| **RESEARCH ARTICLE**

A Corpus-based Analysis of *The Time Machine*: From the Perspective of Literary Stylistics

Hongming Fang
College of Foreign Studies, Jinan University, Guangzhou, Guangdong, China
Corresponding Author: Hongming Fang, **E-mail**: hommy840469792@163.com

| **ABSTRACT**

From a corpus stylistic perspective, this paper analyses H.G. Wells' book *The Time Machine* at the level of punctuation and vocabulary, etc. By utilizing the Sketch Engine and AntConc software, this paper finds that the author excels at using a large number of content words to present information and makes extensive use of transitive conjunctions in the course of the narration in order to present information that is difficult for the readers to anticipate. Based on the above methods, the author creates a contextual atmosphere suitable for the construction of the scientific discourse, eventually achieving an efficient degree of language code transmission with readers.

1. Introduction

Over the past years, literary scholars and linguistics have paid much attention to the study of literary works. In most cases, however, they only apply the respective theory to the literary work itself or simply share their reading insights, which only scratch the surface of literary studies. It would be difficult to draw convincing conclusions without the use of scientific corpus tools to explore the linguistic and textual dimensions of literary works. In this context, since the 1980s, with the rapid development of computer science and network technology, more and more stylistic researchers abroad have combined the tools and methods of corpus linguistics with the analysis of literary works.

The application of corpus linguistics to language research has a long history and has been a hot topic of research in academia. The corpus can be used to conduct research on language acquisition, phonological studies, English grammar, and mechanically controlled corpora. It can also be applied to the study of translation. Nowadays, the number of corpus-based literary studies is also increasing, and this paper, in this background, is an exploration of the theme of fiction using corpus tools. This paper has chosen *The Time Machine* as the textual material for this study. *The Time Machine* is a long science fiction novel written by H.G. Wells, which depicts the two main themes of evolution and class antagonism in human society through the main line of time travel, critiquing and profoundly thinking about human nature and human society. Based on the corpus stylistic method, this paper explores the linguistic and stylistic features of *The Time Machine* through a combination of qualitative and quantitative analyses with the help of corpus research methods.

2. Literature review

As corpus stylistics focuses specifically on the analysis of the linguistic features of literary texts, the communion between corpus linguistics and stylistics is not only found in the general principles of linguistic features that encompass both literary and non-literary texts but also in the relevance of going beyond the conventional analysis of the linguistic features of literary texts (Li & Wang, 2009).

As an emerging corpus-oriented research discipline, corpus stylistics combines corpus linguistics and stylistics in the service of uncovering surface features and deep systems of language. Milic (1967) uses a combination of qualitative and quantitative methods, combined with corpus tools, to provide an in-depth exploration of Swift's work. Burrows (1987) was an early experimenter in the quantitative analysis of texts using multivariate statistics. He summarized the common words used in Austen's works by counting word frequency and conducting significant difference analysis, which deepened readers' understanding and appreciation of the characteristics of his works. Burgess (2000) shows how corpus tools can be used to build syntactic levels and shape authorial style through word frequency lists, co-occurrence index lines, and collocations. Baker (2006) combines linguistic concepts such as discourse meaning and discourse features to integrate corpus linguistic research methods into discourse analysis.

Strictly speaking, corpus stylistics can be subdivided into broad and narrow concepts. The corpus stylistics mentioned in this paper mainly involves stylistics in a narrow sense, that is, the corpus research of literary stylistics. At a macro level, literary stylistics is the study of literary texts using various linguistic tools, typically by applying the analytical methods of linguistic science and a system of categorization to study genres such as poetry, fiction, and prose. This paper utilizes theories from various linguistic sub-disciplines, such as sociolinguistics, pragmatics, cognitive linguistics, and historical linguistics, to describe the phonological, syntactic, and semantic dimensions of language (Wynne, 2006). The advantage of corpus stylistics over previous literary studies, which have remained at the theoretical level, lies in the data visualization of linguistic features. Linguists can use the collection of code-based data from texts to more scientifically validate relevant literary theories. In addition, the common concern with the relationship between form and meaning has further contributed to the integration of stylistics and corpus linguistics. This combination of quantitative and qualitative research provides a more reliable database for the study of corpus stylistics.

From a micro perspective, in the field of corpus stylistics, linguists have often used the terms "stylometrics" or "statistical stylistics". Although there are a few differences in data collection and statistical methods, these branches of research have different emphases. Corpus stylistics often emphasizes the special-purpose use of corpora, i.e., the code-based annotation of corpora, the association of word-symbol classes through word frequency statistics, grammatical assignment, semantic analysis, etc. At the same time, of course, manual annotation and retrieval statistics are also an integral part of the process. It is important to emphasize that such corpus stylistic research can be divided into two main categories: theory-driven stylistic research and theory-based stylistic research. Literary theory plays a more flexible role in research; for example, if we want to study the narrative style of a literary work, we can assume the narrative path of the text before conducting word frequency statistics on narrative words and conjunctions in the text, or we can verify the narrative style in the process of completing the word frequency of the text.

The use of corpus in the analysis of literary works provides us with a new perspective on the analysis and appreciation of literary works. Although this research theory has achieved some success, we still need to be aware of the shortcomings of development in this area of research. For example, most of the studies have used a corpus to analyze the themes of literary works, taking a particular novel as an example. With the help of corpus search software, they use data to show the storyline, linguistic features, and rhetorical techniques of the work, but unfortunately, they do not go far enough, and the data are not clear enough. Therefore, the analysis of textual themes under corpus stylistics needs to be further strengthened. It is necessary to make full use of corpus tools to analyze literary texts from multiple perspectives and further explore the deeper meaning of literary texts. At present, however, the application of corpus stylistics by domestic scholars mainly focuses on literature, especially the fiction genre, but less on science fiction. In this paper, the author plans to expand and study this issue accordingly.

On the whole, this paper will be based on the corpus research approach and explore the stylistic and linguistic characteristics of literary works from the perspective of corpus stylistics. This emerging approach to the study of literary works can be regarded as one of the corpus stylistics. In general terms, corpus stylistics is an emerging field of research that combines the paradigm of corpus linguistics research with stylistic research (Lu & Xia, 2010). It is also because of this scientific principle that the use of corpus tools in the study of literary texts allows scholars to move beyond the theoretical textual level and derive a deeper connotation of the text through its linguistic features themselves. In this sense, therefore, corpus stylistics, in addition to its own embedded systematic communion, also reveals a significant harmonization of linguistic and non-linguistic forms.

3. Research design
This study is based on the research carried out on *The Time Machine*: an analysis of the genre features of the novel based on a corpus tool in accordance with quantitative statistics. The research perspective of the corpus considers the following main aspects:

Wells' science fiction changed the optimistic tendencies and recaptured the anxiety and unease about the future that had characterized English literature; but because there was a certain benevolence in Wells' personality, his pessimistic works were always accompanied by glimmers of hope, and most of them ended on an optimistic note in *The Time Machine*, the main character ends on a path to which he will never return, which is inevitably thought-provoking and should therefore be treated with special consideration when analyzing this work. This paper will therefore collect and analyze data on the frequency of punctuations and

words in the text from a stylistic perspective, including lexical features and syntactic features, to explore the text more comprehensively.

4. Results and discussion

In the following part, this paper will utilize the corpus analysis software AntConc3.4.3 and the Sketch Engine to collect and analyze data on the frequency of words in the text from a stylistic perspective, including punctuation, lexical features, and syntactic features, to provide a more comprehensive discussion on the text.

4.1 Punctuation

Punctuation is a symbol that aids the written record of language and is an integral part of the written word. It is used to indicate the pause in an utterance and the tone of voice. Further speaking, it is also to mark the specific nature and role of certain components (mainly words). After searching and counting, the summary of punctuation marks in the text is as follows：

Table 1: Punctuation Rates of *The Time Machine*

Punctuation	Frequency	Percentage (%)
Comma	3036	46.95
Period	2333	36.08
Double inverted commas	647	10.01
Dash	200	3.09
Semicolon	123	1.90
Colon	58	0.90
Exclamation mark	69	1.07
Total	6466	100

Table 1 shows the corpus data from *The Time Machine* Corpus. It can be seen that this science fiction novel is consistent with other fictional text types in its high frequency of use of periods and commas. However, as a literary work narrated in the third person, the book has a higher use of double inverted commas and colons than other general texts.

4.2 Lexical features

In terms of verbs, the author uses the Sketch Engine to count the frequency of verbs and summarize them as follows:

Table 2: Frequency of Verbs in *The Time Machine*

Word	Frequency	Percentage (%)
be	1498	36.43
have	776	18.87
see	192	4.67
come	181	4.40
do	156	3.79
say	131	3.19
go	131	3.19
feel	108	2.63
think	100	2.43
seem	100	2.43
look	100	2.43
take	95	2.31
make	93	2.26
find	92	2.24
know	80	1.95
get	68	1.65
begin	62	1.51
leave	54	1.31
tell	48	1.17
grow	47	1.14
Total	4112	100

Apart from the verbs "have" and "be", which are verbs of presence or possession, the top verbs in the text are "see", "come", "do", "say", "go", and "feel". For example:

> …*and almost immediately the second. I seemed to reel; I felt a nightmare sensation of falling; and, looking round, I* **saw** *the laboratory exactly as before. Had anything happened? For a moment, I suspected that my intellect had tricked me…*

As shown above, the author likes to use the verb "see" to present information. Below is a graph of the frequency of the verb "see" in the text, with the black vertical lines representing the distribution of its occurrence in the text.

Figure 1: Occurrence of Verbs in *The Time Machine*
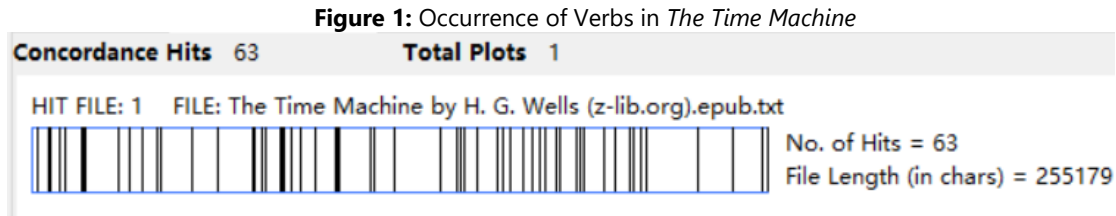
It can be seen that the verbs appear more evenly distributed throughout the text and begin to decrease towards the end of the text (most likely because the main character becomes familiar with the new world over time). It is clear that the text uses these verbs to describe what the main character sees and hears when he visits the new world through *The Time Machine*. Therefore, it is certain that the author's specific choice of verbs makes the narrative clearer and more coherent.

As for nouns, we can see the table selected from the Sketch Engine below:

Table 3: Frequency of Nouns in *The Time Machine*

Word	Frequency	Percentage (%)
time	254	16.83
Andrew	121	8.02
thing	109	7.22
man	102	6.76
machine	93	6.16
hand	87	5.77
way	76	5.04
eye	61	4.04
traveler	61	4.04
night	59	3.91
people	56	3.71
life	55	3.64
day	55	3.64
Weena	54	3.58
face	51	3.38
moment	50	3.31
world	46	3.05
Marie	42	2.78
place	40	2.65
creature	37	2.45
Total	**1509**	**100**

In addition to the important names and the subject words "time" and "machine", the most frequently occurring nouns are biologically related words such as "hand", "eye", "face", and "creature". In *The Time Machine*, the author devotes a great deal of space to a clear and vivid description of the appearance of human beings and new types of humans in the future world, reflecting the author's bold predictive power. The extensive use of physical and sensory language also serves to heighten the readers' authentic emotions and give them a more personal sense of the authors' adventures. For example:

> …*rescuing her from the bottom of one of those water tanks stage magicians immersed their beautiful assistants in, bound* **hand** *and* **foot**, *or as though his sense of direction were so good it saved him from getting lost in the maze like her other lovers…*

Wells excels at describing the differences between the Eloi and the Morlocks in vivid detail: the Eloi live on the ground, feed on fruits and vegetables, and live in groups. They are small in stature, slender in limbs, fair-skinned, and simple-minded, and all they do is play and have fun. The Morlocks, on the other hand, live underground all year round and only come out at night to feed on the Eloi. They are cunning, cruel, and bloodthirsty. By utilizing specific words about the human body, the readers can get a very clear picture of what the Eloi and the Morlocks look like.

As for conjunctions, we can see the table selected from the Sketch Engine below:

**Table 4:** Frequency of Conjunctions in *The Time Machine*

Word	Frequency	Percentage (%)
and	1484	77.82
but	272	14.26
or	123	6.45
nor	9	0.47
either	6	0.31
yet	5	0.26
both	5	0.26
neither	3	0.16
Total	**1907**	100

The text makes extensive use of transitive conjunctions, including "and", "but", etc., for example:

*...His grey eyes shone **and** twinkled, and his usually pale face was flushed **and** animated...*

Wells does well in using transitive conjunctions to subvert conventional perceptions and to make it easier for the author to explain new information. The skilful use of conjunctions, therefore, provides syntactic support for the author's presentation of new information. Here is a graph of how frequently the "and" conjunction appears in the text:

Figure 2: Occurrence of Conjunctions in *The Time Machine*
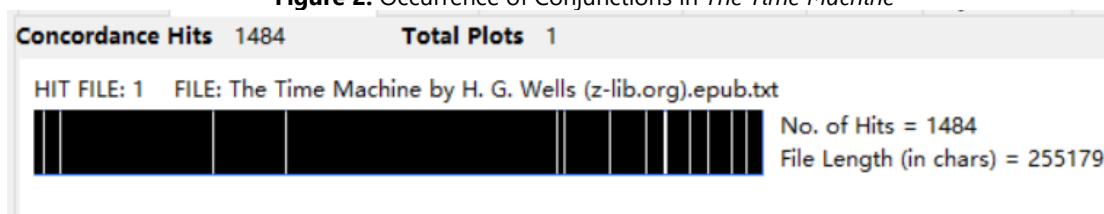
It is clear to see that the conjunctions appear very frequently and cover a wide range of topics in the text. The extensive use of the "and" conjunction also provides syntactic support for the author's narrative.

4.3 Discussion

In order to minimize the impact of the formalist approach to stylistic analysis, we need to look through the text itself to analyze the work's textual features and cultural connotations and try to decipher the deeper meaning of the author's diction.

Wells' science fiction is full of obscurity and pessimism. He is adept at drawing on the fantasy future world to reveal and criticize the realities of late Victorian England society in late Victorian England, expressing his dissatisfaction and anxiety about the dark side of capitalist society and his doubts and worries about the future of humanity (Zhang,2008). Such a starting point requires Wells to have his unique consideration in wording. In terms of punctuation, the fact that most of the sentences in the text are complex sentences with subordinate clauses has led the author to use commas extensively in the text and make them the main punctuation mark in the text. In this work, the author mixes three different literary genres: science fiction, political allegory and outlaw fiction. Each of the three genres gives the novel a special meaning, and together they strengthen the social criticism of the text (Yu, 2020). Therefore, when it comes to the text, existential sentences are present in order to explain and scientifically explain new things more directly. In addition, Wels makes extensive use of content words to explain new things so that the readers can get a clearer picture of *The Time Machine*. In this context, we need to clarify the cultural context of the text. Wells' *The Time Machine* is recognized as one of the earliest true masterpieces of science fiction.

Wells has been trying to further explore the causes and undesirable consequences of the formation of class antagonism in human society by explaining the future state of human existence. And in the second half of the text, Wells begins to envision his plan for social improvement boldly. It is important to note that his approach to improvement is not necessarily one of fierce class struggle or social revolution; rather, his initial intentions are more in favor of positive social moderation and improvement. In his youth, he believed that the Marxist materialist way of struggle was the ultimate destination of all revolutions and that the proletariat could stand up solely by its struggle and efforts. Therefore, in the text, we can clearly see the presence of a large number of real verbs, whose existence is meant to help the author elucidate the process of human antagonistic events. From the above analysis, we see Wells progressively unveiling a more obscure and cold truth than one through the time traveler's deepening observation and understanding of the future society.

In summary, combined with the textual characteristics of science fiction, it is not difficult to find the special intention of Wells in the book. He has been sparing no effort to depict his dream picture of a new, improved society through the use of content words based on maintaining the essential characteristics of science fiction, thus vigorously criticizing the evil nature of capitalism, and this unique textual characteristic also makes this book a classic in the history of human civilization.

5. Conclusion

This paper analyses H.G. Wells' book *The Time Machine* from a corpus stylistic perspective, at the level of punctuation and vocabulary, and finds that the author excels at using a large number of sensory verbs and sensory nouns to present information and makes extensive use of transitive conjunctions in the course of the narration in order to present information that is difficult for the reader to anticipate. There are some shortcomings in this paper: for example, it does not fully explore all lexical forms and fails to analyze the stylistic features of the text in terms of syntactic linguistic representations.

This study opens up a new branch of research in the field of corpus studies and makes a certain contribution to the influence of lexical syntax on the stylistic aspects of literary genres. This study explores in greater depth the synergistic relationship between lexical and syntactic aspects of corpus stylistics, i.e., the relationship of mutual influence and interdependence.

Conflicts of Interest: The author declares no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References
[1]   Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum.
[2]   Burgess, G.J.A. (2000). *Working with German Corpora*. Birmingham University Press.
[3]   Burrows, J.F. (1987). *Computation into Criticism*. Clarendon Press.
[4]   Li, T., & Wang, J. (2009). Corpus Stylistics: Computer-assisted Stylistics Analysis of Literary Texts. *Technology Enhanced Foreign Language Education,* (01),20-24.
[5]   Lu, W., & Xia, Y. (2010). Corpus Stylistics: A New Approach to Literary Stylistics. *Journal of Foreign Languages,* (01):47-53.
[6]   Ma, G. (2005). Corpus-based Stylistic Analysis of Literary Texts. *Journal of Changshu Institute of Technology,* (05):4-6. doi:10.16101/j.cnki.cn32-1749/z.2005.05.002.
[7]   Milic, L.A. (1967). *Quantitative Approach to the Style of Jonathan Swift*. Mouton.
[8]   Wynne, M. (2006). *Stylistics: Corpus Approaches* (2nd ed.). Elseveier Science.
[9]   Yu, J. (2020). *Time Machine* and Its Trans-Genre Narrative Art. *English Studies*, (01):41-50.
[10]   Zhang, Y. (2008). Interpreting the Moral Spirit of Critical Realism in *The Time Machine. Foreign Language and Literature Studies,* (04):277-282. doi:10.19716/j.1672-4720.2008.04.011.