| RESEARCH ARTICLE

A Corpus-based Comparative Analysis of Linguistic Features in *Silent Spring* and *To Kill a Mockingbird*

Li Lin

College of Foreign Studies, Jinan University, Guangzhou, Guangdong, China
Corresponding Author: Li Lin, **E-mail**: 1757469422@qq.com

| ABSTRACT

Silent Spring and *To Kill a Mockingbird* are two novels written by two American women - Rachel Carson and Harper Lee, respectively. The study attempts to comparatively uncover the linguistic features in the two novels by means of the corpus tool Multi-dimensional Analysis Tagger and the statistical tool SPSS. It is found that the text of *Silent Spring* is quite different from the text of *To Kill a Mockingbird*, among which *Silent Spring* is classified into the register of "general narrative exposition", and *To Kill a Mockingbird* is classified into the register of "Imaginative narrative". Besides, the text of *Silent Spring* is characterized by more structures of that relative clauses on subject position, prepositions, attributive adjectives, long words, downtoners, phrasal coordination, agentless passives and conjuncts. However, the text of *To Kill a Mockingbird* is featured by more structures of past tense, verbs, analytic negation, direct WH-questions, first person pronouns, subordinator that deletion and predictive modals.

| KEYWORDS

Corpus; Linguistic features; Comparative analysis; *Silent Spring*; *To Kill a Mockingbird*

| ARTICLE INFORMATION

1. Introduction

Silent Spring is a novel created by Rachel Carson, the noted American marine biologist and groundbreaking environmentalist, which was first published in 1962. It describes that human beings may face a world without birds, bees and butterflies and reveals the harm of chemical pesticides in all directions. *To Kill a Mockingbird* is a novel written by Harper Lee and was published in 1960. It deals with the serious issues of racial inequality. Both of them are works of the 1960s, and the authors are both women.

As corpus analysis has become a valuable tool that allows for a comparison between individual works, collections, or a specific language variety (McEnery and Wilson 2001, 117), the study attempts to comparatively uncover the linguistic features in the two above novels by means of corpus analysis. The study attempts to answer the following questions: (1) What are the general features of *Silent Spring* and *To Kill a Mockingbird*? What are the similarities and differences? (2) What are the differences in the use of internal linguistic features between *Silent Spring* and *To Kill a Mockingbird*? In order to answer these questions, the study will adopt the corpus tool Multi-dimensional Analysis Tagger and then the statistical tool SPSS.

2. Corpus Stylistics

Corpus Stylistics is the quantitative study of language applied to the qualitative study of linguistic style. Corpus stylistics investigates the relationship between meaning and form. Thus, it is similar to both stylistics and corpus linguistics. Whereas stylistics pays more attention to deviations from linguistic norms that lead to the creation of artistic effects, corpus linguistics mainly focuses on repeated and typical uses, as these are what the computer can identify (Hoey et al., 2007).

The study of corpus stylistics abroad started earlier and can be roughly divided into five aspects. The first aspect is the study of the theme and significance of the literary works. Green (2017) investigated the keywords and their distribution of Western literary classics from a macro perspective and found that their thematic significance has a positive bias on the whole, and the weakening of religious elements in the literature shows a trend of cultural omics. The second aspect is the study of the language style in the literary works. Stubbs and Barth (2003) constructed three corpora, FICTION (mainly fictional novels), BELLES (literature, autobiography, memoirs, etc.) and LEARNED (academic works such as natural social sciences, humanities, etc.), and analyzed their frequency and distribution from the aspects of keywords, multi-word strings and repeated lexical-grammatical patterns in order to present the linguistic features of different literary genres. Biber (2010) focuses on keywords, keyword clusters and word collocation and analyzes the unique language style of specific articles or authors. The third aspect is the study of the characters in the novels. Stockwell and Mahlberg (2015) use CLiC tools to obtain the context that the address of "Mr. Dick" appears in non-quotes in David Copperfield and analyze these text types with the characterization model of cognitive poetics to interpret the relationship between Mr. Dick and the narrator. Moustafa (2022) comparatively analyzes the thematic foci as well as characterization aspects and interpersonal relations between characters in the two novels by dint of corpus linguistic tools, i.e., keywords and key clusters. The fourth aspect is the study of the discourse structure and narrative features of the literary works. Jaafar (2019) examines the mind style of the leading character in the 1997 film As Good as It Gets performed by Jack Nicholson as Melvin Udall, a novelist who is characterized as a misanthrope and suffers from obsessive-compulsive disorder (OCD) with the help of the Wmatrix. Busse (2020) develops a new corpus-stylistic approach for systematically analyzing the different narrative strategies of historical discourse presentation in key pieces of 19th-century narrative fiction, thus identifying diachronic patterns as well as unique authorial styles and placing them within their cultural-historical context. There are also studies that focus on symbolism Abdul Ghani and Yasir (2022) and etc.

The study of corpus stylistics started late at home, and thus there are a few studies on it. Overall, they can be divided into three aspects. First, it is the study of the theme and significance of the literary works. For example, Wei and Zhou (2020), taking WordSmith as the tool, conducted a corpus-based cognitive stylistic analysis of Nabokov's *novel The Original Laura* to examine its themes. The results indicate that the themes reflected through keywords in the group echo with those reflected through high-frequently used rhetorical devices, which are fake love and despairing marriage. Second, it is the study of the narrative features of the novels. Under the framework of corpus translation stylistics, Zhang (2022) makes a corpus-based comparative analysis of the stylistic features of Canxue's English translation and Kafka's English translations of novels and finds the similarities between them in terms of text type features and some specific temporal and spatial semantic domains. There are also studies on another aspect, such as Chen and Zhang (2018), comparing punctuations, keywords and clusters in Ulysess and the 20th Century novels.

According to the existing literature, there are few studies on linguistic features, but research in this field is of great importance for both corpus linguistics and stylistics. Therefore, this paper attempts to conduct a corpus-based comparative analysis of linguistic features in *Silent Spring* and *To Kill a Mockingbird*.

3. Research Methodology
3.1 Data
The data used in this study are all the whole chapters in the book *Silent Spring* written by Rachel Carson and the book *To Kill a Mockingbird,* written by Harper Lee.

3.2 Research Instruments
In order to achieve the purpose of the study objectively and scientifically, two kinds of research instruments are adopted in this study. They are Multidimensional Analysis Tagger 1.3.3 and Statistical Product and Service Solutions 26.0.

The Multidimensional Analysis Tagger (MAT) is a program which is based on Biber's (1988) tagger to analyze the variation of genre or text type. The main steps are as follows: it firstly calculates the frequency of 67 features in every text, and then it converts into the standardized frequency per thousand words; finally, factor analysis is produced on the data. The 67 linguistic features are divided into different dimensions according to their occurrence. Therefore, the factor scores of a single text and even the whole register can be calculated. According to the similar linguistic features that occur in different genres, Biber divided them into 8 types, and those are as follows: 1) Intimate interpersonal interaction; 2) Informational interaction; 3) Scientific exposition; 4) Learned exposition; 5) Imaginative narrative; 6) General narrative exposition; 7) Situated reportage; 8) Involved persuasion. Besides, the 6 dimensions are presented as follows: 1) Involved versus Informational Production; 2) Narrative versus Non-narrative Concerns; 3) Explicit versus Situation-Dependent Reference; 4) Overt Expression of Persuasion; 5) Abstract versus Non-Abstract Information. Here, Dimension 6 is not considered in this study because this dimension is related to spoken language.

Statistical Product and Service Solutions 26.0 (SPSS) are used in this study, which plays an important role in the statistical analysis. It can process a large amount of data in social science. In the t-test of SPSS, if the P-value is less than 0.05 (P<0.05), it means there

is a significant difference between the two samples. After the data processing by MAT, the mean value of the two texts is used to make an independent-samples t-test to evaluate whether the differences between the data are significant.

3.3 Research Questions

The study aims to address the following two research questions:

1. What are the general features of *Silent Spring* and *To Kill a Mockingbird*? What are the similarities and differences?
2. What are the differences in the use of internal linguistic features between *Silent Spring* and *To Kill a Mockingbird*?

3.4 Research Procedures

Firstly, the text of *Silent Spring* and the text of *To Kill a Mockingbird* are made into txt plain texts by each chapter. Then, MAT will process the data and report some basic statistics, including the frequency per thousand tokens for all the linguistic variables, the z-scores of the linguistic variables, the scores for the Dimensions and the closest type in each text. Thirdly, SPSS is used to conduct an independent-samples t-test for the mean values of the two texts in different dimensions so as to analyze and explore the significant differences between the two texts. Then the differences between *Silent Spring* and *To Kill a Mockingbird* in the five dimensions can be found. Similarly, the internal linguistic features on every dimension also use the SPSS to conduct independent t-test and the differences in the internal linguistic features between the two texts are discussed.

4. Results and Discussion

4.1 The General Features of Two Texts on Five Dimensions

As the dimension scores are standardized data (Z scores), the author uses SPSS to carry out an independent sample t-test on the dimension scores of the two texts. The results show that there are significant differences ($p < 0.05$) between the text of *Silent Spring* and the text of *To Kill a Mockingbird* in all five dimensions (See Table 1).

Table 1 T-test of scores on five dimensions

	T-value	Df	P-value	MD
Dimension 1	-23.657	42.476	.000	-19.306
Dimension 2	-9.378	46	.000	-4.993
Dimension 3	12.790	46	.000	4.664
Dimension 4	-5.879	46	.000	-2.125
Dimension 5	15.383	46	.000	4.012

As is shown in Figure 1, there are positive and negative differences in Dimensions 1, 3, 4 and 5, except in Dimension 2. The text of *Silent Spring* is quite different from the text of *To Kill a Mockingbird*. More specifically, on Dimension 1, the score of *Silent Spring* is a negative value of -14.09 which indicates that it belongs to the informational production and it has a lot of nouns, long words, prepositions and adjectives. However, the score of *To Kill a Mockingbird* is a positive value of 5.22 which shows that it is involved and interactional. Accompany with obvious colloquial features. It has many verbs and pronouns. On Dimension 2, the mean scores of *Silent Spring* and *To Kill a Mockingbird* are 0.11 and 5.10, severally. They are both positive values, which indicates that both of them are narrative concerns. There is no doubt that the text of *To Kill a Mockingbird* presents an extreme narrative concern. Because it tells a story of justice from the point of view of the author, and thus it presents a high frequency of past tense, third person pronouns, etc. The text of *Silent Spring* describes that human beings may face a world without birds, bees and butterflies and reveals the harm of chemical pesticides in all directions in a scientific and technical way. But some narration can also be found in the text, such as the fable at the beginning. Next, on Dimension 3, the score of *Silent Spring* is a positive value of 3.23 which indicates that indicate that the text is more context-independent (explicit reference), for example, academic prose, and thus presents WH relative clauses and nominalizations etc. However, the score of *To Kill a Mockingbird* is a negative value of -1.44 which shows that the text is context-dependent (situation-dependent reference), as in the case of general fiction, and thus presents time adverbials, place adverbials and adverbs. Then, on Dimension 4, the score of *To Kill a Mockingbird,* which is 0.02, is higher than the score of *Silent Spring* which is -2.1. It indicates that the text of *To Kill a Mockingbird* is more persuasive, and the text explicitly marks the author's point of view as well as her assessment of likelihood and/or certainty, for example, in professional letters. Finally, on Dimension 5, the score of *Silent Spring* is a positive value of 2.22 indicates that the text of *Silent Spring* provides information in a technical, abstract and formal way, for example, in scientific discourse, and it presents many passive clauses and conjuncts among other features. On the contrary, the score of *To Kill a Mockingbird* is a negative value of -1.79.
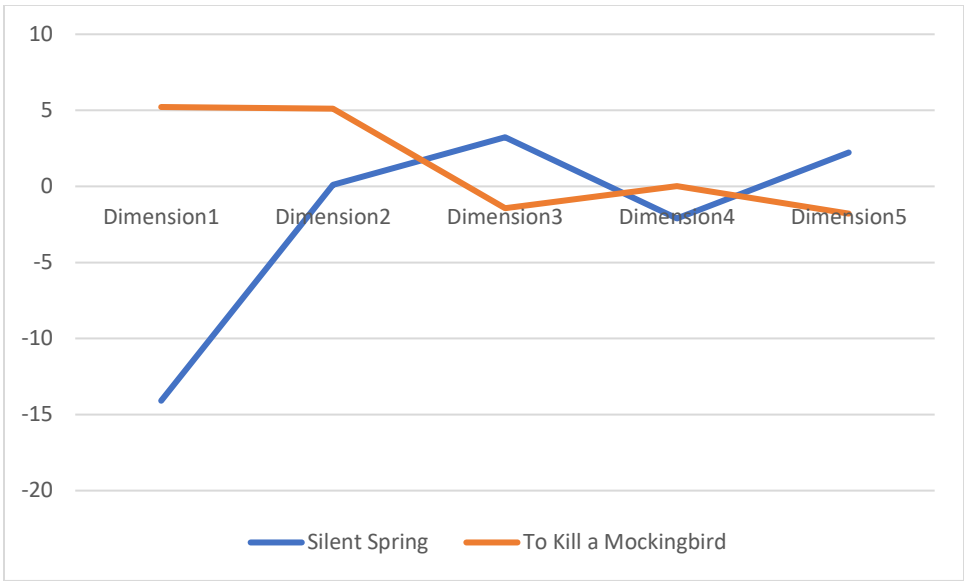
Figure 1 Comparison of scores on five dimensions

MAT finally reports that *Silent Spring* is classified into the register of "general narrative exposition", and *To Kill a Mockingbird* is classified into the register of "Imaginative narrative" (See Figure 2 and Figure 3). The former register contains sub-categories such as "press editorials, press reportage, nonsports broadcasts, biographies and science fiction", etc. The texts belonging to this register have non-literary characteristics and often convey information through narrative explanation. The latter mainly includes "romance fiction, mystery fiction, adventure fiction and general fiction", etc. The texts belonging to this register are fictional and focus on exceedingly narrative and appropriately involved.
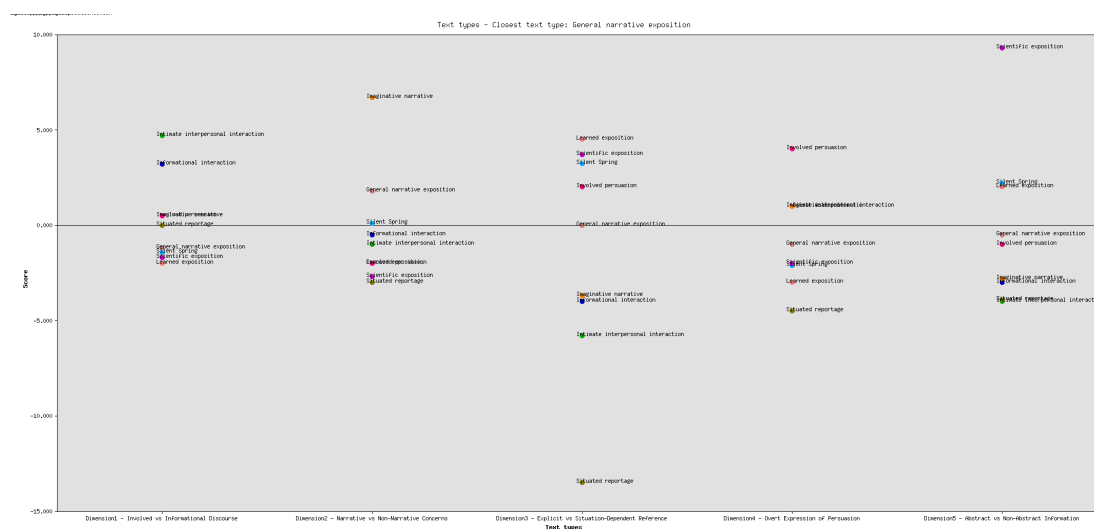

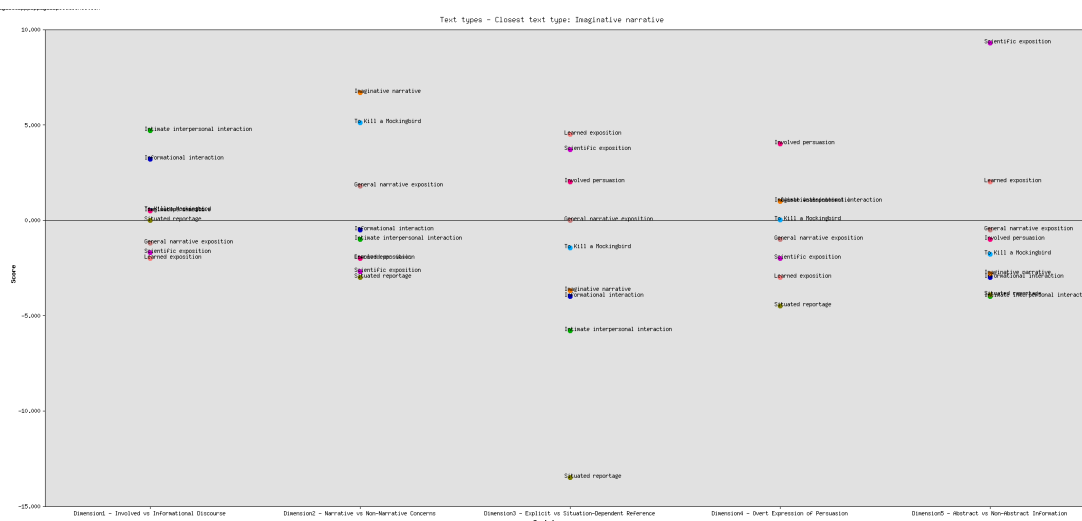**Figure 2 Text type of *Silent Spring***

Figure 3 Text Type of *To Kill a Mockingbird*

4.2 The Internal Linguistic Features of Two Texts on Five Dimensions

As we mentioned above, there are salient differences between *Silent Spring* and *To Kill a Mockingbird* in all five dimensions; In order to further show the differences between the two texts, an independent sample t-test is used to compare the 67 language factors which are involved in the calculation of register dimension. The results show that there are as many as 50 linguistic features with significant differences between them. As space is limited, only the first 15 linguistic features with large differences are listed in the following table (see Table 2).

Table 2 The first fifteen linguistic features with large differences

Linguistic Features	Silent Spring	To Kill a Mockingbird	P-value	T-value	Absolute difference
That relative clauses on subject position (TSUB)	4.47	0.55	.000	15.390	3.92
Past tense (VBD)	-0.35	2.03	.000	-16.568	2.38
Public verb (PUBV)	-0.71	1.50	.000	-18.818	2.21
Prepositions (PIN)	.98	-1.17	.000	23.579	2.15
Attributive adjectives (JJ)	.93	-1.16	.000	16.861	2.09
Average Word Length (AWL)	1.02	-1.05	.000	22.303	2.07
Downtoners (DWNT)	1.75	-0.10	.000	9.046	1.85
Phrasal coordination (PHC)	2.06	0.25	.000	6.031	1.81
Analytic negation (XX0)	-.62	1.18	.000	-16.455	1.8
Direct WH-questions (WHQU)	.68	2.43	.000	-3.811	1.75
First person pronouns (FPP1)	-.80	.90	.000	-18.861	1.7
Agentless passives (PASS)	0.87	-0.79	.000	10.200	1.66
Subordinator that deletion (THATD)	-.54	1.02	.000	-13.118	1.56
Predictive modals (PRMD)	-0.64	0.89	.000	-13.969	1.53
Conjuncts (CONJ)	1.09	-0.39	.000	9.206	1.48

Combined with the previous analysis, it can be seen that the text of *Silent Spring* is characterized by more structures of that relative clauses on subject position, prepositions, attributive adjectives, long words, downtoners, phrasal coordination, agentless passives

and conjuncts. However, the text of *To Kill a Mockingbird* is featured by more structures of past tense, verbs, analytic negation, direct WH-questions, first person pronouns, subordinator that deletion and predictive modals.

5. Conclusion

In summary, the text of *Silent Spring* is quite different from the text of *To Kill a Mockingbird*. Firstly, while *To Kill a Mockingbird* is involved and interactional and context-dependent, the text of *Silent Spring* is informational, context-independent, persuasive and provides information in a technical, abstract and formal way. *Silent Spring* is classified into the register of "general narrative exposition", and To Kill a Mockingbird is classified into the register of "Imaginative narrative". Secondly, the text of *Silent Spring* is characterized by more structures of that relative clauses on subject position, prepositions, attributive adjectives, long words, downtoners, phrasal coordination, agentless passives and conjuncts. However, the text of *To Kill a Mockingbird* is featured by more structures of past tense, verbs, analytic negation, direct WH-questions, first person pronouns, subordinator that deletion and predictive modals. This paper is an attempt in this field. But there are inevitably some limitations to it. The follow-up research can consider multiple factors and try to make a more comprehensive and objective analysis of the linguistic features in novels.

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]   Abdul, G., Nadia & Yasir, H. (2022). Corpus-assisted study of symbolism in Heart of Darkness by Joseph Conrad. *Literary Voice, 18*(1), 48-58.
[2]   Biber, D. (2011). Corpus linguistics and the study of literature: Back to the future? *The Scientific Study of Literature*, 1, 15-23.
[3]   Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
[4]   Busse, B. (2020). *Speech, Writing, and Thought Presentation in 19th-Century Narrative Fiction: A Corpus-Assisted Approach.* London: Oxford University Press.
[5]   Green, G.(2017). Introducing the corpus of the canon of Western literature: a corpus for culturomics and stylistics. *Language and Literature, 26*(4), 282-299.
[6]   Hoey, M., Mahlberg, M., Stubbs, M., & Teubert, W.. (2007). *Text, Discourse and Corpora: Theory and Analysis*. Continuum.
[7]   Jaafar, E. A.. (2019). Melvin's mind style: A corpus stylistic analysis of *As Good as It Gets. Literacy and Translation, 2*(2), 29-41.
[8]   McEnery, T. & Andrew, W. (2001). *Corpus linguistics: An introduction. 2nd ed*. Edinburgh: Edinburgh University Press.
[9]   Moustafa, B. (2022). A comparative corpus stylistic analysis of thematization and characterization in Gordimer's *My Son's Story* and *Coetzee's Disgrace*. *Open Linguistics, 8*(1), 46-64.
[10] Stubbs, M. & Barth, I. (2003). Using recurrent phrases as text-type discriminators: A quantitative method and some findings. *Functions of Language, 10*(1):61-104.
[11] Stockwell, P., & Mahlberg, M. (2015). Mind-modelling with corpus stylistics in David Copperfield. *Language and Literature, 24*(2), 129–147.
[12] 陈建生,张珊.(2018).基于语料库文体学的《尤利西斯》意识流研究.外国语言文学(03),304-317.
[13] [魏利霞,周震.(2020).基于语料库的《劳拉的原型》多元主题解析.外语教学(05),104-108.
[14] 张立柱.(2022).独特的时空叙事：基于语料库的残雪小说英译文体研究.解放军外国语学院学报(04),96-104.