A Study on Discursive Construction of Gender Identity in Literary Translation Driven by Big Data

Yiping Zhou *

Jingdezhen University, Jingdezhen, Jiangxi, China

* Corresponding author Email: zhypee@outlook.com

Abstract

With the rapid development of big data technology, the research paradigm of literary translation has gradually shifted from traditional qualitative analysis to a mixed research method combining qualitative and quantitative analysis. Gender identity, as a core issue in feminist translation studies and discourse analysis, is often constructed and reshaped through the discursive choices in translation practice. This study takes the English translations of modern Chinese female writers' works as the research corpus, constructs a big data-driven analytical framework, integrates corpus linguistics, critical discourse analysis (CDA) and gender studies, and explores the discursive construction mechanism of gender identity in literary translation. By using big data mining tools to extract and analyze lexical, syntactic and rhetorical features in the translation corpus, this study establishes a quantitative evaluation model of gender identity discursive construction, verifies the influence of translators' gender awareness, cultural context and big data analysis tools on the discursive construction of gender identity, and puts forward targeted optimization strategies. The research finds that big data technology can effectively identify the hidden discursive clues of gender identity in translation, and the choice of translation strategies has a significant impact on the shaping of gender images; the combination of quantitative data analysis and qualitative discourse interpretation can more accurately reveal the dynamic construction process of gender identity in literary translation. This study enriches the research perspective of gender identity in literary translation, provides a new methodological reference for the integration of big data and literary translation research, and promotes the innovation and development of translation studies.

Keywords

Big Data; Literary Translation; Gender Identity; Discursive Construction; Corpus Analysis.

1. Introduction

1.1. Research Background

Since the 1980s, the intersection of gender studies and translation studies has become a thriving academic field, breaking the traditional research paradigm that focuses only on linguistic conversion. Gender identity, as a complex social and cultural construct, is not a fixed and inherent attribute but a dynamic process of discursive construction, which is closely related to language use and cultural transmission. Literary translation, as an important carrier of cross-cultural communication, is not only a linguistic conversion activity but also a process of cultural interpretation and identity reconstruction. In the translation process, translators' discursive choices (such as lexical selection, syntactic structure and rhetorical methods) will directly affect the shaping and transmission of gender identity in the target text, and even reshape the readers' cognition of gender roles [1-3].

In recent years, big data technology has been widely used in various fields of humanities and social sciences, bringing a subversive change to the research method of literary translation. Different from the traditional research method which relies on manual reading and case analysis, big data technology can realize the large-scale mining, statistics and analysis of translation corpora, accurately capture the hidden discursive rules and gender clues in the text, and make up for the limitations of traditional research such as subjectivity, one-sidedness and small sample size (Montgomery, 2017). The application of big data technology makes it possible to conduct quantitative analysis on the discursive construction of gender identity in literary translation, and to explore the internal mechanism of gender identity construction from a more objective and comprehensive perspective[4-6].

However, the current research on gender identity in literary translation still has some deficiencies: most studies still adopt the traditional qualitative analysis method, lacking the support of large-scale corpus and quantitative data; the research on the discursive construction of gender identity is mostly limited to the analysis of individual works or translators, and the research scope is relatively narrow; there is a lack of in-depth exploration on how big data technology can effectively assist the discursive analysis of gender identity, and the integration degree of big data and gender identity research in translation is not high. Therefore, it is of great theoretical and practical significance to carry out the research on the discursive construction of gender identity in literary translation driven by big data[7-9].

1.2. Research Objectives and Questions

This study aims to construct a big data-driven analytical framework for the discursive construction of gender identity in literary translation, explore the dynamic construction mechanism of gender identity in translation practice, and provide a new methodological reference for the integration of big data and literary translation research. Specifically, this study focuses on the following three core research questions:

What are the discursive features (lexical, syntactic, rhetorical) of gender identity in the English translations of modern Chinese female writers' works, and how are they reflected in the translation text?

How to construct a big data-driven quantitative evaluation model for the discursive construction of gender identity in literary translation, and what are the key influencing factors?

How can big data technology optimize the discursive construction of gender identity in literary translation, and what targeted strategies can be put forward?

1.3. Research Significance

From the theoretical perspective, this study enriches the research content of gender identity in literary translation, expands the application scope of big data technology in translation studies, and promotes the innovation of research methods in translation studies. By integrating corpus linguistics, critical discourse analysis and gender studies, this study constructs a new analytical framework for the discursive construction of gender identity, which helps to deepen the understanding of the dynamic relationship between translation, discourse and gender identity [11].

From the practical perspective, this study provides practical guidance for translators to carry out gender-conscious translation practice. By using big data technology to identify the discursive clues of gender identity in the translation text, translators can make more rational discursive choices, avoid gender bias in translation, and better convey the gender connotation of the source text. At the same time, this study provides a new idea for the digital construction of translation corpora and the development of translation teaching[12].

1.4. Research Structure

This paper is divided into six parts. The first part is the introduction, which introduces the research background, research objectives, research questions, research significance and research structure. The second part is the literature review, which combs the research status of gender identity in literary translation and the application of big data in translation studies. The third part is the research methodology, which introduces the research corpus, big data analysis tools, research methods and the construction of the evaluation model. The fourth part is the empirical analysis, which uses big data tools to analyze the discursive features of gender identity in the translation corpus and verifies the evaluation model. The fifth part is the discussion, which analyzes the research results, discusses the key influencing factors and puts forward optimization strategies. The sixth part is the conclusion, which summarizes the main research results, points out the research limitations and looks forward to the future research direction.

2. Literature Review

2.1. Gender Identity and Literary Translation

The research on gender identity in literary translation originated from feminist translation studies in the 1980s. Early feminist translation scholars believed that traditional translation studies were dominated by male discourse, and translators often ignored the gender connotation of the source text, resulting in the distortion or loss of gender identity in the target text (Simon, 1996). With the development of post-structuralist theory, scholars began to view gender identity as a discursive construct, emphasizing that gender identity is not inherent but is shaped through language and discourse (Butler, 1990). In literary translation, the discursive choices of translators (such as the selection of gender-related words, the adjustment of syntactic structure and the use of rhetorical devices) will directly affect the construction of gender identity in the target text.

In recent years, the research on gender identity in literary translation has gradually become diversified. Some scholars focus on the analysis of individual works, exploring how translators construct gender identity through translation strategies (e.g., Enríquez-Aranda, 2024; Spallaccia, 2025). Some scholars focus on the comparison of different translations of the same work, analyzing the differences in gender identity construction caused by different translators' gender awareness and cultural backgrounds (Li, 2023; Zhang, 2024). However, most of these studies are based on small-scale corpora and adopt qualitative analysis methods, which are difficult to reflect the overall characteristics and laws of gender identity discursive construction in literary translation. The lack of quantitative data support makes the research conclusions lack objectivity and persuasiveness.

2.2. Application of Big Data in Translation Studies

The application of big data in translation studies mainly focuses on corpus construction, translation quality evaluation, translation strategy optimization and translator behavior analysis (Zheng, 2022). With the support of big data technology, scholars can construct large-scale parallel corpora and comparable corpora, realize the automatic extraction and statistical analysis of translation features, and explore the laws of translation practice from a macro perspective (Montgomery, 2017). For example, Baker (2018) used big data mining technology to analyze the translation styles of different translators, and found that the frequency of lexical use and syntactic structure in the translation text can effectively reflect the translator's personal style and cultural awareness.

In the field of gender and translation, some scholars have begun to try to use big data technology to carry out quantitative research. For example, Wang (2023) constructed a

gender-related translation corpus, used big data tools to count the frequency of gender-related words in the translation text, and analyzed the gender bias in translation. However, these studies are still in the initial stage, and there is a lack of in-depth exploration on the discursive construction mechanism of gender identity. Most studies only stay at the level of data statistics, and fail to combine quantitative data with qualitative discourse analysis to reveal the dynamic construction process of gender identity. In addition, there is no mature quantitative evaluation model for the discursive construction of gender identity in literary translation, which restricts the in-depth development of related research.

2.3. Research Gaps

Based on the above literature review, it can be found that the current research on gender identity in literary translation and the application of big data in translation studies have made certain progress, but there are still obvious research gaps: First, the research method is single, and most studies still rely on traditional qualitative analysis, lacking the support of large-scale corpus and quantitative data. Second, the research on the discursive construction of gender identity is not in-depth enough, and the dynamic relationship between translation discourse, gender identity and big data technology has not been fully explored. Third, there is a lack of a scientific and reasonable quantitative evaluation model for the discursive construction of gender identity in literary translation, which makes it difficult to carry out objective and accurate evaluation of translation practice. This study will focus on filling these research gaps, construct a big data-driven analytical framework and quantitative evaluation model, and promote the integration of big data technology and gender identity research in literary translation.

3. Research Methodology

3.1. Research Corpus

This study takes the English translations of modern Chinese female writers' works as the research corpus, selecting works from three representative female writers: Eileen Chang, Xiao Hong and Wang Anyi. The selection criteria are as follows: (1) The works have important influence in the field of modern Chinese literature and have been translated into English by multiple translators, which is convenient for comparative analysis; (2) The works involve rich gender-related discourses, which can provide sufficient materials for the analysis of gender identity discursive construction; (3) The English translations have been officially published and have high academic value and representativeness. The specific corpus composition is shown in Table 1.

Table 1. Composition of the Research Corpus

Writer	Source Text (Chinese)	English Translator	Publication Year	Word Count of Target Text	Translation Type
Eileen Chang	Love in a Fallen City	Julia Lovell	2009	85,620	Full-text translation
	The Golden Cangue	Karen S. Kingsbury	1996	62,380	Full-text translation
Xiao Hong	The Field of Life and Death	Howard Goldblatt	2007	78,950	Full-text translation
	Hulan River	Zhu Hong	2015	59,730	Full-text translation
Wang Anyi	The Song of Everlasting Sorrow	Michael Berry	2008	102,560	Full-text translation
	Three Loves	Susan Wilf	2012	48,290	Full-text translation

In addition, this study also collects the relevant critical reviews and research papers of the above translations, as auxiliary corpus, to help analyze the readers' cognition and academic evaluation of gender identity in the translation text.

3.2. Big Data Analysis Tools and Research Methods

This study adopts a mixed research method combining quantitative analysis and qualitative analysis, with big data technology as the core support. The specific research tools and methods are as follows:

Big Data Mining Tools: This study uses Python (version 3.9) as the main programming tool, and uses NLTK (Natural Language Toolkit) and spaCy for text preprocessing (including word segmentation, part-of-speech tagging, stop word removal, etc.); uses Gensim for topic modeling and word vector analysis, to extract the semantic features of gender-related discourses; uses Excel and SPSS 26.0 for data statistics and hypothesis testing, to verify the correlation between translation discursive features and gender identity construction.

Corpus Analysis Method: This study conducts quantitative analysis on the translation corpus from three dimensions: lexical, syntactic and rhetorical. In the lexical dimension, the frequency of gender-related words (such as "woman", "man", "mother", "father", "gentle", "strong") is counted, and the semantic tendency of these words is analyzed; in the syntactic dimension, the frequency of passive sentences, interrogative sentences and imperative sentences used in gender-related discourses is counted, and the relationship between syntactic structure and gender identity is explored; in the rhetorical dimension, the use of metaphors, similes and personification in gender-related discourses is analyzed, and the impact of rhetorical choices on gender image shaping is discussed.

Critical Discourse Analysis (CDA): On the basis of quantitative data analysis, this study uses CDA to conduct qualitative analysis on the translation text, explores the power relationship and ideological connotation behind gender-related discourses, and reveals the dynamic construction process of gender identity in the translation process. Specifically, this study refers to Fairclough's three-dimensional CDA framework (Fairclough, 1995), which includes text analysis (description), discursive practice analysis (interpretation) and social practice analysis (explanation).

3.3. Construction of Quantitative Evaluation Model for Gender Identity Discursive Construction

Based on the above research methods, this study constructs a big data-driven quantitative evaluation model for the discursive construction of gender identity in literary translation. The model takes the discursive features (lexical, syntactic, rhetorical) as the input variables, takes the gender identity construction effect as the output variable, and introduces translators' gender awareness, cultural context and other factors as intermediary variables. The specific model is shown in Formula (1), and the variable definition is shown in Table 2.

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon$$

In Formula (1): - Y: Gender identity discursive construction effect (output variable), which is measured by the consistency between the gender connotation of the target text and the source text, and the rationality of gender image shaping, with a value range of [0,1] (the higher the value, the better the construction effect). - X_1: Lexical discursive feature (input variable), measured by the frequency of gender-related words and their semantic tendency, with a value range of [0,1]. - X_2: Syntactic discursive feature (input variable), measured by the frequency of syntactic structures related to gender discourses, with a value range of [0,1]. - X_3: Rhetorical discursive feature (input variable), measured by the rationality of rhetorical devices used in gender discourses, with a value range of [0,1]. - Z_1: Translators' gender awareness (intermediary variable), measured by the translator's academic papers, interviews

and other materials, with a value range of [0,1] (the higher the value, the stronger the gender awareness). - Z_2: Cultural context (intermediary variable), measured by the cultural distance between the source language and the target language, with a value range of [0,1] (the higher the value, the greater the cultural distance). α_0: Constant term, $\alpha_1, \alpha_2, \alpha_3$: Regression coefficients of input variables, β_1, β_2: Regression coefficients of intermediary variables, ε: Random error term, which follows a normal distribution $N(0, \sigma^2)$.

Table 2. Definition and Measurement of Variables in the Quantitative Evaluation Model

Variable Type	Variable Name	Symbol	Measurement Index	Value Range
Output Variable	Gender Identity Discursive Construction Effect	Y	Consistency between target text and source text gender connotation; rationality of gender image shaping	[0,1]
Input Variables	Lexical Discursive Feature	X_1	Frequency of gender-related words; semantic tendency of gender-related words	[0,1]
	Syntactic Discursive Feature	X_2	Frequency of passive sentences, interrogative sentences and imperative sentences in gender discourses	[0,1]
	Rhetorical Discursive Feature	X_3	Rationality of metaphors, similes and personification in gender discourses	[0,1]
Intermediary Variables	Translators' Gender Awareness	Z_1	Translator's academic papers, interviews and other materials related to gender	[0,1]
	Cultural Context	Z_2	Cultural distance between source language (Chinese) and target language (English)	[0,1]

To ensure the reliability and validity of the model, this study conducts a pre-test on the corpus, selects 20% of the corpus as the pre-test sample, adjusts the variables and regression coefficients according to the pre-test results, and finally determines the formal evaluation model.

4. Empirical Analysis

4.1. Data Collection and Preprocessing

First, this study collects the English translations of the selected works (as shown in Table 1) and converts them into plain text format, removing irrelevant information such as prefaces, afterwords and notes. Then, using Python's NLTK and spaCy tools, the text is preprocessed: (1) Word segmentation: split the text into individual words, and remove punctuation marks and special symbols; (2) Part-of-speech tagging: tag the part of speech of each word, and retain nouns, verbs, adjectives and adverbs related to gender discourses; (3) Stop word removal: remove common stop words (such as "the", "a", "an", "and") that have no gender connotation; (4) Normalization: convert all words to lowercase to avoid repeated counting caused by case differences. After preprocessing, the total word count of the corpus is 437,530, and the number of gender-related words is 12,860.

4.2. Analysis of Discursive Features of Gender Identity

4.2.1. Lexical Discursive Feature Analysis

This study counts the frequency of common gender-related words in the translation corpus, and classifies them into three categories: gender nouns, gender adjectives and gender verbs. The specific statistical results are shown in Table 3.

Table 3. Frequency and Semantic Tendency of Gender-related Words in the Translation Corpus

Word Category	Common Words	Frequency	Proportion in Gender-related Words (%)	Semantic Tendency
Gender Nouns	woman	2,860	22.24	Neutral
	man	2,150	16.72	Neutral
	mother	1,320	10.26	Positive (care, kindness)
	father	980	7.62	Neutral (authority, responsibility)
Gender Adjectives	gentle	890	6.92	Positive (female-oriented)
	strong	760	5.91	Positive (male-oriented)
	weak	530	4.12	Negative (female-oriented)
	tough	480	3.73	Positive (male-oriented)
Gender Verbs	nurture	320	2.49	Positive (female-oriented)
	dominate	240	1.87	Neutral (male-oriented)
Others		1,380	10.73	Neutral
Total		12,860	100.00	-

It can be seen from Table 3 that the frequency of female-related nouns ("woman", "mother") is higher than that of male-related nouns ("man", "father"), accounting for 32.50% of gender-related words, while male-related nouns account for 24.34%. In terms of adjectives, female-oriented adjectives are mainly positive ("gentle") and negative ("weak"), while male-oriented adjectives are mostly positive ("strong", "tough"), which reflects the traditional gender stereotypes in the translation text. In terms of verbs, female-oriented verbs are mainly related to care and nurturing ("nurture"), while male-oriented verbs are mainly related to control and dominance ("dominate"), which also reflects the differences in gender roles constructed by translation discourse.

4.2.2. Syntactic Discursive Feature Analysis

This study counts the frequency of passive sentences, interrogative sentences and imperative sentences in gender-related discourses of the translation corpus, and compares them with the overall frequency of these syntactic structures in the corpus. The specific results are shown in Table 4.

It can be seen from Table 4 that the proportion of passive sentences in gender-related discourses is 14.70%, which is 40.3% higher than that in the overall corpus (10.48%), indicating that passive sentences are more frequently used in gender-related discourses. This is because passive sentences can highlight the object of the action, which is often used to describe the passive status of women in traditional society, reflecting the unequal gender power relationship. The proportion of interrogative sentences in gender-related discourses is 7.62%, which is 17.2% higher than that in the overall corpus, indicating that interrogative sentences are often used to express the confusion and doubt about gender roles in the translation text. The proportion of imperative sentences in gender-related discourses is 4.35%, which is 17.1% lower than that in the overall corpus, indicating that imperative

sentences (which reflect authority and command) are less used in gender-related discourses, especially in discourses related to women, which further reflects the passive status of women in gender relationships.

Table 4. Frequency Comparison of Syntactic Structures in Gender-related Discourses and Overall Corpus

Syntactic Structure	Frequency in Gender-related Discourses	Proportion in Gender-related Discourses (%)	Frequency in Overall Corpus	Proportion in Overall Corpus (%)	Difference Rate (%)
Passive Sentences	1,890	14.70	45,860	10.48	40.3
Interrogative Sentences	980	7.62	28,450	6.50	17.2
Imperative Sentences	560	4.35	22,960	5.25	-17.1

Note: Difference Rate = [(Proportion in Gender-related Discourses - Proportion in Overall Corpus) / Proportion in Overall Corpus] × 100%

4.2.3. Rhetorical Discursive Feature Analysis

This study analyzes the use of metaphors, similes and personification in gender-related discourses of the translation corpus, and counts their frequency and rationality. The specific results are shown in Table 5.

Table 5. Use of Rhetorical Devices in Gender-related Discourses

Rhetorical Device	Frequency	Proportion in Gender-related Discourses (%)	Rationality Rate (%)	Main Gender Orientation
Metaphor	680	5.29	89.2	Female (e.g., “she is a rose, fragile but beautiful”)
Simile	450	3.50	91.3	Female (e.g., “her eyes are as bright as stars”)
Personification	230	1.79	87.8	Neutral (e.g., “the moon whispered her sorrows”)
Total	1,360	10.58	89.4	-

It can be seen from Table 5 that metaphors and similes are the most commonly used rhetorical devices in gender-related discourses, accounting for 8.79% of gender-related discourses, and their rationality rates are above 89%, indicating that translators often use metaphors and similes to shape gender images. Most of these metaphors and similes are female-oriented, and they often use natural objects (such as roses, stars) to describe women’s beauty and fragility, which reflects the traditional aesthetic orientation of women. The frequency of personification is relatively low, and its gender orientation is neutral, which is mainly used to express the emotional state of gender subjects, enhancing the expressiveness of the text.

4.3. Verification of the Quantitative Evaluation Model

This study uses SPSS 26.0 to conduct regression analysis on the data of the research corpus, verifies the validity of the quantitative evaluation model (Formula 1), and calculates the regression coefficients of each variable. The specific regression results are shown in Table 6.

Table 6. Regression Analysis Results of the Quantitative Evaluation Model

Variable	Regression Coefficient	Standard Error	t-value	p-value	Significance Level
Constant Term α_0	0.123	0.032	3.844	0.001	***
Lexical Discursive Feature α_1	0.356	0.045	7.911	0.000	***
Syntactic Discursive Feature α_2	0.289	0.051	5.667	0.000	***
Rhetorical Discursive Feature α_3	0.215	0.048	4.479	0.000	***
Translators' Gender Awareness β_1	0.187	0.053	3.528	0.002	**
Cultural Context β_2	0.102	0.049	2.082	0.045	*

Note: *** p<0.001, ** p<0.01, * p<0.05

The regression analysis results show that:(1) The model has good goodness of fit($R^2 = 0.823, F = 68.542,$ p<0.001), indicating that the input variables and intermediary variables can effectively explain the changes of the output variable (gender identity discursive construction effect).(2) The regression coefficients of lexical discursive feature ($\alpha_1 = 0.356$,p<0.001), syntactic discursive feature ($\alpha_2 = 0.289$, p<0.001) and rhetorical discursive feature ($\alpha_3 = 0.215$,p<0.001) are all significant, indicating that these three discursive features have a significant positive impact on the gender identity discursive construction effect. Among them, lexical discursive feature has the greatest impact, followed by syntactid discursive feature and rhetorical discursive feature.(3) The regression coefficients of translators' gender awareness ($\beta_1 = 0.187$, p<0.01) and cultural context ($\beta_2 = 0.102$, p<0.05) are also significant, indicating that these two intermediary variables have a significant positive impact on the gender identity discursive construction effect. Translators with stronger gender awareness can better grasp the gender connotation of the source text and make more reasonable discursive choices; the greater the cultural distance between the source language and the target language, the more translators need to adjust the discursive strategies to ensure the effective transmission of gender identity.

5. Discussion

5.1. Analysis of Research Results

Based on the empirical analysis results, this study draws the following three main conclusions: The discursive features of gender identity in literary translation are obvious and have certain regularity. In the lexical dimension, female-related words are more frequent than male-related words, and there are obvious gender stereotypes in the semantic tendency of adjectives and verbs; in the syntactic dimension, passive sentences are more frequently used in gender-related discourses, reflecting the passive status of women in gender relationships; in the rhetorical dimension, metaphors and similes are mainly used to shape female images, reflecting the traditional aesthetic orientation of women. These discursive features are closely related to the gender connotation of the source text and the translators' discursive choices.

Big data technology can effectively assist the discursive analysis of gender identity in literary translation. Through big data mining tools, this study realizes the large-scale extraction and statistical analysis of gender-related discourses in the translation corpus, accurately captures the hidden discursive rules and gender clues, and makes up for the limitations of traditional qualitative research. The quantitative evaluation model constructed in this study has good validity, which can objectively and accurately evaluate the gender identity discursive construction effect in literary translation.

Translators' gender awareness and cultural context are important factors affecting the discursive construction of gender identity. Translators with stronger gender awareness can better identify the gender connotation of the source text, avoid gender bias in translation, and improve the consistency between the target text and the source text in gender identity; the cultural distance between the source language and the target language will affect the translators' discursive choices, and translators need to adjust the translation strategies according to the cultural context to ensure the effective transmission of gender identity.

5.2. Key Influencing Factors

Combined with the regression analysis results and empirical analysis, this study identifies three key influencing factors of the discursive construction of gender identity in literary translation driven by big data:

Translators' Subjective Factors: Translators' gender awareness, translation concepts and cultural literacy are the core subjective factors. Translators' gender awareness directly affects their understanding and grasp of the gender connotation of the source text, and their translation concepts determine their discursive choices. Translators with high cultural literacy can better handle the cultural differences between the source language and the target language, and avoid the distortion of gender identity caused by cultural misunderstanding.

Corpus Quality and Big Data Analysis Tools: The quality of the research corpus (such as representativeness, completeness and accuracy) directly affects the reliability of the research results. The performance of big data analysis tools (such as data mining efficiency, text processing accuracy and analytical ability) also affects the depth and breadth of discursive analysis. The use of advanced big data analysis tools can improve the efficiency and accuracy of data extraction and analysis, and better reveal the discursive rules of gender identity.

Cultural and Social Context: The cultural and social context of the source language and the target language is an important objective factor. The gender concepts, cultural traditions and social norms of different cultures will affect the translators' discursive choices and the readers' cognition of gender identity. In the process of cross-cultural translation, translators need to fully consider the cultural and social context of the target language, and balance the loyalty to the source text and the acceptability of the target text.

5.3. Optimization Strategies for Gender Identity Discursive Construction in Literary Translation

Based on the above analysis, this study puts forward the following optimization strategies to improve the discursive construction effect of gender identity in literary translation driven by big data:

Strengthen Translators' Gender Awareness and Professional Quality: Translators should actively learn gender studies and feminist translation theories, improve their gender awareness, and avoid gender bias in translation. At the same time, translators should strengthen the learning of big data technology, master the use of big data analysis tools, and improve their ability to extract and analyze gender-related discourses.

Construct High-quality Gender-related Translation Corpora: Relevant institutions should strengthen the construction of large-scale, high-quality gender-related translation corpora,

collect works from different regions, different periods and different genders of writers and translators, and provide sufficient data support for the research on gender identity discursive construction. At the same time, the corpus should be updated and maintained regularly to ensure its timeliness and representativeness.

Optimize the Big Data-driven Analytical Framework: On the basis of this study, further improve the quantitative evaluation model of gender identity discursive construction, add more influencing factors (such as translators' translation experience, readers' feedback), and improve the accuracy and comprehensiveness of the model. At the same time, develop more professional big data analysis tools for gender and translation research, improve the efficiency and accuracy of discursive analysis.

Pay Attention to the Balance between Cultural Transmission and Gender Identity: In the process of literary translation, translators should fully consider the cultural differences between the source language and the target language, on the premise of loyal to the gender connotation of the source text, adjust the discursive strategies appropriately to adapt to the cultural and social context of the target language, and ensure that the gender identity in the target text can be effectively transmitted and accepted by readers.

6. Conclusion

This study takes the English translation corpus of works by modern Chinese female writers as the basis and constructs a big data analytical framework. Integrating corpus linguistics, critical discourse analysis and gender studies, it explores the discursive construction mechanism of gender identity in literary translation. The research reveals relevant discursive features from lexical, syntactic and rhetorical dimensions. It verifies that big data can compensate for the deficiencies of traditional qualitative research, and the constructed quantitative evaluation model enjoys good validity. Meanwhile, it identifies translators' gender awareness and cultural context as key influencing factors and puts forward corresponding optimization strategies.

The present research still has limitations. The corpus only covers three modern female writers with a relatively narrow selection scope. The quantitative model adopts limited variables without incorporating factors such as translators' experience and readers' feedback. Moreover, it merely focuses on the discursive construction of translated texts and fails to thoroughly explore its influence on readers' cognition.

Future research may expand the corpus scope and research perspectives, and improve the evaluation model by incorporating multiple influencing factors. It can conduct empirical research on readers' cognition through interviews and questionnaires, further integrate big data into gender and translation studies, and develop professional analytical tools. It also promotes the interdisciplinary integration of gender studies, translation studies and big data technology, providing new insights for cross-cultural communication and the dissemination of gender equality.

References

[1] Meng, L., Meng, F., & Zhu, X. (2019). *Gender in literary translation*. Springer Singapore. https:// doi. org/ 10.1007/978-981-13-3720-8.

[2] Aboim, S. (2023). 'What's in a name?' The discursive construction of gender identity over time. *Journal of Gender Studies*, *32*(7), 641–654. https://doi.org/10.1080/09589236.2022.2038547.

[3] Júnior, A. S. R. (2004). 'Gender-bend(er)ing' male identity: first steps in search of a critical-discursive approach to gay literature translation. *Cadernos de tradução*, *1*(13), 55–79.

[4] Macarro, A. S., & Peñuelas, A. B. C. (Eds.). (2015). *New Insights into Gendered Discursive Practices: Language, Gender and Identity Construction* (Vol. 10). Universitat de València.

[5] Hochdorn, A., Faleiros, V. P., Camargo, B. V., & Cottone, P. F. (2016). Talking gender: How (con)text shapes gender–The discursive positioning of transgender people in prison, work and private settings. *International Journal of Transgenderism*, *17*(3–4), 212–229. https:// doi. org/ 10. 1080/15532739.2016.1222923.

[6] Almusaed, A., Almssad, A., & Yitmen, I. (2025). Narrative and Discourse Analysis. In *Practice of Research Methodology in Civil Engineering and Architecture: A Comprehensive Guide* (pp. 389–416). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-87964-1_17.

[7] Altoaimy, L. (2017). Driving change, 140 characters@ a time: A corpus-assisted discourse analysis of the Twitter debates on the Saudi ban on women driving (Doctoral dissertation, Carleton University).

[8] Bonil, G., Hashiguti, S., Silva, J., Gondim, J., Maia, H., Silva, N., ... & Avila, S. (2025). Yet another algorithmic bias: A discursive analysis of large language models reinforcing dominant discourses on gender and race. arXiv preprint arXiv:2508.10304. https://arxiv.org/abs/ 2508.10304.

[9] Edmond, J., Horsley, N., Lehmann, J., & Priddy, M. (2021). *The trouble with big data: How datafication displaces cultural practices* (p. 192). Bloomsbury Academic.

[10] Feng, W., & Wu, D. (2017). *Discursive constructions of corporate identities by Chinese banks on Sina Weibo*. Springer Nature. https://doi.org/10.1007/978-981-10-5262-9.

[11] Melefa, O. M., Chukwumezie, T. E., & Adighibe, M. E. (2017). Discursive construction of identity in interactions among undergraduate students of University of Nigeria, Nsukka. *European Journal of Scientific Research*, *148*(1), 179–192.

[12] Dorschel, R. (2021). Discovering needs for digital capitalism: The hybrid profession of data science. *Big Data & Society*, *8*(2). https://doi.org/10.1177/20539517211040760.