A Corpus-based Stylistic Study of Jane Eyre and Wide Sargasso Sea

Mengying Liang, Man Chen

School of Law and Humanities, China University of Mining and Technology (Beijing), Beijing, 100083, China

Abstract: Corpus stylistics is an emerging research field that combines corpus linguistics research with the stylistics. Based on the research method of corpus stylistics, this article builds the corpora of *Jane Eyre* and *Wide Sargasso Sea* and uses F-LOB as the reference corpus. The stylistic analysis of the two works is conducted with the YiCorpus multifunctional corpus retrieval platform and Antconc 3.2.1 software for corpus text retrieval. By comparing the type/token ratio, mean word length, mean sentence length, and keywords, this study explores the different stylistic features of *Jane Eyre* and *Wide Sargasso Sea* in terms of lexical richness, vocabulary and sentence complexity, and keyword lists, providing new paths for appreciating literary works.

1. Introduction

Corpus stylistics is a research method that studies themes, characters, narrative structures, and authorial styles of literary works through techniques such as word frequency analysis, keyword retrieval, and concordance [1]. It combines qualitative, interpretive and experience-based stylistic analysis with quantitative, descriptive, and probability-based corpus statistics using corpora and retrieval software [2]. Corpus stylistics provides a new perspective and a scientific method for the analysis and appreciation of literary works.

The research object of this article is the classic work *Jane Eyre* and its prequel *Wide Sargasso Sea*. *Jane Eyre*, written by British female author Charlotte Bronte, pioneered women literature, with the image of the protagonist Jane daring to fight for freedom and equality, and her love story with Rochester, breaking through prejudice and respecting each other. *Wide Sargasso Sea* was a rewriting of *Jane Eyre* by the Dominican-British author Jean Rhys. Jane Rhys conducted in-depth research on Bertha Antoinette Mason, the mad woman in the original work, and based on her rich experience of the West Indian colonial culture, wrote this companion volume about the ex-wife's experience. *Wide Sargasso Sea* achieved an unexpected success after its publication in 1966, becoming a masterpiece in the post-colonial texts.

Many studies, compared *Jane Eyre* and *Wide Sargasso Sea*, are based on text only from the perspectives of post-colonialism, new historicism, feminism, discourse analysis, etc. There are few researches on stylistic analysis using corpus methods, so the topic of this paper has a positive significance with a large space for research. This article adopts the method of corpus data analysis, self-build the corpus of *Jane Eyre* and *Wide Sargasso Sea*, and selected F-LOB as a reference corpus.

This study uses the Yi Corpus multifunctional corpus retrieval platform and the Antconc software to reveal the similarities and differences in stylistic and thematic expression between the two works.

2. Research methods

2.1 Corpus information

As for self-built corpus, first of all, the electronic versions of *Jane Eyre* and *Wide Sargasso Sea* are required by downloading them online. Then, the two novels are converted form PDF format into Word format. The body part is copied into a plain text document and saved in TXT format. Finally, redundant characters, formulas and charts are eliminated in case of affecting corpus retrieval.

To ensure the objectivity of the study and the representativeness of the keyword lists, this paper uses a reference corpus F-LOB (the Freiburg update of the LOB corpus) which is originated from Brown Corpus. Like the original Brown and LOB corpora, F-LOB contains 500 texts of around 2000 words each, distributed across 15 text categories. The texts were selected carefully to match the LOB corpus as closely as possible, representing the language of the early 1990s. The main aim is to achieve a close comparability with LOB, in order to provide linguists with an empirical basis to study language change in progress.

2.2 Software

The tools used in corpus construction and data analysis are as follows:

YiCorpus multifunctional corpus retrieval platform is a professional corpus retrieval and resource management platform that supports the import and retrieval of multilingual and multi-format files for single-language corpora, multilingual parallel corpora, and terminology databases. It also supports independent storage and management of personal and public language resources. The "Word Frequency" function can automatically generate data of novels, such as token, type, Type-Token Ratio (TTR), and Standardized Type-Token Ratio (STTR).

Antconc 3.2.1 software is a corpus analysis toolkit developed by Laurence Anthony for concordancing and text analysis. It hosts a comprehensive set of tools including a powerful concordancer, word and keyword frequency generators, tools for cluster and lexical bundle analysis, and a word distribution plot. Consequently, AntConc can be applied for a comparative study to explore how words and phrases are distributed as well as their frequency and keyness in any given textual dataset.

3. Results and Discussion

3.1 Type-token Ratio

Corpus-based stylistic research typically analyzes the formal features of texts, and studies vocabulary to help understand specific styles of the text. Generally, studying lexical richness involves analyzing the type/token ratio of a text. "Token" refers to all the words that appear in the text. "Type" is the number of unique symbols not counted repeatedly, i.e. different words in the text. Type/token ratio refers to the ratio between the number of different words used in a corpus and the total number of running words, and is one of the important indicators for studying stylistic features. It can be used to visually examine the richness of the vocabulary [3]. However, since the type-token ratio is easily affected by length of the text, the standardized type-token ratio is usually used as the measurement standard, and the ratio is directly proportional to the richness and diversity of the vocabulary [4].

This article uses the YiCorpus multifunctional corpus retrieval platform to compare the token, type,

type/token ratio, the standardized type-token ratio of the two texts, as shown in table 1. From table 1, it can be seen that the standardized type-token ratio of *Jane Eyre* is 0.50, while the standardized type-token ratio of *Wide Sargasso Sea* is 0.43, which is lower than that of *Jane Eyre*. The difference in data indicates that *Jane Eyre* has a higher degree of vocabulary variation and uses more words than *Wide Sargasso Sea* under the same numbers of words. In addition, it should be noted that although the difference between the standardized type-token ratio of the two books is only 0.07, which seems small in numerical terms, the overall number of tokens in *Jane Eyre* is about three times that of *Wide Sargasso Sea*, so the actual difference in vocabulary diversity between the two books is very large.

Table 1: The type-token ratio and the standardized type-token ratio of the two works

Measurement indicators	*Jane Eyre*	*Wide Sargasso Sea*
Type	16546	4785
Token	139007	47645
TTR	0.12	0.10
STTR	0.50	0.43

3.2 Vocabulary and Sentence Difficulty

Analysis of vocabulary and sentence difficulty usually involves considering the mean word length and mean sentence length of the text. "Mean word length" refers to the ratio of the total number of letters to the total number of tokens in a specific text, and is an important indicator reflecting the formality and difficulty of the text. "Mean sentence length" can be calculated by counting the number of punctuation marks and total words in the text.

Through the analysis of the Yi Corpus, the data of mean word length and mean sentence length of the two works can be obtained, and the results are shown in table 2. This data shows that the mean word length of *Jane Eyre* is 7.34 and the mean sentence length is 13.90; the mean word length of *Wide Sargasso Sea* is 6.21 and the mean sentence length is 11.03. Both sets of indicators for *Jane Eyre* are significantly higher than those for *Wide Sargasso Sea*, indicating that *Jane Eyre* is a book with longer and more difficult vocabulary, more medium and long sentences, more complex and varied sentence forms, and more formal and written in terms of style. The wording used in *Wide Sargasso Sea* is relatively simple, with shorter and easier sentence structures, so the text is more colloquial and highly readable, presenting a concise and plain writing style.

Table 2: The mean word length and mean sentence length of the two works

Measurement indicators	*Jane Eyre*	*Wide Sargasso Sea*
Mean word length	7.34	6.21
Mean sentence length	13.90	11.03

3.3 Keywords List

In corpus stylistic research, keyword retrieval is an important way to identify the author's creative intention and style. By retrieving words with high keyness or high frequency, the story plot of the text can be determined and connected. The greater the keyness, the more important the word is in the text, and the greater its role in driving the development of the story plot. This study analyzes two works using the "Keyword List" function of Antconc software with the F-LOB corpus as the reference corpus. The top 25 keywords of the two works are sorted by keyness and are shown in table 3.

Based on the lexical filtering and sorting, both keyword lists contains various parts of speech, such as personal pronouns, nouns, verbs, adjectives, articles, conjunctions, and adverbs. This article will

conduct a combined quantitative and qualitative analysis from three aspects: personal pronouns, nouns, and verbs.

Table 3: The keyword list of the two works

Keyword list of *Jane Eyre*				Keyword list of *Wide Sargasso Sea*			
Rank	**Keyness**	**Freq**	**Keyword**	**Rank**	**Keyness**	**Freq**	**Keyword**
1	11196.644	7226	I	1	4795.015	2206	I
2	6976.271	1753	a	2	1384.700	538	me
3	4069.960	2045	me	3	1105.358	706	you
4	4047.993	2080	my	4	1018.453	644	she
5	3297.110	2649	you	5	810.349	398	my
6	1432.280	366	Rochester	6	805.777	652	her
7	1008.587	1683	her	7	725.339	437	said
8	929.903	340	Jane	8	669.468	103	Christophine
9	892.166	292	sir	9	606.338	334	She
10	862.228	643	your	10	496.329	78	Antoinette
11	838.166	305	Miss	11	375.434	362	t
12	801.968	6314	and	12	342.300	160	don
13	677.411	1235	she	13	293.598	51	Mason
14	668.988	370	am	14	281.916	171	You
15	541.228	136	Fairfax	15	268.372	784	was
16	536.953	542	Mr	16	253.488	39	Cora
17	531.676	1483	not	17	252.431	67	dress
18	525.725	135	Adele	18	245.818	212	like
19	515.761	2072	it	19	234.283	93	white
20	473.817	126	Reed	20	224.873	36	Baptiste
21	473.134	598	now	21	221.551	77	girl
22	462.471	131	Bessie	22	220.980	40	Aunt
23	416.667	1477	had	23	220.367	121	away
24	397.962	100	Thornfield	24	218.052	94	tell
25	389.453	2509	was	25	217.829	149	know

(1) From the perspective of personal pronouns

Jane Eyre and *Wide Sargasso Sea* both use first-person pronouns such as "I", "me", and "my" most frequently in their narrations. The keyness and frequency of second-person pronouns like "you" and third-person pronouns like "she" are much lower than that of first-person pronouns.

This shows that the two works are very consistent in adopting a first-person perspective and presenting readers with an immersive reading experience in the form of autobiography. Consequently, readers can directly contact the internal thoughts and feelings of the narrators, adding authenticity and credibility to the two stories.

(2) From the perspective of nouns

The nouns in the top 25 keyword list in *Jane Eyre* are, in order of keyness, "Rochester", "Jane", "Fairfax", "Adele", "Reed", "Bessie", and "Thornfield". Among them, the keyness of "Rochester" and "Jane" is much higher than those for other names, manifesting that the main characters of the story are "Rochester" and "Jane". Other nouns such as "Fairfax" (the housekeeper of Thornfield), "Adele" (Rochester's adopted daughter), "Reed" (the surname of Jane's uncle's family), "Bessie" (the maid of Jane's uncle's family), and Thornfield (Rochester's estate) also rank high on the list of themes. Their appearances drive the development of the plot and allow readers to understand Jane's difficult experience as an orphan and her series of self-reliance and touching stories. Although the

data on the list shows that the keyness and frequency of "Rochester" are higher, this does not mean that there is more plots about Rochester than Jane, or that Rochester is the core of the characters. Because the entire work is told from Jane's perspective in the first person narration, and Rochester only appears on page 165 (one-fifth of the book). Therefore, the plot of *Jane Eyre* revolves mainly around the governess Jane, just as its title suggests.

The nouns in the top 25 keyword list in *Wide Sargasso Sea* are sorted by keyness as follows: "Christophine", "Antoinette", "Mason", "Cora", "Baptiste", "girl", "Aunt". These characters are Antoinette's father's second wife (who has taken care of Antoinette), Antoinette herself, Antoinette's stepfather, Antoinette's aunt (who is also "Aunt"), the housekeeper of Antoinette's family, and a local girl Antoinette encounters. These clearly show that the central character of *Wide Sargasso Sea* is Antoinette. But the original male protagonist, Antoinette's husband Rochester, is reduced to a minor cameo in *Wide Sargasso Sea*, and his plot is not even as much as that of Antoinette's housekeeper. This arrangement of characters reflects the author's complete subversion of the original work. By bringing the marginal character back into the spotlight and describing in detail Antoinette's entire life from childhood to marriage and death, Jean Rhys put Antoinette and her family at the centre of the stage.

In addition to names, there is also a noun "dress" that has high keyness in the keyword list of *Wide Sargasso Sea*. Combining with the "Concordance plot" function of Antconc software, it can be seen that "dress" appears respectively in the first, middle and last three parts of the book (as shown in figure 1).

Figure 1: The concordance plot of "dress"

"Dress" first appeared frequently in the novel referring to Antoinette's skirt, which was secretly stolen by her only local friend Tia. After the collapse of slavery in the West Indies in the 19th century, Antoinette, as a mixed-race Creole and the daughter of a former slave owner, often faced hostility from the locals and disdain from white aristocrats. Although she was used to enjoying herself in the shabby and difficult environment, Tia's act of stealing her skirt made her feel deeply betrayed and despised, and also made her understand that even sincere friendship could not eliminate the everlasting hatred between races. The word "dress" was intensively mentioned for the second time in the book, referring to the skirt of Christophine. When the hem of Christophine's dress dragged on the ground making a rustling sound as she walks, Rochester thought it was an unclean habit, and extended this to Christophine's lack of etiquette and laziness. Rochester and Antoinette's conflicting views on the hem of the dress illustrated their cognitive differences, indicating that their marriage is hasty and unsuitable. The third frequent appearance of "dress" came at the end of the novel. When Antoinette was delirious as a result of being imprisoned in the attic by Rochester, the thing what she repeatedly referred to was a red dress. The dress was as red as "the colour of fire and sunset" [5]. Rochester once commented that Antoinette "looked intemperate and unchaste" in it[5], but Antoinette thought that wearing the red dress would make her brother Richard recognize her and show her true self. After being found, this dress exuded a strong scent of "vetivert and frangipanni, of cinnamon and dust and lime trees" and "the smell of the sun and the smell of the rain" which were unique to the West Indies [5]. The home that used to be in the West Indies was a place that Antoinette loved and hated, where Antoinette was insulted and bullied by the locals and whites, and eventually the whole house was deliberately set on fire by the locals. But there also made Antoinette feel the beauty of the natural scenery and the love and care from her mother and Christopher. West Indies became a ray of hope when she was in mental breakdown. The skirt was as fiery red as flames spreading in the room, and drove Antoinette to ignite the house that once imprisoned her in the confusion between dream and

reality.

(3) From the perspective of verbs

The verbs in both keyword lists are in the past tense except for "am", indicating that both works are recording experiences that have already happened, which is consistent with the autobiographical style of the novel.

The verbs in the keyword list of *Jane Eyre* are "am", "had" and "was". Based on the "Concordance" function of Antconc, it can be seen that "am" appears in both "I am" sentence structure and the direct speech dialogues (as shown in figure 2). This indicates that Jane is in a relatively strong and dominant position in conversations with others, and has a strong desire to make her own voice, corresponding to her personality of self-reliance, fearlessness of authority and willingness to fight.

364	¡°And there is enchantment in the very hour I am now spending ı	Jane Eyre.txt
365	e at hand: you talk of my being a fairy, but I am sure, you are	Jane Eyre.txt
366	ut I am sure, you are more like a brownie.¡± ¡°Am I hideous, Jaı	Jane Eyre.txt
367	velling these last three days, and I believe I am tired. Good nː	Jane Eyre.txt
368	concentrated in my Jane¡¯s tongue to my ear (I am glad it is noı	Jane Eyre.txt
369	e Vulcan, sir.¡± ¡°Well, you can leave me, ma¡¯am: but before yı	Jane Eyre.txt
370	Eyre, I repeat it, you can leave me. How often am I to say the ː	Jane Eyre.txt
371	I have given you notice to quit?¡± ¡°Because I am comfortable tł	Jane Eyre.txt
372	I forget that you have formed a new tie. But I am not a fool ¡ª	Jane Eyre.txt
373	as an iceberg. He is not like you, sir: I am not happy at	Jane Eyre.txt
374	le down the manly cheek. My heart swelled. ¡°I am no better than	Jane Eyre.txt
375	yer ¡ª if ever I wished a righteous wish, ¡ª I am rewarded now.	Jane Eyre.txt
376	t I know whose voice it was ¡ª replied, ¡®I am coming: wait	Jane Eyre.txt
377	requent, correspondence ever since: he hopes I am happy, and trı	Jane Eyre.txt
378	ever since: he hopes I am happy, and trusts I am not of those ı	Jane Eyre.txt
379	st beyond what language can express; because I am my husband¡¯s	Jane Eyre.txt
380	e. No woman was ever nearer to her mate than I am: ever more abː	Jane Eyre.txt
381	so very close: for I was then his vision, as I am still his rigł	Jane Eyre.txt

Figure 2: The concordance of "am"

According to the "Concordance Plot" function of Antconc, it can be seen that the density of "had" is very high (as shown in figure 3). "Had" is used throughout the novel to indicate past tense, and also shows Jane's subjective initiative, for example, "I had the greatest inclination to follow up my advantage to purpose", "To speak truth, I had not the least wish to go into company", "The fact was, I had other things to think about"[5]. After the death of her parents, Jane was sent to live with her uncle and aunt enduring 10 years of discrimination and abuse. Later, she suffered both mentally and physically at Lowood School, where poor conditions led to the deaths of many students. After leaving Rochester, she was homeless and begged for food and shelter. Therefore, readers can deeply understand Jane's strong mind, bravery, as well as control over her own fate and soul.

HIT FILE: 1   FILE: Jane Eyre.txt

No. of Hits = 1490
File Length (in chars) = 1037021

Figure 3: The concordance plot of "had"

The verbs in the keyword list of *Wide Sargasso Sea* are "said", "was", "like", "tell" and "know" in order of keyness. The book is divided into three parts, and although they are all narrated in the first person, the subjects of the three narratives are Antoinette, Rochester and Antoinette respectively. "Said" and "tell" indicate that the plot is not only the subjective narrative of the two narrators Antoinette and Rochester, but also includes many dialogues and persuasions from other characters around Antoinette. The abundance of third party discourse adds other voices to the text of first perspective, making the narration about the childhood and marriage of the insane Antoinette more objective and authentic. The word "was" mainly reflects characters' identity, state, and passive

behavior in the novel. "Was" drives the story forward while also indicates that many of Antoinette and Rochester's actions deviate their real thoughts with helplessness, especially as two complete strangers are forced to become a couple. Although Antoinette believed that marriage could help her escape the marginalized life and fulfill her dream of living in England, and although the second son Rochester, who had no inheritance rights, obtained a large sum of money from Antoinette's stepfather through marriage, they still showed confusion, anxiety, and even fear in the hasty marriage. After they became a couple, they also experienced a lot of struggles and pains in the face of the inappropriateness of the ideology and the provocation of the villains. Finally, under the incitement of racial prejudice and rumors, Rochester imprisoned Antoinette in the attic despite her explanation and entreaties. "Like" and "know" express the feelings and attitudes of the narrative subjects, and according to the function of "Concordance Plot" (as shown in figure 4 and figure 5), these two verbs are evenly distributed throughout the book. This indicates that both Antoinette and Rochester had explored the preferences and thoughts of themselves and those around them, had paid attention to the emotional changes in their hearts, and had been filled with deep humanistic care. But later they still violated their own will and forced themselves to marry and live with each other. In the final part of the book, Antoinette finally understood Rochester's cruelty and madness, realized the fact that she was locked in the attic with no escape, and knew that Rochester would have a new marriage and a new wife. Therefore, under the hint of dreams, she knocked down all candles and gazed at the the flames spreading throughout the house.

Figure 4: The concordance plot of "like"

Figure 5: The concordance plot of "know"

4. Conclusion

This study uses YiCorpus multifunctional corpus retrieval platform and Antconc 3.2.1 software for corpus text retrieval to analyze the stylistic characteristics of *Jane Eyre* and *Wide Sargasso Sea* in terms of lexical richness, vocabulary and sentence difficulty and keyword lists. The results show that *Jane Eyre* is higher in lexical diversity, lexical difficulty and syntactic difficulty. Although both works are popular fictions, the style of *Jane Eyre* is more formal and written, and *Wide Sargasso Sea* is more colloquial.

From the perspective of the keywords, both novels adopt the first-person autobiographical narrative perspective. The narrative subject of *Jane Eyre* is Jane herself, and the whole book revolves around her life and experiences, depicting a female image with a distinct consciousness, independence and proactiveness. The narrative subjects of *Wide Sargasso Sea* are Antoinette and Rochester. It is worth noting that Antoinette, who was once a "hyena", "monster" and "maniac" in Jane's eyes and Rochester's mouth, has become a sympathetic and pitiful tragic character under the pen of Jean Rhys. Rhys' emphasis on Antoinette and her belittling and weakening of the original male character Rochester reflects not only the huge changes in character shaping and plot focus of *Wide Sargasso Sea*, but also the author's desire to make a new voice and reshape history through the lenses of new historicism and post-colonialism.

To sum up, YiCorpus and Antconc corpus software enable the characteristics of the literary styles, narrative perspectives, plot development, and character portrayal of *Jane Eyre* and *Wide Sargasso*

Sea to be revealed, providing data verification for traditional qualitative literary research and compensating for the subjectivity and abstraction of pure text research. Corpus Stylistics provides a scientific research perspective for literary research, especially for comparative literature research in both horizontal and vertical dimensions.

References

[1] Weizhong Lu, Yun Xia. Corpus Stylistics: A New Approach to Literary Stylistics [J]. Journal of Foreign Languages, 2010, 33(01): 47-53.
[2] Huizhong Yang. An Introduction to Corpus Linguistics [M]. Shanghai: Shanghai foreign language education press, 2002:242-243.
[3] Baker M. Towards a Methodology for Investigating the Style of a Literary Translator [J]. Target, 2000(2):241-166.
[4] Scott M. & Tribble C. Textual Patterns: Keyword and Corpus Analysis in Language Education [M]. Amsterdam & Philadelphia: John Benjamins Publishing Co., 2006.
[5] Rhys J. Wide Sargasso Sea [M]. New York: W. W. Norton& Company, 1991.