# Sentiment Analysis of Reader Reviews on the English Version of To Live Featuring Text-Emoticon Blending

**Zixuan Jiang, Xiaohong Li**
*College of Foreign Languages, North China University of Science and Technology, Caofeidian, Tangshan, Hebei, China*

**Abstract: Yu Hua's novel *To Live* portrays the transformation of Chinese society through the tragic life of its protagonist Fugui. Its English translation has sparked extensive discussions overseas, where reader comments not only directly reflect emotions but also contain profound insights into cross-cultural interpretation. To address the multilingual and multimodal characteristics of these international reader reviews, this study integrates textual content with emoticons, employing Python-based word frequency and word cloud analysis to conduct sentiment analysis on Amazon reviews of the English version of *To Live*. By examining the layered and multidimensional expressions of emotions through vocabulary and emoticon usage, this research provides a nuanced interpretation of readers' emotional tendencies.**

**Keywords: *To Live*; Sentiment Analysis; Word Frequency Statistics; Emoticons**

## 1. Introduction

In the field of literary criticism, many readers express their reflections on novels through online platforms. Particularly with the widespread adoption of the internet and the acceleration of globalization, the dissemination and exchange of literature across borders have become more convenient. Outstanding domestic literary works have entered overseas markets through translation and copyright transactions, reaching a broader readership. Analyzing the emotional evaluations of foreign readers toward domestic literary works can effectively enhance the quality and visibility of these works, strengthen cross-cultural communication, and foster the creation of internationally influential literature.

Yu Hua's novel *To Live* has been translated into more than 40 languages and published in over 40 countries and regions, including the United States, the United Kingdom, South Korea, Japan, Australia, France, Germany, Italy, Spain, Portugal, and the Netherlands[1]. Through the tragic life of its protagonist Fugui, *To Live* reflects the transformations in Chinese society, and its English translation has sparked extensive discussions overseas. These comments not only directly convey emotions but also carry deeper implications for cross-cultural understanding. However, the multilingual and multimodal nature of overseas reader reviews poses challenges to traditional sentiment analysis.

Currently, a search on the China National Knowledge Infrastructure (CNKI) yields 126 publications related to research on the translated versions of Yu Hua's *To Live*, including 40 journal articles, 71 theses and dissertations, and 4 newspaper articles and books. Most of these studies focus on culture-loaded terms, translation strategies, eco-translatology, translation reception, metaphor translation, translation quality assessment, field theory, and contextual culture. Among these 126 publications, only one paper—titled "A Study on the Reception of the English Translation of *To Live* in the United States" by Dr. Zhou Xiaoling of South China Normal University [2]—addresses sentiment analysis. This study employs Python's TextBlob for tasks such as part-of-speech tagging and sentiment analysis, conducting discourse- and sentence-level sentiment analysis of reader reviews to understand their perceptions of the work.

However, when posting comments on social media, people often use not only text but also emoticons, images, animations, and even videos to express emotions[3]. How to integrate these multimodal data for sentiment analysis remains a key challenge. This paper simultaneously analyzes the emotional polarity expressed in both textual comments and emoticons to gain deeper insights into readers' emotional attitudes toward the translated version of *To Live*.

## 2. Research on Translated Work Reviews

 Copyright @ STEMM Institute Press

With the increasing frequency of global cultural exchanges, literary translations have occupied a growing share in the digital reading market. Platforms such as Amazon Crossing and the foreign literature section of Douban Books have accumulated a large number of online reader reviews of translated novels. These comments not only reflect readers' impressions of the works themselves but also contain unique judgments on translation quality. Conducting sentiment analysis on these reviews of translated novels can help understand readers' acceptance of the translations. Additionally, by examining different dimensions of reader feedback, it is possible to reconstruct a translation quality assessment system and further improve the quality of translations. With the continuous advancement of deep learning technologies, researchers are constantly updating methods to explore sentiment analysis in reviews of translated novels.

Sentiment analysis of textual reviews using machine learning or deep learning has become a current research hotspot. Literature[4] pioneered the use of Python's NaiveBayes Analyzer for sentiment analysis of reviews on *The Three-Body Problem*; Literature[5] employed the PatternAnalyzer sentiment analyzer to analyze reader reviews of Mo Yan's translated novels; Literature[6] utilized Python's Natural Language Toolkit 3.5 to perform sentiment and semantic analysis on reader review texts, examining word frequency and part-of-speech sentiment scores to analyze readers' star ratings, emotional attitudes, and review volume trends; Literature[7] applied Python's lexicon-based semantic analysis method to conduct sentiment analysis on online reader reviews of the English translation of *Life and Death Are Wearing Me Out*; Literature[8] combined ANTCONC and Python's TEXTBLOB to evaluate reader reviews of the English translation of A Thought of Eternity, analyzing the focus, attitudes, and emotions of overseas online readers toward the translated work.

Existing studies have increasingly employed advanced computational methods to examine reader responses to literary translations.[9] introduced a novel deep learning-based framework for sentiment analysis of translation reviews, achieving 94.69% classification accuracy through training on the Amazon Review Polarity dataset. While demonstrating superior capability in capturing nuanced emotional expressions, this approach demands substantial computational resources and annotated data. [10] implemented BERTopic modeling to automatically identify and cluster thematic keywords from overseas readers' comments on the English version of Wolf Totem, subsequently analyzing their overall sentiment orientation.[11] conducted a comprehensive evaluation of various sentiment analysis models using Amazon book review data, comparing traditional machine learning algorithms (Naive Bayes, KNN, CART), deep learning models (LSTM), and transformer-based architectures (RoBERTa). [12] developed an integrated NLP and sentiment analysis system, constructing a multilingual corpus encompassing news, social media, and literary texts to train deep learning models for accurate identification of sentiment tendencies, metaphorical expressions, and culture-specific lexicon across languages.

While these studies have advanced the application of machine learning and deep learning techniques, they have yet to incorporate multimodal elements such as images, videos, and emoticons in their sentiment recognition frameworks. To address this gap, the present study proposes an integrated analysis of textual content and emoticons, employing machine learning techniques to conduct sentiment polarity analysis of reader comments on the English translation of Yu Hua's To Live. This approach aims to provide a more comprehensive understanding of reader responses by capturing both verbal and non-verbal emotional expressions.

## 3 Data Processing

### 3.1 Data Collection

This study collected English book review data for *To Live* from the globally renowned online book platform Goodreads. Using Python's requests library to simulate client requests, POST requests with authentication tokens were sent to the GraphQL endpoint provided by Goodreads. By designing query templates and pagination mechanisms in the request body, complete review information was obtained, including reviewer names, review dates, ratings, and review content, which was then organized into a structured format. A total of 1,103 book reviews from June 2007 to March 2025 were collected, providing a substantial data foundation for subsequent textual analysis.

Copyright @ STEMM Institute Press 

### 3.2 Data Preprocessing

To ensure the quality of the English book review data for *To Live* and meet the requirements for textual analysis, all review data underwent preprocessing. First, data cleaning was performed: duplicate data were directly removed. For missing values, records with missing values were deleted if the proportion of missing values was low. If the proportion was high, interpolation or other reasonable methods were considered to fill in the missing values, ensuring data completeness; Next, data formats were standardized: review dates were converted into a standard datetime format to facilitate chronological analysis and visualization of reviews. For rating data, they were uniformly converted into numerical format, and the rating range was validated to ensure ratings fell within a reasonable interval (e.g., 1-5 points). Out-of-range abnormal ratings were corrected or removed; Then, text preprocessing was applied to the review content: HTML tags, special characters, and other irrelevant information were removed, retaining only plain text. All text was converted to lowercase to reduce word variants (e.g., "Good" and "good" were standardized as "good"), improving the efficiency and accuracy of text processing; Finally, the text underwent tokenization, where continuous text was segmented into individual lexical units for subsequent tasks such as word frequency statistics.

## 4 Results and Analysis

### 4.1 Sentiment Analysis Based on Word Frequency

To better analyze the sentiment polarity in reader reviews of the translated novel To Live, we conducted word frequency statistics on the collected data. The results are presented in Table 1 and Figure 1.

**Table 1. Word Frequency Statistics of Review Data**

| No. | Word | Frequency | No. | Word | Frequency |
|---|---|---|---|---|---|
| 1 | Book | 80 | 11 | Time | 14 |
| 2 | Read | 35 | 12 | Chinese | 14 |
| 3 | One | 24 | 13 | Better | 13 |
| 4 | life | 24 | 14 | Reading | 12 |
| 5 | Good | 24 | 15 | Movie | 11 |
| 6 | Heart | 23 | 16 | Stars | 11 |
| 7 | Story | 21 | 17 | Great | 10 |
| 8 | Sad | 15 | 18 | Man | 10 |
| 9 | Br | 15 | 19 | Finished | 10 |
| 10 | Really | 14 | 20 | Novel | 9 |

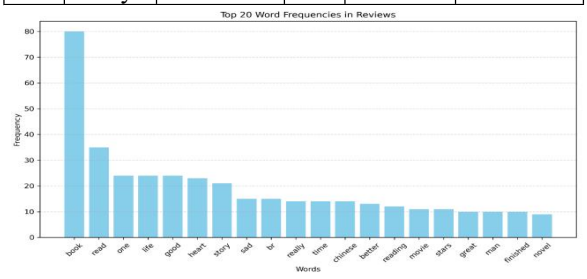


**Figure 1. Word Frequency Distribution in Reader Reviews**

As shown in Table 1 and Figure 1, the most frequently mentioned word by readers is "book" (80 occurrences), indicating its widespread recognition as the core subject of the work. This is followed by high-frequency words such as "read" (35 times), "life," "good," "heart," and "story," reflecting readers' concentrated attention on the reading experience, themes of life, emotional resonance, and narrative structure. Emotional vocabulary including "sad," "tear," "death," "devastating," and "heartbreaking" appears with particularly high frequency, demonstrating how the work's portrayal of human suffering and vicissitudes of fate evokes strong emotional 共鸣 among readers. Additionally, the words "translation" and "Chinese" are mentioned multiple times, revealing readers' heightened sensitivity to the quality of linguistic transformation and the socio-cultural context of China.

Figure 2 visually presents keywords from international readers' comments on Yu Hua's "To Live" through a word cloud. Prominently featured in large font size are words such as "book," "read," "life," "story," and "heart," indicating their extremely high frequency in the comments and their focus on the work's content, plot structure, and life themes. Emotionally charged words like "sad," "tear," "death," "devastating," and "tragedy" surround these core terms, reflecting the compassionate and empathetic psychology evoked by the work. Meanwhile, terms such as "movie," "translation," and "Chinese" are also frequently mentioned, suggesting that the book's cinematic adaptation and cultural context have become important points of discussion in the comments. The overall word cloud appears compact and semantically concentrated, clearly demonstrating the profound intellectual stimulation and emotional reverberations "To Live" has generated among international readers, while

 Copyright @ STEMM Institute Press

also highlighting its cultural penetration power in global literary dissemination.

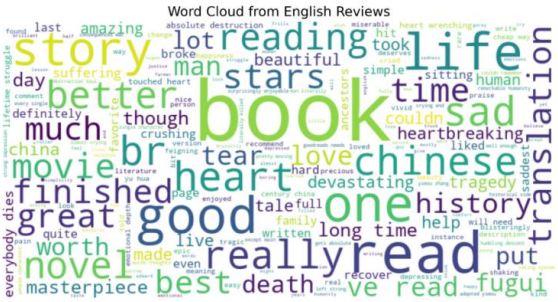


**Figure 2. Word Cloud Visualization of Review Keywords**

### 4.2 Sentiment Analysis Based on Emoticons

To gain deeper insights into readers' emotional responses, we conducted statistical analysis of emoticons appearing in the book reviews. The results are systematically presented in Table 2.

Table 2 Statistical Analysis of Emoticon Usage

| Emoji | Count | Emoji | Count |
|---|---|---|---|
| 😭 | 64 | 🥺 | 35 |
| 💔 | 47 | ✨ | 33 |
| 😱 | 39 | 📖 | 31 |
| ❤️ | 38 | 🌧️ | 31 |
| 😳 | 37 | 🥹 | 30 |
| 😞 | 35 | 🥺 | 26 |

As shown in Table 2, the crying face emoji "😭" ranked first with a frequency of 64 occurrences, followed by other common emotional symbols such as the broken heart "💔" and the shocked face "😱". The most frequently used emojis were predominantly those expressing sadness, heartbreak, surprise, or astonishment, indicating a prevalence of content with negative or complex emotional tones in the dataset.

Furthermore, we conducted an analysis of readers' emoji usage across different sentiment polarities, presented in bar charts as shown in Figures 3 to 5.

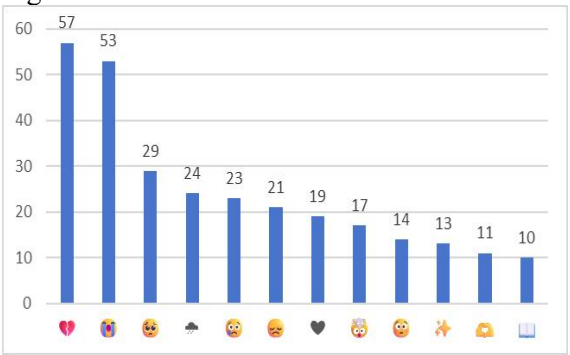


**Figure 3. Usage Distribution of Emoticons in Negative Sentiment Contexts**

From Figure 3, the data reveals that "💔" (broken heart) and "😭" (crying face) are the two most frequently used emojis in negative sentiment reviews, with occurrence counts of 57 and 53 respectively, significantly higher than other emojis. This indicates users predominantly employ "broken heart" and "crying" symbols when expressing sadness, grief, or other negative emotions. Additional emojis including "😞" (disappointed face), "🥺" (tearful face), and "🌧️" (cloud) also demonstrate relatively high usage frequency, further confirming that sadness and depression constitute the primary characteristics of negative emotional expression in these reviews.

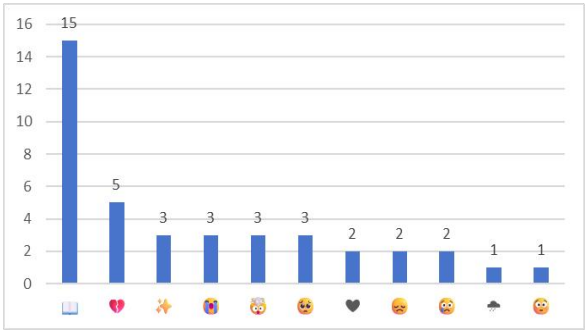


**Figure 4. Usage Distribution of Emoticons in Neutral Sentiment Contexts**

As shown in Figure 4, the "📖" (book) emoji ranks first with a frequency of 15 occurrences, significantly higher than other emojis. This indicates that users tend to prefer reading- and writing-related symbols when making neutral expressions.

Other emojis such as "💔" (broken heart), "✨" (sparkles), and "😭" (crying face) appear less frequently and are relatively evenly distributed. Overall, emoji usage in neutral sentiment contexts tends to be diverse but avoids strongly emotional symbols.

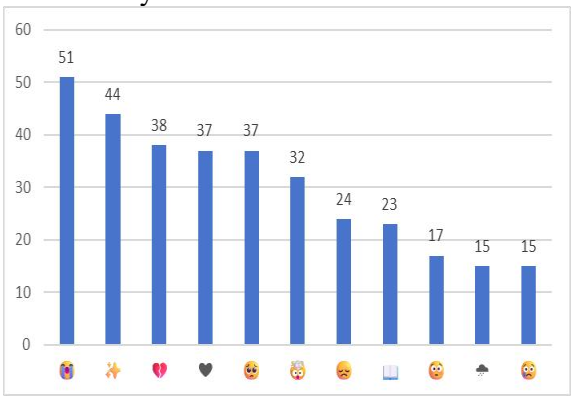


**Figure 5. Usage Distribution of Emoticons in Positive Sentiment Contexts**

As shown in Figure 5, in positive sentiment expressions, the "😭" (crying face) emoji ranks first with 51 occurrences, followed by "✨" (sparkles, 44 times) and "💔" (broken heart, 38 times). Notably, while "😭" is typically associated with sadness, its high frequency in positive contexts likely conveys meanings such as "moved to tears" or "tears of joy."

Additionally, positive symbols like "✨" (sparkles), "❤" (heart), and "🫶" (heart hands) also appear prominently, reflecting the diversity and emotional richness of emoji usage in positive sentiment expressions.

## 5. Conclusion

As an increasing number of Chinese literary works gain global recognition, understanding readers' emotional tendencies and focal points holds significant reference value for the overseas dissemination of translated Chinese literature. This study comprehensively analyzes reader sentiments by incorporating both textual content and emoticons from reviews, leveraging word cloud visualization and Python-based sentiment analysis techniques to gain nuanced insights into readers' perceptions of the work. While this research provides a detailed examination of reader emotions, future studies could expand the scope by incorporating sentiment analysis of images, animations, and videos in reviews. Such an approach would enable a more in-depth and holistic exploration of readers' underlying emotional attitudes.

## References

[1] Zhong, Q., Ding, L., Liu, J., et al. (2023). Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis. IEEE Transactions on Knowledge and Data Engineering, 35(10), 10098-10111.

[2] Zhou, X., & He, S. (2024). A study on the reception effects of the English translation of To Live in the United States. Foreign Language and Literature (Bimonthly), 40(4), 130-143.

[3] Jiang, X., Ma, J., Xiao, G., et al. (2021). A review of multimodal image matching: Methods and applications. Information Fusion, 73, 22-71.

[4] Zhang, L. (2019). Understanding overseas readers' reception and evaluation of Chinese translated literature through Python sentiment analysis: A case study of the English translation of The Three-Body Problem. Foreign Language Research, 4, 80-86.

[5] Shi, C., & Deng, L. (2020). A sentiment analysis-based study on Western reception of Mo Yan's novels in English translation. Foreign Language and Literature (Bimonthly), 36(3), 91-96.

[6] Wang, Y., & Shao, M. (2022). Research on the international dissemination effects of Chinese classics translation based on Python: A case study of the English translation of The Art of War. Foreign Language and Literature (Bimonthly), 38(3), 116-129.

[7] Li, J., & Wei, J. (2022). Research on overseas reception of the English translation of Life and Death Are Wearing Me Out based on online book reviews. Journal of PLA University of Foreign Languages, 45(3), 144-151.

[8] Research on overseas readers' reception of Chinese web literature: A case study of the English translation of A Thought of Eternity. (2023). Journal of Mudanjiang College of Education, 1, 12-16, 120.

[9] Wang, S., Cai, G., & Lv, G. (2024). Aspect-level multimodal sentiment analysis based on co-attention fusion. International Journal of Data Science and Analytics, 1-14.

[10] Deng, K., & Deng, Y. (2025). Research on the dissemination effects of the English translation of Wolf Totem based on BERTopic model. Technology Enhanced Foreign Language Education, 2, 70-77, 107.

[11] Yang, H., Zhao, Y., & Qin, B. (2022). Face-sensitive image-to-emotional-text cross-modal translation for multimodal aspect-based sentiment analysis. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 3324-3335).

[12] Liu, Z., Lin, Y., Cao, Y., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10012-10022).

 Copyright @ STEMM Institute Press