A Corpus-based Stylistic Analysis of *The Great Gatsby*

Qian Ju[*]

Nanjing Normal University, Nanjing, China
[*]Corresponding author: vivianju12@163.com

Abstract: *The Great Gatsby is a timeless exploration of the illusory nature of the American Dream, decadence in the 1920s, and the complexities of love and desire. Fitzgerald's ability to vividly capture the spirit of an era while addressing universal themes makes The Great Gatsby a literary masterpiece deserving of its enduring acclaim and ongoing literary analysis. Based on theories of corpus stylistics, this paper uses corpus search software AntConc, TagAnt and UAM CorpusTool to analyze the stylistic features of this novel in four aspects: lexical, phrase, syntax and rhetorical features. This paper concludes that F. Scott. Fitzgerald excels in using past tense words, phrase collocations, syntactic deviations and rhetorical devices to portray various negative characters and reveal the wealthy but indifferent and hypocritical society of 1920s America.*

Keywords: *Corpus stylistics; The Great Gatsby; AntConc; TagAnt; UAM CorpusTool*
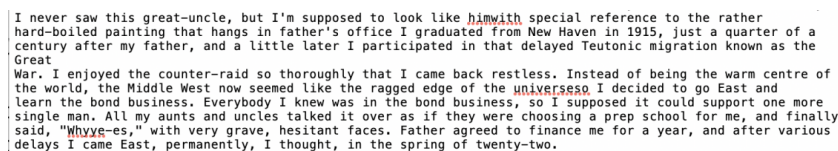
1. Introduction

The Great Gatsby is a novella written by American novelist F. Scott. Fitzgerald in 1925. [1] It captures the reality of American social life in the 1920s and describes the hypocrisy and destruction of the so-called "American Dream". The novel has been a great success in the literary world and has been analyzed by many scholars at home and abroad from different perspectives, using a variety of theories to analyze its literary and aesthetic value, most of which are too subjective due to the lack of objective examples. However, quantitative research based on corpus can improve this deficiency and become a breakthrough in methodology.

Corpus stylistics aims to quantitatively describe and qualitatively analyze the stylistic features of various texts with the help of corpus tools. This paper will use AntConc, TagAnt and the UAM CorpusTool to extract the necessary data, analyze the stylistic features of the novel in terms of vocabulary, phrase, syntax and rhetoric, and provide quantitative data for the interpretation and appreciation of the works.
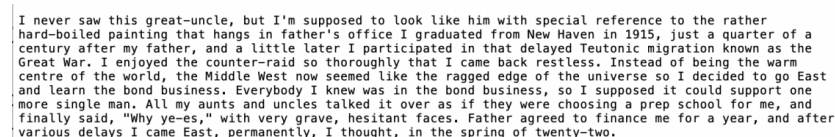
2. Text Source and Tiding-Up of Text

This paper selects *The Great Gatsby* by F. Scott. Fitzgerald as the content of text analysis.

To make the text orderly, this paper has done the collation before analysis, such as deleting redundant spaces and carriage return, separating the incorrectly connected words, modifying the wrong spelling, etc. After manual correction, the total tokens of the text changed from 49,992 to 50,353. Selected manual correction results are presented in Figure 1-2.

```
I never saw this great-uncle, but I'm supposed to look like himwith special reference to the rather
hard-boiled painting that hangs in father's office I graduated from New Haven in 1915, just a quarter of a
century after my father, and a little later I participated in that delayed Teutonic migration known as the
Great
War. I enjoyed the counter-raid so thoroughly that I came back restless. Instead of being the warm centre of
the world, the Middle West now seemed like the ragged edge of the universeso I decided to go East and
learn the bond business. Everybody I knew was in the bond business, so I supposed it could support one more
single man. All my aunts and uncles talked it over as if they were choosing a prep school for me, and finally
said, "Whyye-es," with very grave, hesitant faces. Father agreed to finance me for a year, and after various
delays I came East, permanently, I thought, in the spring of twenty-two.
```

Figure 1: Selected text before modification

```
I never saw this great-uncle, but I'm supposed to look like him with special reference to the rather
hard-boiled painting that hangs in father's office I graduated from New Haven in 1915, just a quarter of a
century after my father, and a little later I participated in that delayed Teutonic migration known as the
Great War. I enjoyed the counter-raid so thoroughly that I came back restless. Instead of being the warm
centre of the world, the Middle West now seemed like the ragged edge of the universe so I decided to go East
and learn the bond business. Everybody I knew was in the bond business, so I supposed it could support one
more single man. All my aunts and uncles talked it over as if they were choosing a prep school for me, and
finally said, "Why ye-es," with very grave, hesitant faces. Father agreed to finance me for a year, and after
various delays I came East, permanently, I thought, in the spring of twenty-two.
```

Figure 2: Selected text after modification

3. Tools Used

(1)AntConc (version 4.0.11): AntConc is a freely-available desktop corpus search and concordancing program available for Windows, Mac, and Linux. AntConc is easy to use but offers many powerful functions for analyzing corpora, including concordancing, collocation search, word frequency, keyword, and n-gram lists. It also, importantly, supports the use of regular expressions in search queries.

(2)TagAnt (version 1.3.0): TagAnt is a lightweight piece of software that can simplify how to assign POS tags to a text or an input list of text files that are UTF-8 encoded.

(3)UAM CorpusTool (version 41.0.2272.76): The UAM CorpusTool is a state-of-the-art environment for annotation of text corpora. It can be used to annotate a corpus as part of a linguistic study or build a training set for use in statistical language processing according to the needs of users.

4. Research Method

In this research, AntConc is the main corpus retrieval software, in which KWIC, Plot, Word and Keyword functions are used to analyze the four stylistic features of lexical features, phrase features, syntax features and rhetorical features. TagAnt is used to automatically assign parts of speech tags to words and other tokens from the given plain text content to assist AntConc in analyzing lexical features, while the UAM CorpusTool is used to assist AntConc in phrase collocation retrieval.

5. Results and Discussion

5.1. Lexical Features

5.1.1. Past Tense and Perfect Tense Words

An important method to analyze literary works is to search the high-frequency words in the works. Graeme Kennedy proposed that with the auxiliary means of the corpus, we can conduct detailed and substantial analyses of various texts and achieve efficient and useful vocabulary descriptions. [2] Click the option "Word" in AntConc software, and we can check the frequency of each word in the collected corpus, so as to determine the words that are valuable for research. General functional words are not included in this list. Although functional words corresponding to content words appear frequently in the text, the analysis of these words can not enlighten people. This study can omit functional words when analyzing high-frequency words in text. The results of the most commonly used 20 high-frequency words (excluding function words) are presented in Table 1.

Table 1: 20 High-frequency words of The Great Gatsby (excluding function words)

N	Word	Freq.	N	Word	Freq.
1	I	1379	11	him	274
2	he	839	12	my	243
3	was	768	13	we	233
4	you	520	14	said	232
5	his	492	15	there	214
6	she	412	16	Tom	189
7	her	395	17	Daisy	183
8	had	380	18	one	177
9	Gatsby	370	19	were	172
10	me	326	20	they	166

It can be seen from Table 1 that high-frequency words include a large number of past tense words, such as "was", "had", "said", and "were", which have appeared 768, 380, 232, and 172 times in this text respectively. Categories for further analysis were generated mainly simply by scrutinizing the trees.[3] To further understand the proportion of past tense words in the text, TagAnt is used to tag part of speech, and then AntConc is used to retrieve the total hits of VVD, VVN, VBD, VHD, VHP and all the verbs in the text. The results are presented in Table 2.

According to Table 2, it can be found that 49.42%, namely nearly half of the verbs are in the past tense or perfect tense. This can explain why the story in the novel happened in the past. It is the

memory and description of the characters of the past. Many things have already happened. What happened in the past cannot be made up. It also lightly reveals a sad and nostalgic emotional tone.

Table 2: Hits, sum, proportion

	VVD	VVN	VBD	VHD	VHP	V(Regex)
Hits	3026	1101	987	426	127	11467
Sum	5667					11467
Proportion	49.42%					

5.1.2. Color Words

Table 3: Statistics on the use of color words in The Great Gatsby

Rank	Freq.	Color Words	Rank	Freq.	Color Words
1	48	white	5	17	green
2	22	blue	6	11	black
3	22	yellow	7	10	red
4	18	gray	8	6	silver

In literary works, various colors have their unique connotations. This novel skillfully uses a large number of color words to depict the character, and the characters in the play are vivid and full of beauty. According to above Table 3, "white", "blue", "yellow", "gray", "green" and other color words frequently appear in the novel, which play a certain role in promoting the characterization and plot development of the novel. The Row Index of "white" in *The Great Gatsby* is shown in Figure 3.

	File	Left Context	Hit	Right Context
17	The...	nnected this Gatsby with the officer in her	white	car. When Jordan Baker had finished telling all t
18	The...	way places to which they had driven in her	white	car. Just as Daisy's house had always seemed to
19	The...	' called Gatsby. We slowed down. Taking a	white	card from his wallet, he waved it before the mar
20	The...	omewhere, though its feet were startlingly	white	changed hands and settled down into Mrs. Wilse
21	The...	ny shadow westward as I hurried down the	white	chasms of lower New York to the Probity Trust.
22	The...	s Island a limousine passed us, driven by a	white	chauffeur, in which sat three modish negroes, t'
23	The...	d the child calmly. "Aunt Jordan's got on a	white	dress too." "How do you like mother's friends?"
24	The...	ith and bought him a blue coat, six pair of	white	duck trousers, and a yachting cap. And when th
25	The...	tcher on which lies a drunken woman in a	white	evening dress. Her hand, which dangles over th
26	The...	His heart beat faster and faster as Daisy's	white	face came up to his own. He knew that when he

Figure 3: Row index of "white" in The Great Gatsby

"White" is a symbol of purity, but it is often associated with the selfish, self-serving rich in the novel, whose outer whiteness cannot hide their inner darkness, but makes the readers feel their hypocrisy and indifference more deeply.[4] A search of "Plot" for "white" reveals that the name "Daisy" is derived from a flower, which has a yellow center surrounded by white. It is symbolic: the yellow heart symbolizes Daisy's philosophy that money is the most important thing in life, while the white color accompanies her throughout the novel. For example, she lives in a white house in the palace, likes to wear white dresses and often drives a white car. White, beautiful and hollow, while white to the extreme, is cold and terrifying. It is another reflection of Daisy's selfish character and the emptiness and indifference of high society.

In addition, "gray" is another important color word in the novel. It is the dominant color of the ruins of the village, where everything is gray, symbolizing the spiritual negativity and death of the villagers. The word "gray" is mostly used in conjunction with Wilson, a man with an ashen complexion who lives a dull day-to-day life in a car dealership with no purpose or meaning. After the tragic death of his wife, he is instigated to kill Gatsby, and the spiritual death suggested by the gray tones becomes clear after Gatsby's death.

Other color words like "yellow" and "silver", which appear several times in the text, are mostly used to refer to the wealth of the upper class, describing the extravagant life of the rich. "Blue" and "green" are a symbol of purity in contrast to "white". "Red" refers to bloodshed, and violence appears frequently in many scenes of the novel, implying that it is caused by the mental disorder of people in an unstable society.

The frequent use of color words not only deepens the readers' understanding of the characters, but also gives the text a sense of color and beauty.

5.2. Phrase Features

The Great Gatsby is rich in male and female characters, with Tom and Gatsby as the main male characters and Daisy, Jordan Baker and Myrtle Wilson as the main female characters. The whole novel is based on simple dialogue. By searching the collocations of "he" and "she" with different verbs through the UAM CorpusTool and analyzing their actions and expressions when speaking, we can find out the author's intention in portraying the male and female characters. The results are presented in Figure 4 and Figure 5.

Figure 4: "he + v." phrase collocation in The Great Gatsby

Figure 5: "she + v." phrase collocation in The Great Gatsby

According to the figures, "he" is often associated with "approved", "remarked", "advised", "reformed", "had", "demanded", "killed" and other dominant verbs, reflecting the sense of male

superiority that prevailed at that time. In contrast, "she" is often collocated with "cried", "laughed", "complained", "snapped", "nodded", "whispered", "hesitated" and other verbs with strong emotional or submissive overtones, reflecting the over-emotional nature of women and the fact that women's fates are often controlled by men under the oppressive patriarchal culture, with certain tragic connotation.

5.3. Syntax Features

Reflecting a text-linguistics approach, Halliday and Hasan proposed the notion of cohesion, such as the connective word "and", whereby explicit linguistic devices relate one element in the text to another and thus create texture through cohesive ties between structurally unrelated propositions.[5]

The word "and" appears 1,549 times in this 50,353-word novel. In addition to maintaining coherence, "and" is used deliberately in many instances to achieve a purpose, that is what Halliday refers to as "out-of-convention or below-convention loss of coordination in frequency".[5] For example, in the fourth paragraph (2 sentences) of Chapter 3, "and" occurs 15 times. In Chapter 4, the author devotes several paragraphs to the extravagance of the Gatsby House party, among them, "and" appears 11 times in paragraph 4 (3 sentences), 14 times in paragraph 7 (3 sentences), and 17 times in paragraph 9 (4 sentences). "And" appears in a number of passages at an unconventional and high frequency, as if a string of broken beads. In the novel, guests arrive one after the other to join in the all-night revelry. This reflects, on the one hand, Gatsby's desire to be part of high society and, on the other, the emptiness and loneliness of the upper classes in the 1920s.

5.4. Rhetorical Features

In order to emphasize the theme of the novel, the writer uses a variety of rhetorical devices, including metaphor, simile, hyperbole, symbolism, repetition, parallelism and contrast. The most common figurative words are "as" and "like", which are used to create vivid characters, such as the arrogant Tom, the selfish Daisy and the obsessive Gatsby. This study will analyze the characteristics of the row index of the word "like" and several are excerpted in Table 4.

Table 4: Row index of "like" in The Great Gatsby

Rank	Text
1	In his blue gardens men and girls came and went /like/ moths among the whisperings
2	broke down entirely and resorted to flank attacks at intervals she appeared suddenly at his side /like/ an angry diamond
3	Daisy and Jordan lay upon an enormous couch, /like/ silver idols weighing down their own white dresses against the singing breeze of the fans.

In the first example, the writer compares the men and women at the party to moths, because they have the same way of life as moths, preferring to gather in bright places and live blindly in parasitism. At night, when the lights are on, it flies to feed; when the lights are off, it is nowhere to be found. The simile reveals the snobbish nature of the various diners at Gatsby's banquet, the luxurious and dissipated lifestyle of the lost youth of the Jazz Age, and the disgust of the narrator, Nick.

In the second example, the novel uses the term "angry diamond" to refer to the angry wife who questions her husband's love for her because he is a party-goer and has no desire to return home. The diamonds are traditionally regarded as a symbol of love and loyalty, but here the "angry diamond" is an ironic reference to the seemingly precious but hypocritical love between the rich, reflecting the hypocrisy of the upper class.

In the third example, Daisy and Jordan are compared to "silver idols", rich in life but empty in spirit. The emptiness of high society is evident in their abundance of food and clothing, their lack of purpose in life and their inability to make decisions for their own lives.

From the above, we can learn that Fitzgerald likes to use objects with negative overtones or imply negative meanings as vehicles, by which he modifies the novel's upper-class people and deepens the readers' negative impression of them. It is undeniable that the use of rhetoric creates vivid characters, contributes to the novel's plot and has a significant impact on the interpretation of the novel.

6. Conclusion

In conclusion, the article adopts a corpus research method, using AntConc, TagAnt and UAM CorpusTool as research tools to examine the rich linguistic features in *The Great Gatsby* by analyzing its stylistic features in terms of lexical, phrases, syntax and rhetorical features, and then to explore F. Scott. Fitzgerald's unique writing style.

Through the above analysis, conclusions can be drawn that Fitzgerald likes to use past tense words to shape the background of the times and color words to create negative characters, such as the obsessive Gatsby, the arrogant Tom, the selfish Daisy and the aimless Wilson. Fitzgerald is also good at using different phrase collocations to portray males and females, reflecting the patriarchal nature of the time. In the use of syntactic deviations and rhetorical devices, Fitzgerald deliberately portrays a large number of negative characters, revealing the wealthy but indifferent and hypocritical society of 1920s America, setting the scene for the novel's tragic ending, and reflecting his rigorous writing style.

By using corpus as a supplementary tool to analyze this novel, it can be found that the corpus search software can provide a more objective perspective on the work by analyzing the literature through quantitative methods, and also provide a new perspective and method for future literary criticism of the novel *The Great Gatsby*. However, the disadvantage is that this study only examines one representative work of the author, and a quantitative corpus analysis of the author's other works can be carried out subsequently to explore other characteristics of the author's writing and its changing development.

References

[1] F. Scott. Fitzgerald. The Great Gatsby. http://www.qcenglish.com/ebook/927.html
[2] Kennedy G. An Introduction to Corpus Lin- guistics [M]. Beijing: Foreign Language Teaching and Research Press, 2000: 91.
[3] Wei L. Syntactic Markedness as a Stylistic Feature in the Great Gatsby [J]. Studies in Linguistics and Literature, 2018, 3(1).
[4] Raksangob W. Corpus stylistics and colour symbolism in The Great Gatsby and its Thai translations [J]. Language and Literature, 2022, 31(3).
[5] Halliday M. A. K., & Hasan R. Cohesion in English. New York: Longman, 1976.