus research method, using Charlotte Brontë’ s representative work Jane Eyre as the research subject. The research findings reveal that Jane Eyre is a novel narrated in the first person, and it contains a significant amount of dialogue. The novel predominantly employs the past tense, indicating that it largely consists of reminiscence and narration of past experiences. Additionally, the data derived from corpus analysis tools also aids readers in anticipating the plot development of the novel.

Reference

1. Chen Chan. (2014) Cluster Characteristics and Functional Analysis in Alice Munro’ s Novels: A Corpus-based Stylistic Study. Journal of PLA University of Foreign Languages, 37(03): 151-159.
2. Chen Chan, Cheng Le. (2014) A Corpus-based Analysis of Mo Yan’ s Novel Frog. Journal of Zhejiang Gongshang University, 05: 26-34.
3. Culpeper, J.K. (2009) Words, parts-of-speech and semantic categories in the character-talk of Shakespeare’ s Romeo and Juliet. International Journal of Corpus Linguistics, 14(1): 29-59.
4. Hu Kaibao, Yang Feng. (2019) Corpus-based Literary Studies:Connotations and Implications. Journal of Zhejiang University(Humanities and Social Sciences), 49(05):143-156.
5. Mahlberg M. (2013) Corpus stylistics and Dickens’ s fiction. Routledge.
6. QIN Hongwu. (2021) Digital Humanities for Literary Studies:Theory and Analytical Methods. Foreign Languages in China, 18(03): 98-105.
7. Starcke B. (2006) The phraseology of Jane Austen’ s Persuasion: Phraseological units as carriers of meaning. ICAME journal, 30(87): 104.
8. Starcke B. (2009) Keywords and frequent phrases of Jane Austen’ s Pride and Prejudice: A corpus-stylistic analysis. International Journal of Corpus Linguistics, 14(4): 492-523.
9. Stubbs M. (2005) Conrad in the computer: examples of quantitative stylistic methods[J]. Language and literature, 14(1): 5-24.
10. Wang Qin. (2012) Words,Collocations and Lexical Chunks: On Corpus Stylistics of The Sun Also Rises. Journal of Huaihua University, 31(06): 81-84.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.