Lecture Notes on Language and Literature (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/langl.2025.080504
ISSN 2523-5869 Vol. 8 Num. 5

Exploration of Readability of Shuang Xuetao's Works in English Taking Rouge Street Translated by Jeremy Tiang as an Example

Junzhe Zhao

School of Foreign Studies, Liaoning University of International Business and Economics, Dalian, 116052, Liaoning, China

Keywords: Novel Northeastern Literature, Shuang Xuetao, Readability Analysis

Abstract: In recent years, the Northeast China has continued to produce high-quality works, ranging from short plays and drama to literature and art. Creators from there have continuously surprised people with their creativity, creating a cultural boom. Among them, Shuang Xuetao's works have been published in multiple languages. This article uses the readability analysis tool Readability Analyzer 2.0 to explore the readability of the translation by Jeremy Tiang's work—*Rouge Street*, assesses the language proficiency required by readers, and provides experience for the overseas dissemination of new Northeastern literature, helping to revitalize the Northeast.

1. Introduction

In recent years, Northeastern cultural and artistic works, from sketches and dramas to music and ballet, and now to today's "death rock" and the "short video," have filled the daily lives of the people. It not only demonstrates the prosperity of Northeastern culture and art but also enlivens the literary and artistic atmosphere of the Northeast, providing solid foundation for the rise of "New Northeast Literature."

The writers who were born during 1970s-1990s in Northeast China are beginning to make their mark, producing literary works with new qualities and different from the traditional realism of their predecessors. These works offer fresh perspectives on life and life, history and reality, providing a rich textual foundation for the formation of "New Northeast Literature." With the rise of writers like the "Tiexi Trinity" and extensive attention from the academic community, "New Northeast Literature" has gradually emerged from the cocoon of the "Northeast Renaissance" to become an independent topic and literary phenomenon. "New Northeast Literature" not only includes the writer group and works represented by the "Tiexi Trinity" but also encompasses a broader range of writers and more abundant literary texts. Its creative methods and themes also show multidimensional and heterogeneous characteristics, breaking the original binary pattern of rural literature and industrial narrative under the dominance of traditional realistic creation methods.

The overseas dissemination of New Northeast Literature not only helps to enhance the understanding and recognition of Chinese culture among overseas readers but also promotes exchanges and integration between different cultures. The overseas dissemination of New Northeast

Literature is bound to further stimulate Northeast revitalization and attract more attention.

2. Text Readability

Text readability is also known as readability, defined as "the ease or difficulty of understanding or comprehending caused by the style of writing." Text readability also refers to the text elements that affect readers' understanding, reading fluency, and interest levels.[1] These text components involve several dimensions of text attributes, including lexical, syntactic, and conceptual dimensions[2], which also affect text readability.

In the 1950s, readability research began to emerge. Rudolf Flesch (1948) proposed the Flesch readability formula, emphasizing the impact of sentence length and lexical difficulty on text readability. This formula became the basis for subsequent research. In the 1960s, G. C. L. Harris (1965) further developed the Flesch formula, proposed the concept of readability scores, and explored their application in education and translation. This period of research focused mainly on the establishment of mathematical models and formulas. In the 1980s, readability research began to pay attention to readers' reactions. Kintsch and Van Dijk (1978) proposed the construction-integration model, emphasizing the importance of readers' background knowledge in the process of understanding. This theory provided a psychological perspective for subsequent readability research[3]. In the 1990s, with the development of computer technology, automated readability scoring tools gradually appeared. McLaughlin (1969) proposed the SMOG index (Simple, Meaningful, Organized, and Grammatical language), which was widely applied to the evaluation of different types of texts.[4] At this time, researchers began to pay attention to the readability of different types of texts such as legal, medical texts. At the beginning of the 21st century, readability research was gradually applied to practical translation and education. Baker (2006) pointed out that readability in translation not only affects the effectiveness of text communication but also relates to the acceptability of readers. She emphasized the importance of considering the target readers in the translation process.[5] In addition, M. A. C. De Beaugrande (1997) proposed the text linguistic theory, which provided a new perspective for readability research, emphasizing the role of text structure and language function in readability (De Beaugrande, 1997).[6]

In recent years, with the development of big data and artificial intelligence, readability research has entered a new stage. Researchers have begun to use machine learning and NLP technologies to analyze the readability of texts. At the same time, the relationship between readability and multimodal texts has also been of concern. Kress and van Leeuwen (2006) proposed that the design of multimodal texts should consider the readability of different media, which provided a new research direction for visual elements in translation technology.[7]

This study will utilize mainstream formulas, including the following 6 indicators: 1) Flesch Reading Ease 2) Flesch-Kincaid Grade Level 3) Automated Readability Index 4) Coleman-Liau Readability Score 5) Gunning Fog 6) SMOG.

2.1 Flesch Reading Ease

Flesch Reading Ease is a measure of text readability, with a numerical range from 0 to 100. The higher the number, the easier the text is to read. This formula takes into account sentence length and syllable count, is typically applicable to English, and is currently the most widely used readability indicator. The formula is as follows:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right) \quad (1)$$

The difficulty-ease table is as follows Table 1:

Table 2 Flesch Reading Ease Difficulty Table

Score	Level
0-30	Very Difficult
30-50	Difficult
50-60	Harder
60-70	Hard
70-80	Less Easy
80-90	Easy
90-100	Very Easy

2.2 Flesch-Kincaid Grade Level

The Flesch-Kincaid Grade Level is a metric that corresponds the readability of text to the American school grade system, with higher scores indicating a higher author's level and also implying higher reader's level and requirements. The scoring formula calculates the average sentence length and the average word length per sentence of the text, thus determining a grade level, which is divided into 12 levels, roughly corresponding to the reading levels of American primary and secondary school students. For example, a score of 8.0 indicates that the text is suitable for 8th-grade students to read. The formula is as follows:

$$0.39\left(\frac{\text{total words}}{\text{total sentences}}\right)+11.8\left(\frac{\text{total syllables}}{\text{total words}}\right)-15.59 \tag{2}$$

2.3 Automated Readability Index

The Automated Readability Index (ARI) is a measure of text readability calculated by considering the number of characters, words, and sentences. The result is a number indicating the grade level the text is suitable for. This index is particularly sensitive to short texts and is often used in conjunction with other readability tests in practical applications. The formula is as follows:

$$4.71\left(\frac{\text{characters}}{\text{words}}\right)+0.5\left(\frac{\text{words}}{\text{sentences}}\right)-21.43 \tag{3}$$

2.4 Coleman-Liau Readability Score

The Coleman-Liau Readability Score primarily evaluates readability based on the number of words and sentences in the text, designed by Meri Coleman and T. L. Liau. Unlike other readability formulas, it uses the number of letters instead of syllables, making it easier to calculate. The score also indicates the appropriate grade level and is suitable for quickly assessing the difficulty of the text. The formula is as follows:

$$CLI = 0.0588L - 0.296S - 15.8$$

2.5 Gunning Fog Index

The Gunning Fog Index is a readability measure that assesses the complexity of text by calculating the proportion of complex words and the average sentence length. A higher score indicates a more complex text, typically suitable for students in higher grades, and the number roughly reflects the number of years of formal education required to understand the text. It is

important to note that the Gunning Fog index is based on the assumption of starting school at the age of 6, for example, a score of 7 means that a child has received formal school education for 7 years, and a 13-year-old student can understand the text. The formula is as follows:

$$0.4\left[\left(\frac{\text{words}}{\text{sentences}}\right)+100\left(\frac{\text{complex words}}{\text{words}}\right)\right] \tag{4}$$

2.6 SMOG

The SMOG is a readability test primarily used to assess the understandability of public texts. The number reflects a similarity to the Gunning Fog index, with the score representing the required years of education. It calculates the score by counting the number of polysyllabic words in the text, and is usually applicable to documents that require broad understanding. The formula is as follows:

$$\text{grade}=1.0430\sqrt{\text{number of polysyllables}\times\frac{30}{\text{number of sentences}}}+3.1291 \tag{5}$$

The above are the mainstream readability calculation formulas and their basic meanings. In this study, the Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning Fog Index, SMOG Index, and Automated Readability Index will be used to observe and judge the consistency of the readability of three sub-sections (three independent works by Shuang Xuetao) in the book *Rouge Street*, namely, *The Aeronaut*, *Bright Hall*, and *Moses on The Plain*.

3. Research Content and Research Approach

3.1 Research Questions

Readability can measure the difficulty of understanding text. This paper aims to explore the following two questions: 1). *Rouge Street* within different chapters, does readability maintain basic consistency? 2). How do the influencing factors and characteristics of readability differ within the chapter?

3.2 Research Object

Henry Holt & Company in the United States has combined Shuang Xuetao's three works, *Moses on the Plains*, *Bright Hall*, and *The Aeronaut* into a collection of short stories, translated by Jeremy Tiang, titled *Rouge Street* . This study takes the Jeremy Tiang translation as the research object and conducts readability analysis on the text level.

4. Research Methods and Process

4.1 Research Methods

This study adopts a quantitative research approach. Readability is measured through formulas. Readability is not as undesirable as some scholars believe. As long as it is reasonable and effectively controlled to determine the social, cultural variables of the readers and the variables of the language itself, and as long as the formula is based on a scientific theoretical foundation, it is completely possible to measure the readability of English through a formula (Lin Zheng, 1995). Therefore, this study uses readability analysis tools to calculate the readability of the translation and to analyze and interpret the readability values of different calculation formulas is feasible; this

article uses the BFSU Hugemind Readability Analyzer 2.0 to compare the readability of Rouge Street from multiple perspectives and indicators; and uses SPSSAU to test the level differences of words and sentences within different chapters.

4.2 Research Process

The research first conducts corpus cleaning on the collected translated texts, divides the articles into three parts based on titles, and places each part into the BFSU Hugemind Readability Analyzer 2.0, a readability analysis tool, to analyze the readability of the texts separately. The readability level is evaluated from multiple numerical indicators such as word length, sentence length, Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning Fog Index, SMOG Index, and Automated Readability Index. The minimum language proficiency or years of formal education required for different parts of the text is analyzed, and a one-sample t-test is conducted using SPSSAU to compare horizontally whether there are differences in word length, sentence length, and readability level between different parts.

4.3 Data Statistics

4.3.1 Word Length

It's needed to input the three components of Rouge Street, namely The Aeronaut, Bright Hall, and Moses on The Plain, into Readability Analyzer 2.0 to obtain the total word count, word length, and difficult word ratio. The specific data is shown in Table 3 Word Length Statistics:

Table 3 Word Length Statistics

Text	Total Word Count	Word Length (characters per word)	Proportion of Difficult Words
The Aeronaut	21909	4.21	6.13%
Bright Hall	22347	4.14	4.94%
Moses on The Plain	22530	4.1	6.28%

The total word count for the three parts is 21,909, 22,347, and 22,530 words, respectively. The average word length, which is the number of letters in each word, is 4.21, 4.14, and 4.1, respectively. The analysis of difficult word ratios is 6.13%, 4.94%, and 6.28%, respectively. Due to the small sample size and the lack of a control group, the study used SPSS-AU for a one-sample t-test, with the median of each group as the default reference value. The p-values for the three groups are 0.690, 0.785, and 0.499, respectively. The conclusion values are all greater than 0.005, indicating no significant difference. This further demonstrates that the total word count, word length, and difficult word ratio of the three parts in this translation are at the same level, which is conducive to text reading.

4.3.2 Sentence Length

The specific data of the total number of sentences and average word count for the three parts in Readability Analyzer 2.0 are shown in Table 3:

Table 3 Sentence Length Statistics Table

Text	Total Number of Sentences	Average Word Count
The Aeronaut	1434	13.99
Bright Hall	1973	10.31
Moses on The Plain	1646	12.38

The total number of sentences in the three parts are 1434, 1973, and 1646, respectively, with

average word counts of 13.99, 10.31, and 12.38. A single sample t-test was conducted using SPSSAU, with t-values of 0.245 and -0.144, and p-values of 0.830 and 0.899, respectively. Generally, the closer the t-value is to 0, the less significant the difference, and when p-values are greater than 0.05, the results indicate that there is no significant difference between the sample mean and the assumed mean.

4.3.3 Readability Index

In addition to the statistics of word length and sentence length mentioned above, this section will focus on the core data module of this article, namely the text readability index. In order to make the text readability data more comprehensive and reliable, this study selects five mainstream application indicators to display the readability of the text, avoiding the occasions caused by single data. The specific values of each indicator are shown in Tables 4 and 5:

Table 4 Readability Index Statistics I

Text	Flesch Reading Ease	Flesch-Kincaid Grade Level	Gunning Fog Index
The Aeronaut	81.93	5.50	6.85
Bright Hall	94.15	2.90	5.03
Moses on The Plain	83.15	5.00	6.26

Table 5 Readability Index Statistics II

Text	SMOG Index	Automated Readability Index
The Aeronaut	7.90	7.00
Bright Hall	6.80	4.90
Moses on The Plain	7.80	5.80

From the above data, the overall differences are not significant. Further inspection of the sample differences by SPSSAU shows that the t-values for the 5 indicators are 0.839, -0.670, -0.398, -0.854, and 0.164, respectively, and the p-values are 0.490, 0.572, 0.729, 0.483, and 0.885, respectively. The absolute values of the t-values for the 5 groups of data are close to 0, and the p-values are all greater than 0.05, indicating no significant differences. This suggests that there is no significant difference in readability between different parts, and they are generally at the same readability level. From the perspective of individual indicators, specific age stages or readability levels still show differences. The specific readability levels of the three parts are shown in the table 6:

Table 6 Readability Results of Different Parts

Text	Flesch Reading Ease	Flesch-Kincaid Grade Level	Gunning Fog Index	SMOG Index	Automated Readability Index
The Aeronaut	Easy	Grade 5.50	12.85 years old	6.85 years	Grade 7
Bright Hall	Very Easy	Grade 2.90	11.03 years old	5.03 years	Grade 4.90
Moses on The Plain	Easy	Grade 5	12.26 years old	6.26 years	Grade 5.80

Overall, *Rouge Street* requires the highest reading level and ability, followed by *Moses on The Plain*, and the lowest is *Bright Hall*. Readers over 13 can understand the text, suitable for grades 5-7 and above or readers with an education duration of nearly 7 years. Readers under 13 or with an education duration of no more than 6.85 years may still encounter obstacles while reading, but the obstacles are relatively minor.

5. Research Results

The data from the previous chapter ensures that the three parts of *Rough Street* are of the same difficulty from various angles and indicators, with no significant differences, as shown in Table 1. The readability indicators also demonstrate that the Jeremy Tiang's version of the translation has high readability and fewer reading obstacles, allowing the target language readers to understand the novel content more easily and have a basic grasp of the Liaoshen region and even the Northeast China under the historical context. This undoubtedly lays a good linguistic foundation for the overseas dissemination of Shuang Xuetao's works and expands the reader base. In 2017, Zhuang Qinfang pointed out in the article that the dissemination of the English text mainly focuses on "readability," and the most widely used Flesch formula mainly considers "reading readability" and "human interest." "The number of personal pronouns per 100 words" (PW) and "the number of personal pronouns per 100 sentences" (PS) are the two most important variables in "human interest." The word length of the entire text is calculated by Readability Analyzer 2.0, with a word length of 4.25 and 12.2.

6. Conclusion

This paper takes the English translation of Shuang Xuetao's collection *Rouge Street* as the research object, exploring the interactive relationship between the linguistic structure, cultural load, and subjective readability of literary translations through a mixed method of readability formula analysis and reader interviews. The study finds that traditional readability indicators can predict the linguistic difficulty of translations to some extent, especially in the article *Bright Hall*, where the formula scores are highly consistent with reader feedback. However, in texts like *The Aeronaut*, readers' subjective experiences reveal cultural understanding barriers, narrative complexity, and emotional immersion that the formula scores cannot capture, indicating that readability is not only an objective attribute at the language level but also a dynamic interactive process between readers and the text.

References

[1] Dale, E., & Chall, J. S. (1949). The Concept of Readability. Elementary English, 26, 23.
[2] Pan, W., Li, X., Chen, X., & Xu, R. (2025). Textual form features for text readability assessment. Natural Language Processing, 31(3), 800–841. doi:10.1017/nlp.2024.50
[3] Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85(5), 363–394.
[4] McLaughlin, G. H. (1969). SMOG grading: A new readability formula. Journal of Reading, 12(8), 639–646.
[5] Baker, M. (2006). In Other Words: A Coursebook on Translation (2nd ed.). Routledge.
[6] ROBERT DE BEAUGRANDE, Theory and Practice in Applied Linguistics: Disconnection, Conflict, or Dialectic? Applied Linguistics, Volume 18, Issue 3, September 1997, Pages 279–313
[7] Kress, G., & Leeuwen, T.V. (2020). Reading Images: The Grammar of Visual Design (3rd ed.). Routledge.