ISSN: 2688-8653

A Corpus-based Study on the Stylistic Features of Tess of the D'urbervilles

Yan Wu, Gengsheng Xiao* and Huimin Jiang

School of Languages and Literature, University of South China, Hengyang 421001, China

Abstract

The research aims to examine the text of Thomas Hardy's Tess of the D'urbervilles, and analyze the stylistic features of the novel with the help of corpus and its retrieval software, mainly focusing on the description of environment, psychological description, character language style and character. In terms of environmental description, the study found that this work used a large number of real place names of places of interest in England, and all the background descriptions were based on Dorchester in England. In terms of psychological description, research shows that the author generally uses positive and optimistic words to describe Tess, even though she faces various hardships and setbacks in the second half of the novel. In terms of the character's language style, the analysis shows that the language style mostly adopts the dialect at that time. In addition, in terms of character, it is found that Tess, the protagonist, has the excellent qualities of a laborer in terms of words and body movements: simple, kind, gentle, pure, filial and brave. With such statistics and data provided by corpus, the understanding of the language features, characters in Tess of the D'urbervilles can be strengthened and the research can be a supplement to the traditional literary studies of Tess of the D'urbervilles.

Keywords

Tess of the D'urbervilles; Corpus stylistics; Antconc; Keywords.

1. Introduction

With the vigorous development of corpus linguistics, the concept of corpus stylistics emerges. Corpus stylistics is an interdisciplinary subject which combines literary text analysis with literary criticism by means of corpus linguistics. Ma Guanghui, a Chinese scholar, believes that the research methods of corpus in linguistics can provide a set of effective methods and tools for literature research, enabling people to make a more detailed, in-depth and specific description of literary works.

Firstly, the corpus retrieval tool can calculate the frequency of words in literary texts and the collocation between words, as well as the degree of density of keywords (characters, things, etc.) in each chapter. Secondly, the corpus retrieval tool can also conduct rational processing analysis and statistics on the theme characters and plot development according to the different characteristics of the plot in the literary text, extract the framework from the huge text base, and then combine the text meaning to study the literary text from a new perspective. In short, whether it is lexical feature analysis, plot construction, character characterization, or writing features of literary texts, it can be clearly seen through corpus retrieval and analysis.

Therefore, this study aims to take Thomas Hardy's novel *Tess of the D 'Urbervilles* (Hardy, 2007) as an example, mainly to investigate the language style and personality of the characters in the context of psychological description, and analyze the stylistic characteristics of the novel with the help of corpus and its retrieval software.

ISSN: 2688-8653

2. Literature Review

2.1. Previous Studies on Corpus Stylistics

Corpus linguistics is a mixed subject directly connected to computer science. It collects, stores, processes and analyzes natural language contexts to guide linguistics study and to develop a natural language information processing system on the basis of the object and detailed language evidence provided by corpus (Yang, 2002). Corpus stylistics is a newly subject combining corpus linguistics with stylistics, which gives people a more object and precise statistics to analyze almost all kinds of texts' stylistics features from a scientific and totally new perspective (Zhang, 2019). There are four main branches of the research on corpus stylistics including research on literary works, on non-literary works, on translation and on teaching (Liu & Huang, 2010). During the past two decades when the rapid development of corpus linguistics, the integration of literature and corpus linguistics has also been accelerating. As for the development prospects of corporatology, Wynne points out that with the emergence of important international standards on text coding such as XML and TEI, the quality and credibility of literary text electronic libraries are also improving, so corpus linguistics will become a useful tool for stylists (Wynne, 2006). Nowadays, the majority do their studies on corpus stylistics in a rather easy and fixed model, lacking the innovative spirit. And we have few available corpus materials at hand. The corpus analysis of literary works is still in its infancy, lacking systematic research. And it confines to just some single type of research, whose scope needs to be expanded. In addition, corpus stylistics mainly focuses on novels, leaving poems and plays almost untouched.

2.2. Antconc and its Related Studies

Antconc is an integrated retrieval software invented by Laurence Anthony, a Japanese doctor in Waseda University. It is user-free and easy to operate. It has a simple interface and has all the functions that Word Smith has. Moreover, it can be operated in various systematic environments, such as Windows, Linux, and Macintosh osx, etc. The biggest advantage of Antconc is the quantitative analysis. There are seven fundamental functions of Antconc, which are the generation of frequency Word List, Concordance tool, Concordance Plot tool, the extraction of Keyword list, KWIC (Key Word in Context) tool, Collocates tool, Clusters tool and File View tool.

The first essays on corpus linguistics using Antconc as a retrieval tool in China is Xu's dissertation in 2007 (Xu, 2007). The author conducts a study on college freshmen's master of English verbs' and nouns' collocations based on the corpus principles, which reveals the freshmen's characteristics in using the verbs and nouns. And in 2019, Huang wrote a journal to explore the application of parallel corpus in the teaching of science and technology translation based on the analysis of the characteristics of science and technology English in vocabulary, syntax and discourse (Huang, 2019).

2.3. Tess of the D'urbervilles and its Related Studies

Tess of the D'urbervilles is a novel that considered as a well-known English novel in nineteenth-century by English novelist Thomas Hardy i.e. Hardy's fictional masterpiece. In literature, there are two of his novels, *Far from the Madding Crowd* and *Tess of the d'Urbervilles*, were listed in the top 50 on a BBC's survey--- "The Big Read". With Hardy's growing fame and recognition, the number of the studies of his works in literary world quickly increased, and the research perspectives became increasingly rich.

As to studies on *Tess* in China, 236 papers about *Tess* are found published in cnki.net from 1982 to 2019, ranging from eco-feminism to tragedy consciousness, narrative perspective, image

ISSN: 2688-8653

analysis, color metaphor, conceptual metaphor, contrast analysis of different translated versions, conflicting reproductive strategies and Hardy's modernist literary spirit (Liu, 2004).

In 2009, Zhang You analyzed the artistic effects in this novel and suggested the corpus-based approach is a new way to appreciate the literary works (Zhang, 2009). In 2016, Sun Juan published her journal about a contrast study on the semantic rhyme of the word "life" based on the corpus linguistics (Sun, 2016) and found that "life" mostly has a negative meaning in the *Tess*, which makes Tess's destiny more tragic.

Most of those previous scholars just made literary interpretation of *Tess of the D'urbervilles* and overlooked other perspectives like corpus stylistics. Therefore, this research will combine the corpus retrieval software Antconc3.5.8, taking corpus linguistics and corpus stylistics as the theoretical guidance, to analyze the linguistic characteristics of Hardy's representative novel *Tess of the D'urbervilles* from four kinds of words frequency and its keywords list.

2.4. Summary

Although the corpus stylistics is just a newly emerging subject, it has a wide range of application and a promising prospect. The literary research based on the corpus provides a new perspective for future research with its strict and accurate statistical method. This study attempts to reinterpret *Tess of the D'urbervilles* to support the existing literary interpretation, to explore the new text meaning, and to verify the effectiveness and possibility of the research method based on the corpus in revealing the explicit and implicit content of literary works.

3. Research Methodology

3.1. Corpus-based Approach

To study the linguistic features of *Tess of the D'urbervilles*, this research will use the corpus-based analytical method. There are two major approaches for the corpus-based analysis of language styles i.e. the corpus-based approach and case study. The corpus-based approach is a combination of quantitative and qualitative methods. For quantitative analysis, it means the use of corpus and the retrieval of data; for the qualitative analysis, it means the interpretation of the data generated by the quantitative analysis (Biber & Conrad, 2001).

The first step is to prepare the corpora and the software. The novel *Tess of the D'urbervilles* is used as the target corpus and the other six novels written by Hardy, *Jude the Obscure, the Mayor of Casterbridge, Far from the Madding Crowd, Under the Greenwood Tree, the Return of the Native, and the Woodlanders* as the referential corpora. The establishment of referential corpora aims to highlight the prominent stylistic features by comparing the target corpus to reference corpora. Antconc3.5.8 is used as corpus software.

The next step is the procedure of data collection, which means using Antconc to generate word list and keyword list. And the stylistic features of *Tess of the D'urbervilles* will be fully interpreted in details relying on the generated data. Wang proposed that a stylistic analysis of language covers three levels — "description, interpretation and evaluation", and "the three levels are logically ordered" (Wang, 2000:76).

The third step is to interpret those high-frequency words and the keywords in the text, followed by the conclusion part drawn from the analysis. To sum up, in this research, quantitative and qualitative approaches are of vital importance for the author to provide new and creative interpretations after generation of the data.

3.2. Case Study

Case studies are also known as case investigations. It is a study of a particular individual, unit, phenomenon, or subject. Such studies collect extensively the related information to learn more details about and analyze the generation and development processes of the subjects, about the

ISSN: 2688-8653

internal and external factors and their interrelationships. In case study, research can form in-depth and comprehensive understanding and make conclusions on the relevant issues. In this research, the author takes Hardy's novel *Tess of the D'urberviles* as the study example and uses corpus-based approach to analyze the stylistic features of the novel from the Word List and Keyword List. In the analyzing process, the author has read others' literatures on corpus stylistics and the writing style of Hardy and formed her own understandings on the language features, plot development, writing style of the novel and personalities of the heroine.

4. Data-retrieval and Interpretation of Results

4.1. Type/Token Ratio (TTR)

This part aims to inspire common imagination on making fresh explanations of the language characteristics of *Tess of the D'urbervilles* and to reveal study findings on the basis of retrieved statistics. Token refers to the individual words that can be calculated repeatedly while type refers to the distinct word form which means different morphological tokens with the same meaning are counted as the same type (Liang, Li & Xu, 2010). TTR is a common standard to evaluate the diversity and richness of a text's words. The higher the TTR is, the more variety of words it will be. It is easy to tell that the shorter sentence has a higher TTR than the second one in that there are more Types in it. In this thesis, the Types and Tokens of both target corpus and referential corpora have been shown in the following table (see Table 1):

Table 1. TTR of target corpus and referential corpora

Novels	Types	Tokens	TTR
Tess of the D'urberviles	12138	156129	7.7
Far From the Madding Crowd	11762	143652	8.1
The Mayor of Casterbridge	10266	122538	8.3
Under the Greenwood Tree	6867	63115	10.0
Jude the Obscure	10866	152119	7.1
The Woodlander	10609	142029	7.4
The Return of the Native	10441	148469	7.0

As can be seen from the table, the other six novels written by Hardy have almost the same TTR as *Tess of the D'urberviles.* Those TTRs are just slightly lower or above that of the target corpus, suggesting that the author of this study who takes them as the referential corpora is reasonable and appropriate in that their word richness and language expressions are nearly at the same level.

4.2. Word List and High-frequency Words

To get knowledge of the complete stylistic features of *Tess of the D'urberviles*, the word list and keyword list need to be generated. The author collects the top 50 words by their frequency which are listed in the table below. (see Table 2)

ISSN: 2688-8653

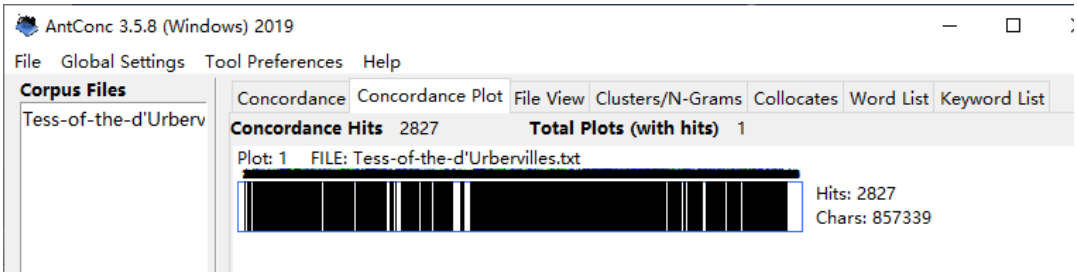
Table 2. Word list of *Tess of the D'urberviles*

Word List Results 2 — Word Types: 12138

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	8918	the	20	1050	with	39	549	no
2	4486	of	21	1042	at	40	525	me
3	4460	and	22	911	s	41	519	there
4	4199	to	23	909	on	42	518	t
5	3291	a	24	902	tess	43	514	been
6	2827	her	25	889	but	44	510	if
7	2526	in	26	878	they	45	508	would
8	2232	she	27	828	be	46	504	all
9	2127	was	28	807	by	47	491	one
10	1929	i	29	776	which	48	488	or
11	1895	that	30	718	this	49	468	could
12	1835	he	31	714	from	50	438	when
13	1780	had	32	713	have			
14	1743	it	33	627	him			
15	1613	you	34	619	so			
16	1593	as	35	605	their			
17	1260	his	36	605	were			
18	1250	not	37	598	said			
19	1083	for	38	586	is			

Among these top 50 high-frequency words, there are 16 pronouns, 10 prepositions and 9 verbs. And personal pronouns and possessive pronouns occupy the largest part of pronouns. The pronoun "her" is the highest frequency words which appears in the novel 2827 times, which, in Li's words, is the detached authorial narrative to make the story more tragic (Li, 2008). Then the author uses the Concordance Plot to analyze the distribution of "her" in the text, as seen from the bar code (the black part signifies the search word's distribution), and finds that the word "her" runs through the whole novel. (see Table 3)

Table 3. Concordance plot of "her" in the novel



Furthermore, verbs like "was" "had" "be" "were" "said" are the most frequent in the text and "was" is the most frequent one which appears 2127 times, which means the target corpus *Tess of the D'urberviles* mainly applies the past tense to tell the story.

In short, from the word list above, it is easy to tell that Hardy wrote many dialogues to develop the novel's plots. And the high frequency of pronouns signifies the core of the novel is the relationships between humans. But in the word list, there are no nouns, adjectives but only one notional verb. That's why in the next part the author will analyze those parts of speech accordingly. The author will retrieve the top 18 words respectively and then will try to interpret them objectively basing on the data collected.

Following is the main results obtained in the study: (1) A large part of these words is related to mental activities like "like" "know" "think" "love", which means the writer uses many psychological descriptions; And verbs like "said" "asked" "tell" are signs of oral words in daily communication, which testified that there are many conversations happened between the main figures; One point needs to highlight, that's to say, the verb "work" has a higher frequency than

ISSN: 2688-8653

"love", from which readers could get the great feature of Tess that she is always hard-working and diligent towards her jobs; (2) The time noun "day" appears more frequent than "night", implying most of the events in the novel happened during the day; And the body nouns "face", "eyes" and "hand" repeated for many times, combining with Antconc's concordance function, it is obvious that most of those body nouns are used to modify Tess. In *Tess of the D'uurberviles*, Tess is a very beautiful, sincere and innocent woman. That also means Hardy uses many descriptions on human's eyes to show their personalities and psychological activities; On the other hand, "man" appeared 241 times while "woman" 164 times. It seems kind of contradictory to Hardy's theme of the novel at the first thought, but it is the evidence to his ideology of patriarchal fatherhood deep in his mind; (3) Many adjectives are of negative mood, like "little", "old", "long", "poor", adding a sad atmosphere on the novel, which is known as a typical tragedy. And the other adjectives to modify good quality like "good", "young", "white", "new", signify Tess's vigorousness and kindness. The word "white" is the symbol of Tess's innocence and purity, and "better" and "new" show Tess's optimism towards her own bad experiences.

4.3. Keyword List

This research will analyze respectively the relationships between those keywords and the environmental description, psychological description, the corpus language style and the characteristics of the main novel character. Interesting though, the first five keywords are related to the three main characters in the novel---Tess of the D'urbervile, Angel Clare and Alec Durbeyfield. That vividly shows all the plots are developed around those three people.

With a further analysis on the File View, it's not hard to find that Hardy likes using color words and passive adjectives to describe environment. Moreover, it is not just description on environment but also bears a strong connection to Tess's feelings.

There are also a lot of other psychological descriptions. As a whole, Hardy is a master of psychology who is good at describing main character's mental activities by drawing conflicts between them.

There are many spoken words and dialects in these picked out conversations happened between the Durberfields. More importantly, most of them are dialogues between Tess's parents, who have received little education at that time. That's just another feature of this novel, that is, the usage of local dialects when describing the conversation, making the people more real and making the novel more of local characteristics.

5. Conclusion

This research discussed the stylistic features of *Tess of the D'urbervilles based on corpus*. Through the analysis of the target corpus's words list and high-frequency notional verbs, nouns and adjectives, many dialogues and Tess's purity, diligence and kindness are vividly shown to readers.

At the same time, the analysis of keywords list at the comparison of the other six reference corpora of Thomas Hardy, the relationships between those keywords and environmental description, psychological description, the conversation style and the characteristics of the three main characters are revealed respectively. As to the portrait on the three main characters, Hardy brings the readers a pure Tess, a traditional and conservative Angel Clare and a wicked Alec D'urberville.

In addition, some limitations and suggestions for future studies in this field will also be put forward. First of all, all of the high-frequency words obtained haven't been studied in depth. Other parts of speech haven't been studied except verbs, nouns and adjectives. Secondly, this research just takes *Tess of the D'urberviles* as the target corpus to analyze the overall writing style of Thomas Hardy, without taking Hardy's other masterpieces into consideration. Thirdly,

ISSN: 2688-8653

though the claims are made on the basis of objective corpus data, more studies in this field need to be concluded in order to further confirm the research findings demonstrated here.

Acknowledgements

This work was supported by grants from the Hunan Provincial Foundation for Philosophy and Social Sciences (No.18WLH33), the Hunan Provincial Situation and Decision-making Consultation Research Project (No.2015BZZ046), and the Foundation for Philosophy and Social Sciences of University of South China (No.2017XGY05).

References

[1] Biber, D. & Conrad, S. Corpus-Based Research in TESOL: Quantitative Corpus-Based Research: Much More Than Bean Counting[J]. TESOL Quarterly, 2001, 23(2): 331–336.

[2] Hardy, T. Tess of the D'Urbervilles[M]. New York: Signet Classics, 2007.

[3] Wynne, M. Stylistics: Corpus Approaches[J]. Literary and Linguistic Computing, 2006(7): 546-549.

[4] Huang, K. L. The application of bilingual Parallel Corpus in Science and Technology Translation Teaching[J]. Overseas English, 2019(13): 156–157.

[5] Li, X. H. On Hardy's language art from Tess of the D 'Urbervilles [J]. The Science Education Article Cultures, 2008(03): 161.

[6] Liang, M. C., Li, W. Z. & Xu, J. J. Using Corpora: A Practical Coursebook[M]. Beijing: Foreign Language Teaching and Research Press,2010.

[7] Liu, J. & Huang, L. B. Corpus Stylistic introduction[J]. Foreign language teaching and Research, 2010, 42(3): 236–239.

[8] Liu, M. S. A Review of Hardy's Novel Research in China in recent 20 years[J]. Foreign Literature Studies, 2004, (6):147-151.

[9] Sun, J. A Comparative Study on the Semantic rhyme of "Life" based on corpus -- A Case study of the novel Tess of the D 'Urbervilles [J]. Modern Chinese (Language Studies Edition), 2016(11): 147–151.

[10] Wang, S. Y. An overview of English stylistics[M]. Jinan: Shandong University Press, 2000.

[11] Xu, J. A corpus-based study on verb/noun collocation for first-year college students[D]. Beijing: Master's thesis, Beijing University of Posts and Telecommunications, 2007.

[12] Yang, H. Z. & Wei, N. X. An Introduction to Corpus Linguistics[M]. Shanghai: Shanghai Foreign Language Education Press, 2002.

[13] Zhang, L. The achievements and development of Corpus Stylistics applied research in China[J]. English Teachers, 2019(03): 6–8.

[14] Zhang, Y. A New Mode of Artistic Effect Appreciation in Tess of the D 'Urbervilles -- Text Analysis based on corpus Retrieval Technology[J]. Novel Review, 2009(S2): 160–162.