A Measured Stylistic Study of the Lexical Styles of Yu Hua's and Mo Yan's Long Stories

Haoxian Bai

Central China Normal University, Wuhan; China

E-mail: baihaoxian00@163.com

Abstract. This paper takes the twelve long novels of Yu Hua and Mo Yan as the research objects, and analyses the differences in the writing styles of the two writers from the perspectives of vocabulary length, vocabulary richness, special word classes and special lexical items, and so on. It is learnt that Mo Yan's works are more complex in language, richer in vocabulary, broader in text creation, and more intuitive and strong in emotional expression, while Yu Hua's works are simpler and easier to understand, less rich in vocabulary, less broad in text creation, and more objective and implicit in emotional expression. In addition, Mo Yan's style of writing is more stable than Yu Hua's.

Keywords: Yu Hua, Mo Yan, Measurement Stylistics, Vocabulary Style.

1 Introduction

Yu Hua (b. 1960), after failing the university entrance examination in 1978, joined a health centre as a dentist, and persisted in writing in his spare time, completing his first short story, The First Dormitory, in 1983. He went on to write a series of outstanding full-length novels, as well as a large number of essays, miscellaneous essays, and short and medium-sized novels. Mo Yan (b. 1955), completed his debut novel Falling Rain on a Spring Night in 1981, and has since constructed a diverse and rich literary world, including full-length works such as The Red Sorghum Family, which was adapted into a film of the same name and won the Golden Bear, and The Frogs, which won the Mao Dun Literature Prize, as well as dozens of short and medium-sized novels.

As the two writers are of similar age and their works have similar themes, many comparative studies of their works have been done. For example, Xu Wenming (2008) compares the differences in the characteristics of the narrative focus of "violent narration" that the two authors have worked so hard on[1]. Of course, there are also comparisons of language styles, such as Wu Lili (2014), who compared and analysed the "differences" in language styles and the "similarities" in the creative tone and thinking characteristics of some of their children's perspective novels[2]. However, scholars' descriptions of language styles are either subjective or independent. Until Tu Mengchun and Liu Ying (2019) analysed the difference between Yu Hua's and Mo Yan's language styles by using five of their novels as the corpus through the method of measurement

S. Zhu et al. (eds.), *Proceedings of the 3rd International Conference on Education, Language and Art (ICELA 2023)*, Advances in Social Science, Education and Humanities Research 831,
https://doi.org/10.2991/978-2-38476-214-9_91

and statistics[3]. However, due to the time reason it did not involve Yu Hua's The City of Letters. Therefore, the corpus can be expanded.

2 Selection and Pre-Processing of the Corpus

Vocabulary is the building material of language, and a comparative study of the vocabulary styles of the two authors' novels using the method of metrical stylistics is of vital importance for us to analyse the differences between the languages of their novels. In this paper, all six of Yu Hua's long novels as well as Mo Yan's long novels with similar corresponding publication years are selected as the corpus, and then they are cleaned into the form of one sentence and one line by using the Python programme, so as to carry out the subsequent segmentation and analysis.

3 Measured Stylistic Comparisons of Vocabulary

3.1 Average Length of Vocabulary

In metrical stylistic studies of vocabulary, average vocabulary length is a key indicator that helps us to see differences in the ease of comprehension of writers' fictional works. Generally speaking, average word length and ease of comprehension are inversely proportional. That is, the larger the average word length of a novel is, the less easy the novel is to be understood by readers, and vice versa. After using the Jieba lexical package for cleaning documents to obtain the total number of words of the work, the total number of words and the total number of words to do the quotient, you can find the average word length of the twelve corpus novels (the results are retained to three decimal places, rounded up), see Figure 1 below:

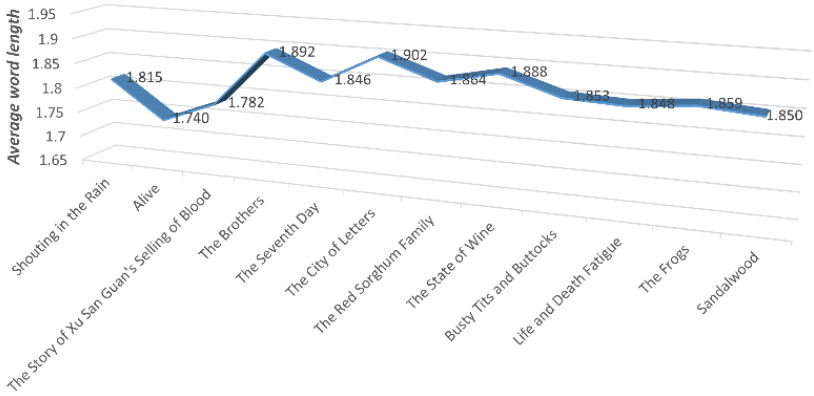

Fig. 1. Average word length of Yu Hua and Mo Yan novels

On the whole, the average length of vocabulary in these twelve works tends to be close to 2, i.e. two-syllable words. Therefore, all of them are relatively easy to understand, and basically do not cause any obstacles to the readers in terms of reading

comprehension. This is also related to their literary genre. After all, novels are centred on character portrayal, reflecting social life through complete storylines and environmental descriptions, focusing on narrative, which is different from essays, poems and dramas.

In contrast, the average length of Mo Yan's vocabulary is higher than that of Yu Hua's, indicating that the language of his novels is more complex, while Yu Hua's novels are a bit more accessible in comparison. Through vertical comparison, we can also find that Yu Hua's novels in different periods are quite different, with the average word length of his later works significantly higher than that of his earlier works, which also objectively reflects the subtle changes in his writing style; while the average word length of Mo Yan's six novels is more stable, close to the average value of 1.860, which indicates that his word habits have not changed much.

3.2 Vocabulary Richness

3.2.1 Comparison of Type-Token Ratio

Type-Token Ratio (TTR, Type-Token Ratio) is a ratio representing the level of lexical richness of a text (proportional) between word types (Type) and word examples (Token). However, over time, scholars have identified some shortcomings of the Type-Token Ratio (TTR), such as problems in measuring the ratio due to differences in text length. Therefore, Baker pointed out that it can be normalised by every 10,000 formants, and the average value can be obtained after calculating the TTR separately[4]. He called this average value Standard Type-Token Ration (STTR), and its calculation process can be expressed as follows:

$$\text{STTR} = \frac{Type}{Token} \times 10000 \tag{1}$$

Ure J. (1971) and Stubbs M. (1986) proposed an alternative method of calculating only the ratio of real morphemes to the total morpheme ratio[5] [6]. Here, we take the first approach. After the computational processing of the twelve novel texts using the above formula, we arrive at their standard type-example ratios as shown in Figure 2:

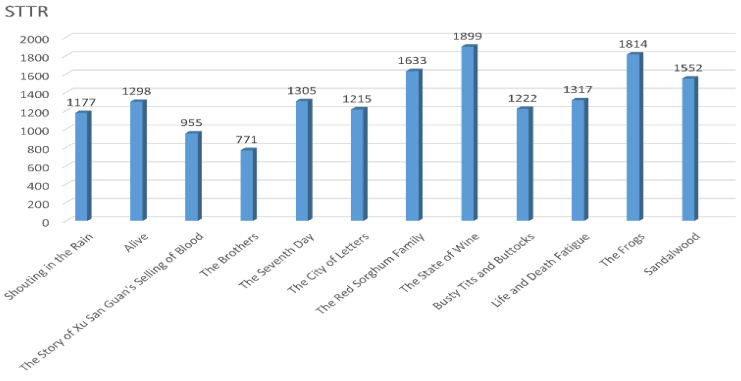

Fig. 2. Comparison of standard type examples of Yu Hua's and Mo Yan's novels

Overall, it seems that Yu Hua's The Brothers has an STTR of 771 and Mo Yan's The State of Wine has an STTR of 1,899, which are at the end and the top of the dataset, respectively, and represent the two extremes of the level of vocabulary richness. In addition, the mean STTR of Mo Yan's six novels is significantly higher than that of Yu Hua, indicating that the former uses a greater level of vocabulary richness in its creation than the latter.

3.2.2 Comparison of Single-Occurrence Words

Single-occurrence words refer to words that occur only once, and they are also an important indicator of the lexical richness of a text. In a text with a certain volume (normalised by 10,000), the larger the proportion of single-occurring words, the fewer words are repeated in the text, and the higher its lexical richness. The single-occurring word percentages of the twelve corpus novels are now presented in the form of radar charts:

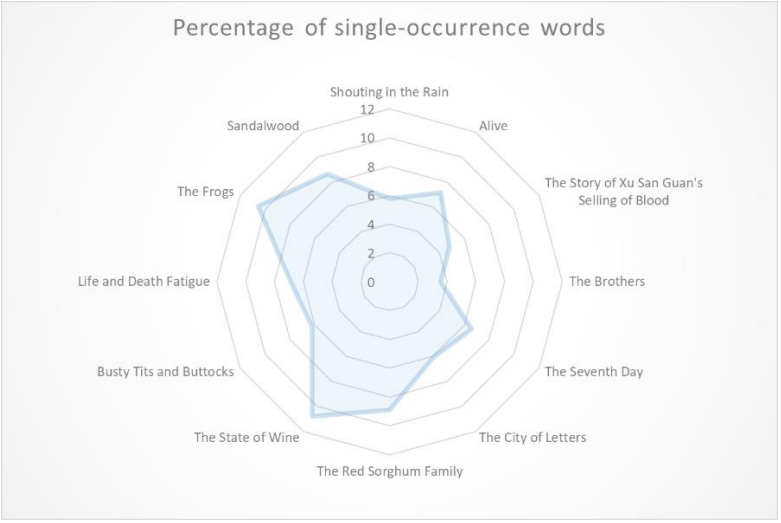

Fig. 3. Radar chart of the proportion of single-occurrence words in Yu Hua's and Mo Yan's novels

The radar chart clearly shows that the proportion of single-occurrence words in Mo Yan's novels (lower left half) is significantly higher than that in Yu Hua's novels (upper right half), which is also in line with the data comparison of the standard example ratio above, and once again demonstrates that Mo Yan's novels have a higher degree of lexical richness. Specifically, Mo Yan's Wine Country and Frogs occupy the first and second places in the ranking of the proportion of single-occurrence words, which should be related to their more novel themes. Parallel to the main story in Jiu Guo are the author's correspondence with Li Yidou, a young literary man in Jiu Guo, about his literary creation, and Li Yidou's novels, which are stylistically diverse and can be called a "full-course meal" of novel genres. Frogs, on the other hand, consists of four long letters and a nine-act play, a very unique structure.

3.3 Special Word Classes

In the field of natural language processing and text analysis, the number of place nouns can often reflect the breadth of a writer's text creation to a certain extent. The top ten location nouns in the frequency of occurrence in the respective novels of the two writers are now counted, and the results are shown in the following table:

Table 1. Top ten place names in the frequency of Yu Hua's and Mo Yan's novels

writer	Top ten place nouns in terms of frequency
Yu Hua	Creek Town, Shanghai, Suzhou and Hangzhou, Welfare Factory, Japan, China, Chenghuangge, Beijing, Xili Village, Beijing
Mo Yan	Gaomi Northeast Township, Japan, Germany, China, Xingyuan, Qinhe, Beijing, USA, Jiaogao, Shazhou

From the statistics of table 1, we can see that the high-frequency location nouns shared by both are: Japan, China, and Beijing. This is related to the background of the stories in the works, most of which take place during the Anti-Japanese War, and also reflects some of their commonalities as folk novels. In addition, the high-frequency foreign place names in Mo Yan's works include "Germany, the United States", etc., while Yu Hua's high-frequency place names are more domestic, which indicates that the breadth of Mo Yan's textual creation is greater than that of Yu Hua.

3.4 Special Lexical Items

Overlapping words refer to the combination of the same words into a new word, in the expression of the effect can usually be used to express a stronger tone, richer feelings or to increase the rhythm of the text, grammatical effect can indicate the degree of high and low or the size of the measure and other meanings. Overlapping words have many forms. The main forms are:AA, AAB, ABB, AABB, AABC, ABAC, ABCC. Now we make a statistical comparison of the use of overlapping words in Yu Hua's and Mo Yan's novels (the results are shown in the table below) in order to analyse the similarities and differences of the two in the expression of emotions.

Table 2. Overlapping words in Yu Hua and Mo Yan's novels

overlapping form	Yu Hua	Mo Yan
AA	4,714	10,808
AABB	103	100
AABC	3	7
ABAC	3	8
percentage	**8.61 per cent**	**11.92 per cent**

From the statistical results in the above table 2, it can be seen that the proportion of overlapping words used in Mo Yan's novels is higher than that of Yu Hua. This also indicates that Mo Yan's emotional expression in his novels is more intuitive and strong.

On the other hand, the extensive use of overlapping words can also make the novel catchy and have a strong sense of rhythm.

4 Conclusions

After taking the twelve long novels of Yu Hua and Mo Yan as the corpus, and conducting a comparative analysis of their vocabulary styles in terms of vocabulary length, vocabulary richness as well as special lexical categories and special lexical items by means of metrical stylistics, it can be concluded that Mo Yan's vocabulary has a larger average length, his novels' language is more complex, and his vocabulary richness as well as the breadth of the textual creation are higher than Yu Hua's. This is evidenced by the Type-Token Ratio as well as the comparative metrics for the Single-occurrence words. Moreover, the expression of emotions in Mo Yan's works is more Mo Yan's works are more intuitive and strong, and the textual language has a stronger sense of rhythm. In addition, Yu Hua's writing style is more variable and unstable than that of Mo Yan. Due to the limitation of space, this paper only compares and analyses a few key indicators in vocabulary, and the research on the measurement style of other aspects, such as sentence paragraphs, chapters and emotions, will be the next research direction.

References

1. Wenming Xu (2008) Landscape of Death, Master's thesis, Henan University.
2. Lili Wu (2014) "Children's Perspectives in the Works of Contemporary Male Writers," Master's thesis, Anhui University.
3. Mengchun Tu, Ying Liu (2019) Econometric Statistics and Analyses of Yu Hua's and Mo Yan's Long Fiction, Journal of Chinese Information, Vol. 2:132-137.
4. Baker M. Towards a methodology for investigating the style of a literary translator[J]. Target, 2000, 12 (2) :241-266.
5. Ure Jean. (1971) Lexical density and register differentiation[A]. Applications of linguistics: selected papers of the Second International Congress of Applied Linguistics[C], eds. By G.E. Perren and J.L.M Trim, Cambridge: Cambridge University Press.443-452.
6. Stubbs M. (1986) Language development, lexical competence and nuclear vocabulary[A]. In K. Durkin(ed). Language Development in the School Years[C], Croom Helm.13.