Scientific Research Publishing

Open Journal of Social Sciences, 2025, 13(9), 209-216
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

What Data Can Reveal in Analyzing Literature

—A Corpus-Based Study on Two Chinese Translations of *A Room of One's Own* at the Lexical Level

Ting Hu, Ziyue Feng

School of Foreign Languages, Shanghai Institute of Technology, Shanghai, China
Email: huting@sit.edu.cn

How to cite this paper: Hu, T., & Feng, Z. Y. (2025). What Data Can Reveal in Analyzing Literature. *Open Journal of Social Sciences, 13,* 209-216.
https://doi.org/10.4236/jss.2025.139013

Received: August 22, 2025
Accepted: September 19, 2025
Published: September 22, 2025

Abstract

Quantitative methods have empowered researchers with new perspectives to assume, test, and interpret what can hardly be analyzed through manual reading. This study employs corpora and data processing techniques to analyze and compare the discourse of *A Room of One's Own* by Virginia Woolf and its two Chinese translations by Zhou Yingqi and Jia Huifeng, respectively. WordSmith and Python are used for data processing and statistical analysis for quantitative examination at lexical levels. The study disclosed a paradox of Woolf's stylistic complexity versus her low lexical diversity that may not have been adequately clarified in previous studies. The statistical findings offered a data-driven explanation for this apparent contradiction. The comparison also unveiled the diverse translation strategies employed by different translators due to their varying interpretations of the source novel and the distinct ideological positions or communicative goals that have informed their choices.

Keywords

Corpus, *A Room of One's Own*, Translation, Strategies

1. Introduction

The growing corpora and data analysis technology are transforming translation comparison studies by providing empirical, large-scale, and systematic evidence. The corpus-based approach can offer a more objective way to identify patterns in translation that anecdotal observation and manual analysis can hardly uncover (Heine & Narrog, 2015). The frequencies and statistical information about the structures and rules of texts that are invisible when looking into a single context can be retrieved, providing an overall picture out of thousands of usages (Atar & Erdem, 2019). This study employs corpora techniques to analyze and compare the

two translations of Virginia Woolf's A Room of One's Own by two translators, Zhou Yingqi and Jia Huifeng, at the lexical level, with an aim to disclose some trends that are unseen before.

First published in 1929, A Room of One's Own was composed based on two lectures given by the author in 1928 at Newnham College and Girton College, two higher education institutions for women at Cambridge at that time. By stating the status of women and women artists in particular, Woolf presents her central theme that if a woman were to write, having money and a room of her own was essential. Women are not short of creativity, but were undermined by centuries of discrimination and disadvantages in finance and education imposed and practiced by the culture and society of that time.

In the novel, metaphors are used to illustrate social injustices and women's lack of free expression. A fish is used to exemplify how a woman with some ideals for writing is rigidized and loses her inspiration eventually, for the society labels women as mere domestic child bearers, ignorant and chaste. Woolf's arguments are extended to many political circumstances later, in which case the shape of social problems has shifted, but the absence of opportunity still causes isolation and inequality. These arouse social attention and awareness of women's suffering from the pressures of society that limited women. Women have been marginalized for decades. In literature, a male-dominated field, women's writing and publishing symbolize the rise of women and challenges to the discrimination.

Virginia Woolf was thus recognized as a prominent literary figure in history, with A Room of One's Own as the most famous non-fiction work on feminism. She is acknowledged as one of the best and brightest writers of the 20th century not only for her impressive descriptive powers, which are characterized by lively, graceful prose, but also for her efforts to push the boundaries of activism and literature (Transue, 1986).

2. Data Collection and Processing

Texts of Virginia Woolf's A Room of One's Own and two translations from Jia Huifeng and Zhou Yingqi are first cleaned in Word to remove garbled characters, spaces, tabs, carriage returns, and redundant information. WordSmith, primarily designed to tokenize based on whitespace, is used for Woolf's text in English for word counting or clustering in either alphabetical or frequency order. Concordances are created in the Concord module to facilitate the search of previously defined text, statistical profiles of the words or word forms are unveiled in the Wordlist module, and keywords can be listed in the KeyWord module to show their frequency.

However, as the Chinese language does not use spaces to separate words, tokenizing Chinese in WordSmith is a known challenge. Thus, Jieba, a Python library for Chinese text segmentation, is used with the jieba.cut() function on the two Chinese translation texts. The pre-processed texts with spaces between words are then imported into WordSmith to perform analysis.

3. Results and Discussion

Linguistic variables of tokens and types are statistical terminologies used in corpus linguistics to unpack the model of texts. Words in different contexts are to be calculated and then compared to identify patterns of language use. Different analyses using different tools under different variable definition criteria may yield quite different results. Therefore, in this study, a token is defined as a word counted as one token that is separated by white space or punctuation. A type refers to how many different word forms exist in the text. The occurrence of the same word is counted as one type.

3.1. Token and Type

As shown in Table 1, the tokens of Jia's translation and Zhou's translation are respectively 34,543 and 34,502, while the number of types of these two Chinese translations are respectively 7,151 and 6,221, compared with Woolf's 37,655 and 5,573. It indicates that there is no significant difference in the two translations from Woolf's essay. However, the distinction in type is obvious between the translations and Woolf's original work, and between the two translations. Word forms in Jia's translation are significantly more diverse than Zhou's, with the difference being 1,596, or 28.7%, and 930, or 11.9%, respectively. This signifies that the diversity of words in Jia's translation is higher than Zhou's (See Table 2).

Table 1. Token and type.

Text	Token	Type
Woolf's	37,655	5573
Zhou's	34,502	6221
Jia's	34,543	7151

Table 2. Tokens by chapter.

Chapter	Woolf's	Jia's	Zhou's	Jia vs. Woolf's	Zhou vs. Woolf's	Jia vs. Zhou
Chapter 1	7509	6857	7083	652	426	226
Chapter 2	5300	4778	4746	522	554	32
Chapter 3	5383	4915	4976	468	407	61
Chapter 4	7000	6447	6412	553	588	35
Chapter 5	5498	5254	5033	244	465	221
Chapter 6	6965	6292	6252	673	713	40
Total	37,655	34,543	34,502	3112	3153	41

Tokens in each chapter are then calculated to observe the dispersion of words in the essay and translation. Table 2 shows that the two translations are similarly close in tokens with two exceptions, Chapter 1 and Chapter 5. The first and fifth

chapters introduce and reiterate the essay's theme that a woman should have her own room for writing, with evidence exemplifying the inequality women endure. The divergence of the word count in these two chapters indicates that the two translators use different words and word lengths in translation.

To investigate how words are diverse in addition to word count distribution, tokens by word class are investigated. As **Table 3** shows, in general, verb and noun are the two word classes used most in the Chinese translations, much higher than adverb and adjective, almost 3 to 3.5 times higher, while in Woolf's work, verb and noun are 3 to 3.5 times higher than adverb and adjectives. This echoes what previous studies have found. The proportion of nouns and verbs in Chinese speakers is evidently higher than that of English speakers, even from childhood (Choi, 2000). Despite individual characteristics and activity context of children, the extent of nouns or verbs in Chinese children's vocabularies is equally significant. When comparing noun and verb specifically, English speakers contain a much higher proportion of nouns in their vocabulary than verbs, while Chinese speakers use almost the same number of nouns and verbs. The proportion of verbs used by Chinese speakers is much higher than that of English speakers (Tardif, 2006). The same patterns are found in two Chinese translations.

Table 3. Tokens by word class.

Word Class	Zhou's	Percentage	Jia's	Percentage	Woolf's
Noun	6901	9.14%	6928	8.98%	8726
Verb	6981	9.25%	7214	9.35%	6622
Adjective	2518	3.34%	2505	3.25%	2369
Adverb	2659	3.52%	2967	3.84%	2298
Pronoun	4214	5.58%	3729	3.84%	3565
Conjunction	1363	1.81%	1456	4.83%	1877
Preposition	1101	1.46%	1044	1.89%	5536
Numerals	807	1.07%	904	1.35%	332
Total	26,544	35.17%	26,747	35%	25,943

As widely recognized, nouns and abstract entities are often prioritized in Woolf's prose to create a rhythm of contemplative stasis (Ahmed, Ahmad, & Afzaal, 2024). Such economy of verbs embodies her feminist discourse in that the nominal weight highlights women's material and intellectual space while verbs symbolize masculinized narratives of action and conquest. By contrast, a greater density of verbs and adverbs in the Chinese translations is firstly due to Chinese linguistic structure. Verb-driven constructions are more entitled in Chinese than in English, which illustrates the use of domestication strategies in translation to clarify logical relations and adverbs to intensify evaluative stance. However, it may soften Woolf's stylistic resistance that she produces and demands for women (Huang, 2020).

3.2. Type/Token Ratio and Standardized Type/Token Ratio

To investigate how different words are used to explore their meanings for communication, lexical diversity statistics are often calculated to determine whether a wide range or a limited range of vocabulary is used in the text. TTR (type/token ratio) is the proportion of different word forms versus running words. A higher TTR value usually points to a more lexically varied text. STTR (standardized type/token ratio) is the text divided into standard-size segments, here in 1,000 words, to calculate the mean value of the TTRs for each segment. TTR is calculated as Formula 1:

$$\frac{\text{type}}{\text{token}}\text{ratio} = \frac{\text{number of types in text or corpus}}{\text{number of tokens in text or corpus}} \quad (1)$$

As shown in Table 4, the STTR (Standardized Type/token ratio) in Jia's translation (55.95%) is relatively higher than in Zhou's (53.01%), but the discrepancy (2.94%) is small, indicating no significant divergence in translations. Notably, the TTR and STTR in the Chinese translations are significantly higher than in Woolf's work, with the difference being 9.25% and 12.19%, respectively.

Table 4. TTR and STTR.

Text	TTR	STTR
Woolf's	14.80%	43.76%
Zhou's	18.12%	53.01%
Jia's	20.70%	55.95%

The small difference between the two translations highlights how both the inherent linguistic distinctions of Chinese and English, and the specific strategies employed by the translators, contribute to the final translated text.

3.3. Keywords

Key words are investigated to explore why there is a paradox of Woolf's reputation for stylistic complexity but with lower TTR and STTR compared with its translations. As Woolf's A Room of One's Own is about women and advocates equality and feminism, the world "women" and "equality" are examined. Table 5 shows that Woolf used the word "women" in her essay 196 times and "female" 6 times while in the two Chinese translations. Table 6 shows that "女性" is used 159 times and 269 times, and "女人" is used 127 times and 83 times respectively by Jia and Zhou. A chi-square test shows a statistically significant difference between Zhou's and Jia's usage of "女性" and "女人", with $p < 0.00000005$.

As the statistics indicate, Woolf's prose relies heavily on the repetition of such key words as 'room', 'money', 'women', and 'freedom'. The recurrence produces the rhetorical backbone of her feminist argument and secures thematic cohesion across the texts. That also explains the paradox. Meanwhile, the higher lexical

Table 5. Keyword "women, girl, female" in Woolf's text.

Word	TTR
women	196
Zhou's	15
Jia's	6

Table 6. Keyword "women" in translations.

Word	Zhou's	Jia's
女性	269	159
女人	83	127
女子	4	31
女士	6	2
女孩	10	5

diversity, as illustrated in TTR and STTR in the two translations, is caused by synonymic variation and grammatical compensation. "女性", "女人", "女子" and "女孩" are used alternately in the Chinese translation to reduce repetition according to context. This increases the TTR/STTR. Meanwhile, meanings are often compressed in English nominalization, leading translators to unpack the nouns with verbs, adjectives, and adverbs to specify and elaborate, to account for Woolf's abstract terms and to bridge cultural gaps. These extensions add lexical variety.

What is interesting is the contrast in the number of "女性" in the two translations. To explore the specific meaning of "woman" and "female," definitions from several dictionaries are examined. In the Oxford English Dictionary, "woman" is first a noun, referring to "an adult female human, the counterpart of man," while the word "female" "describes the presence of ovaries and a uterus in some object that experiences sexual differentiation (human, mammal, lizard, dolphin, etc.)" (Oxford Dictionary). The word "woman" and its plural form "women" are used as opposed to "men," while "female" emphasizes sexual distinction from a more biological view. That explains why in Woolf's original work, the word "women" was used most of the time. In the English-Chinese Cambridge Dictionary, the word "woman" is translated as "成年女子，女人," while "female" is translated as "雌性动物；女人，女性" as a noun. "女性" being specifically used in translating "female" rather than in "woman" signifies that in English, the word "女性" represents more sexual significance. In the Chinese Dictionary (汉语大词典), the definition of "女" is "女性，与'男'相对," which is closer to "woman" in English in terms of literary meaning.

The linguistic alternation of 'women' and 'female' is deliberately used in Woolf's prose to distinguish identity and critique. To Woolf, "woman" is used to advocate women's historical and social identities, who are deprived of education, financial independence, and access to the literary canon (Moi, 1985; Marcus,

1981) solely because of gender regardless of individual difference (Showalter, 1977). "Female" is used to signal the perspective imposed to reduce women to their biological sex rather than recognizing their intellectual or artistic subjectivity (Gubar, 1981; Lanser, 1992), highlighting the objectification and categorization of women by male-dominated institutions. The rhetorical shifts are used to demonstrate how language itself is engaged in the construction and limitation of gender identities (Marcus, 1981).

For the translations, Zhou's preference for "女性," which carries stronger academic, sociological, and feminist implications in modern Chinese discourse, illustrates the intention of maintaining consistency with a feminist, academic register that highlights women's social identity instead of their biological role. This aligns with Zhou's gender-oriented translation strategy.

In comparison, Jia varies "women" into "女人", "妇女" and "女性" to produce a more diverse lexical field. "女人" indicates the existence of women in everyday life with a broader cultural framing, and "妇女" implies the demographic presentation in Chinese, though it is used more in social policy contexts. This synonymic variation suggests an orientation toward accessibility to facilitate the readability of the general public, a domestication strategy.

4. Conclusion

The use of corpora and data analysis has provided empirical grounds on which researchers can interpret, test, or exemplify what could never be analyzed through manual reading. Some subtle, non-obvious patterns become apparent, and paradox is explained in interpreting the statistical results. Woolf's repetition of thematic keywords to increase feminist solidarity decreases its lexical diversity. The language structural differences in Chinese and English result in the disparity of word class in Woolf's prose and the two Chinese translations. The distinctive translation strategies adopted by the two translators are manifested, with Zhou following more of the feminist strategy while Jia chooses domestication to avoid monopoly. The findings complement the theoretical assumptions with empirical evidence. The quantitative methods will advance our understanding of discourse and translation from a more diverse perspective.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Ahmed, S. N., Ahmad, S., & Afzaal, K. (2024). Exploring Subjectivity and Modernist Language Techniques: Time, Gender, and Identity in Virginia Woolf's A Room of One's Own. *Journal of Social Sciences Review, 4,* 169-180. https://doi.org/10.54183/jssr.v4i4.439

Atar, C., & Erdem, C. (2019). The Advantages and Disadvantages of Corpus Linguistics and Conversation Analysis in Second Language Studies. 140-142.

Choi, S. (2000). Caregiver Input in English and Korean: Use of Nouns and Verbs in Book-Reading and Toy-Play Contexts. *Journal of Child Language, 27,* 69-96. https://doi.org/10.1017/S0305000999004018

Gubar, S. (1981). What Ails Feminist Criticism? *Critical Inquiry, 8,* 255-273. https://doi.org/10.1086/448153

Heine, B., & Narrog, H. (2015). *The Oxford Handbook of Linguistic Analysis.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199677078.001.0001

Huang, Z. (2020). A Room of One's Own in Suzhou: Analyzing the Dissemination of Woolf's Feminism in China. *Advances in Social Science, Education and Humanities Research, 638,* 118-121.

Lanser, S. S. (1992). Fictions of Authority: Women Writers and Narrative Voice. Cornell University Press, 304.

Marcus, J. (1981). *New Feminist Essays on Virginia Woolf.* Macmillan. https://doi.org/10.1007/978-1-349-05486-2

Moi, T. (1985). *Sexual/Textual Politics: Feminist Literary Theory.* Routledge.

Showalter, E. (1977). *A Literature of Their Own: British Women Novelists from Brontë to Lessing.* Princeton University Press. https://doi.org/10.1515/9780691221960

Tardif, T. (2006). But Are They Really Verbs? Chinese Words for Action. In K. A. Hirsh-Pasek, & R. M. Golinkoff (Eds.), *Action Meets Word: How Children Learn Verbs* (p. 477). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.003.0019

Transue, P. J. (1986). *Virginia Woolf and the Politics of Style.* SUNY Press.