BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Interdisciplinary Stylometric Analysis of Shakespeare's Translations: Unveiling Gender Dynamics through Lexical, Sentential, and Metrical Layers

Huan Wang

Foreign Studies College, Northeastern University, Shenyang 110169, China

ABSTRACT

While Shakespeare's works have been extensively analyzed, the influence of translator gender on stylistic features remains underexplored, and most certainly not via computational linguistics. This study aims to address this gap by investigating the intersection of gender and translation in the stylistic choices made by translators of Shakespearean plays into three different languages, addressing a notable gap in quantitative research within Translation Studies. The research uses advanced computational methods such as Python's sklearn library for Principal Component Analysis and hierarchical clustering to analyze the lexical, syntactic, and metrical elements involved in the translation process, focusing on data processing, tokenization, and stemming from studying the impact of gender on translation patterns. The findings reveal that translations cluster distinctly by language but exhibit significant stylistic differences based on the translator's gender. Female translators use more diverse words and sentence patterns, which shows they favor creative language and detailed interpretation of original texts. Male translators display stronger fidelity to Shakespeare's iambic pentameter in their Spanish sonnet translations. The research reveals that translators' gender choices play a role in selecting words and understanding and presenting the original content. These findings show how translators' gender backgrounds inform their translation choices, giving us fresh ways to understand old texts amid modern views on gender.

Keywords: Gender; Drama Translation; Shakespeare; Stylometric Analysis; Machine Learning; Principal Component Analysis

*CORRESPONDING AUTHOR:

Huan Wang, Foreign Studies College, Northeastern University, Shenyang 110169, China; Email: wanghuanlw@163.com

ARTICLE INFO

Received: 8 February 2025 | Revised: 10 March 2025 | Accepted: 18 March 2025 | Published Online: 19 March 2025
DOI: https://doi.org/10.30564/fls.v7i3.8692

CITATION

Wang, H., 2025. Interdisciplinary Stylometric Analysis of Shakespeare's Translations: Unveiling Gender Dynamics through Lexical, Sentential, and Metrical Layers. Forum for Linguistic Studies. 7(3): 927–941. DOI: https://doi.org/10.30564/fls.v7i3.8692

1. Introduction

Literary translation is often described as the translation of style, as it conveys linguistic meaning and the author's unique voice, tone, and intention [1]. It is a complex act of stylistic interpretation and reconstruction, where the translator's voice, consciously or unconsciously, shapes the target [2]. According to Baker's findings, translation choices reveal more than textual meaning; they display translators' beliefs, backgrounds, and gender biases [3]. Scholars now focus on how translators use gendered methods in literary translation as an important part of gender research [3, 4]. New scholarship shows us that gender appears in our everyday language exchanges and social behaviors [5, 6]. A translator brings their gender perspective into text translation, combining their language skills and cultural understanding to serve their audience. Translators use their gender experiences to adapt cultural gender roles in their work, which alters the reader's perception of story characters.

1.1. Research Objectives

While experts are paying more attention to how gender matters in translation jobs, few studies measure exactly how differently male and female translators handle gendered language patterns. Research into gender effects on literary translation has not yet been sufficiently studied [3, 7]. Due to his broad translation history, Shakespeare helps researchers study the role of gender in writing style and its effect on his plays' characters and cultural settings. The way translators handle text affects every part of Shakespeare's stories, including how characters evolve and interact through their words while changing the audience's understanding. When translators use gender-based methods, they can either maintain or reshape the rules and relationships of genders in the source material. Research shows that male and female translators handle translation complexities in distinct ways that must be better understood. More studies are needed to see how different male and female translators bring their gender views to translating Shakespeare's plays. This raises important research questions:

RQ1. How do male and female translators differ in their lexical, syntactical, and metrical choices when translating Shakespeare's works?

RQ2. To what extent can computational methods, such as Principal Component Analysis (PCA) and hierarchical clustering, uncover gendered stylistic distinctions in literary translations?

RQ3. How do these gendered choices reflect broader translation theory and practice trends?

1.2. Research Significance

By addressing these questions, this research is significant for several reasons. First, this study helps grow gendered translation studies by showing us numbers that show how men and women translators make different stylistic decisions. Second, this study introduces a new way to apply computer-based language analysis in literary translation, which connects modern technology to classic literary research methods. Finally, the study's results matter to several different fields: gender studies, linguistics, and literary theory. These results question old ideas about translation and show new ways that gender changes when adapting literature. Our study reveals translator choices by examining gender effects on translating Shakespeare's text. The study helps us better understand how social gender rules and cultural expectations influence the movement of literature through different periods and places.

2. Literature Review

2.1. Gender and Translation Studies

The intersection of gender and translation studies has evolved considerably from its initial feminist inquiries into the role of women translators. The first steps in this research area were taken by Simon, who examined how female translators were pushed to the sidelines in literature [8]. Starting from then, scholars began looking at how gender impacts all aspects of how translation works. According to Baker's research, gender affects translators' decisions and determines the ideological content in translations, while unconscious biases affect gender representation [3]. Our recent study matches findings from other researchers, showing that translators can either maintain or question standard gender beliefs through word choices [9, 10]. In addition, gender changes based on context and depends on social and language interactions to keep forming. Our research shows that in translation work, translators can change how gender is seen by choosing words that match new cultural meanings [11].

Current research keeps improving our understanding by examining how gender affects translation work. Research by Mikros and Boumparis showed different translation techniques based on their gender identities when gender-neutral language presented difficulties [12]. According to Ishige and Misiou and Spoturno, translators' gender backgrounds influence their text analysis methods and their handling of gendered content [13, 14]. New ways of studying translations about LGBTQ+ identities have been developed, pushing forward discussions about how these communities are shown in translated texts [9, 15]. New research shows that people now better understand how translators influence gender views in the culture they are translating for [16, 17]. Studies show that translator identity, text analysis, and cultural norms are closely connected and reveal that gender plays a role in translation techniques - a topic that requires more profound research [18, 19]. These changes show we are improving at seeing translation as a place where gender gets redefined and adapted in our connected world.

2.2. Computational Approaches to Stylistic Analysis

The latest progress in computational linguistics lets researchers perform more structured translation studies through data analysis. Stylometry uses statistics to find language patterns and provides essential insights into how translators make word choices at every language level. Studies using computational tools to analyze extensive text collections show how gender background affects translation results within language limits. Recent stylometric research found that male and female translators exhibit unique gender-based differences in word selection, affecting their translation content and aesthetic style [20]. Research shows that PCA successfully groups translation styles to show how translators differ in their work across languages and text types [21]. Our knowledge of digital platforms' effects on language and cultural exchange has grown with technology-based research methods. These results show that computational methods can detect new translation behavior patterns, which helps us better understand the connection between gender and translation style.

2.3. Gender Dynamics in Shakespeare's Translations

Recent research shows how Shakespeare's translators deal with complex gender issues when they bring his plays to modern audiences. Researchers show how modern translators adapt gender roles in Shakespeare's plays, especially for female characters. For instance, Duan and Tian examine how Anne Tyler's Vinegar Girl, a modern retelling of The Taming of the Shrew, negotiates the tension between feminist empowerment and patriarchal constraints, reflecting a more ambivalent portrayal of gender than Shakespeare's original [22]. Translating this work into Chinese further complicates the feminist message, as commercial motivations oversimplify the novel's gender dynamics, ultimately leading to a "failed" feminist narrative in the Chinese market. Similarly, studies by Finn and Gallimore explore how the transgressive elements of female characters in Shakespeare's tragedies, such as Eleanor Cobham and Margaret of Anjou, continue to resonate in translations, revealing how historical gender politics are reimagined for modern audiences [23, 24].

The research explores translation methods used to adapt the gender language and stage roles in Shakespeare's works. According to the research by Ingelbien, translators work to balance Shakespeare's gender language with his social observations during translation [25]. When translators adjust gender pronouns and titles between languages, they change how male dominance and female defiance appear in the text. In contrast, Kingery highlights the impact of cultural and linguistic context on the gendered dynamics of Shakespeare's works, pointing out that even well-meaning feminist interpretations can inadvertently reinforce traditional gender roles when transplanted into non-Western or post-colonial societies [26]. Sakib et al. argue that the growing interest in gender-inclusive translations of Shakespeare's works can challenge hegemonic interpretations and offer more empowering representations for modern audiences, while Zavrl emphasizes the role of translation as a medium for both subverting and reinforcing gendered narratives in Shakespearean drama [27, 28]. Studies show that translation is a powerful method for handling gender issues since translators naturally insert their cultural beliefs about gender into their work.

The bond between gender and translation processes includes clear choices and hidden influences determining how we rewrite literature in different languages. Traditional translation research depended on qualitative analysis, but new computational linguistics methods like principal component analysis now help us spot differences in translation styles between genders. In the context of Shakespeare's transla-

tions, these differences are particularly significant, as the playwright's exploration of gender roles and relationships provides a rich terrain for examining how translators negotiate and perform gender through their linguistic choices. Through this research, we want to add to existing studies by using computer-based analysis of translation styles to understand how translators of Shakespeare's work choose words, sentence structures, and verse patterns in ways that reflect gender differences across various languages and cultures.

3. Datasets

We chose Shakespeare's plays for this study because they attract worldwide audiences, and numerous translators of both genders have rendered them into many languages over time. The translations offer researchers a unique chance to study how gender influences literary interpretation. The dataset holds equal male and female translations of 19 files, including Chinese, English, German, Italian, and Spanish versions of Romeo and Juliet, Coriolanus, and Macbeth [29–45]. As detailed in **Table 1**, the dataset includes diverse linguistic metrics such as word count, character count, average word size, and the number of unique words, categorized by translator gender: The dataset includes translation work by male translators (M), female translators (F), and teams of both male and female translators (M&F). Our large dataset offers extensive resources for stylometric study with 341,176 words and 2,256,062 characters distributed across 79,254 unique words. Additional details about the dataset are available in the supplementary file, *19 texts*.

4. Methodology

4.1. Stylometry and Computational Methods in Translation Studies

Stylometry uses data to analyze writing style, which helps us study translation decisions using a precise scientific method. Stylometry combines different text analysis methods to let computers study how gender affects translation work. Our study benefits from this method because it measures writing patterns in a factual way that can be duplicated to detect delicate stylistic features. The study includes several male and female translations from three famous Shakespeare plays. With the help of Python packages, the texts received standard preprocessing treatments to prepare them for evaluation. Our research team collects language attributes within three different measurement areas. Through Most Frequent Words (MFW) analysis, researchers discover vocabulary trends, and by measuring sentence length variability with syntactic structure analysis, they detect how texts are constructed. Measuring metrics shows how rhythm and sound systems vary in text. The analysis of translation styles becomes clear when we use hierarchical clustering, and PCA results are displayed through dendrograms and scatter plots. Our method reveals unique translation patterns from different genders and shows how these differences impact stylistic choices, so stylometry is the best research method.

4.2. Data Preprocessing

Before stylometric analysis, we must prepare our textual data because data preprocessing makes it better suited for analysis. Tokenization splits raw text into basic linguistic units, which serve as foundations for advanced text analysis. We reduce words to their base form during stemming and merge related word variations into one representation. Our system tags each word in a sentence by adding detailed syntactic data to show its specific grammatical function. Our system filters text to eliminate unnecessary characters and spaces so that only meaningful language data remains for analysis. Every text element needs to stay lowercase to prevent problems from case sensitivity. Our text preparation process follows a fixed method to prepare data for practical stylometric analysis. See the supplementary notebook *'Python notebook'* for details.

4.3. Feature Extraction

Feature extraction then processes the refined data from text into numerals, showing different linguistic features. The research team structured this phase into three connected categories processed through Python. Hoover's research shows that the words we use most often in our writing are essential functions that reveal our unique writing style, even though they lack specific content significance [46]. Within the author's unique usage of function words, analysis at scales such as 5 and 100 MFW can expose highly stylized discriminants of a translator's signature style. Syntactic features help us detect the hidden rules that shape how texts are put together by

Table 1. Linguistics metrics of Shakespeare's plays across multiple languages.

Play	Gender	Language	Word Count	Character Count	Average Word Size	Unique Words
Macbeth	**M**	**English**	**18089**	**106176**	**4.53**	**5349**
	F	German	17436	111864	5.00	6096
	M	German	19483	125299	5.24	6407
	M	Chinese	1973	34056	16.26	1735
	M&F	Chinese	1482	37816	24.48	839
Coriolanus	**M**	**English**	**29016**	**160309**	**4.52**	**6880**
	F	German	27358	175116	5.14	8149
	M	German	28021	170291	5.07	7968
	M	Chinese	3079	56887	17.45	2806
	M&F	Chinese	2338	62022	25.50	1277
Romeo and Juliet	**M**	**English**	**25448**	**145195**	**4.48**	**6618**
	M	German	24387	148006	4.93	7731
	F	German	10477	64935	5.02	2941
	F	Italian	30787	179421	4.83	8662
	M	Italian	30520	186264	4.84	9025
	F	Spanish	42390	246705	4.71	12034
	M	Spanish	23934	141970	4.67	7219
	M	Chinese	3231	53231	15.45	2563
	M&F	Chinese	1727	50499	28.15	989

measuring sentence length variation and analyzing sentence structure. The method detects unique sentence structures used by different writers or documents. The study uses syllable patterns to analyze text rhythm and sound patterns while revealing finer text details. The article uses different feature extraction methods to capture all aspects of language style, making the stylometric analysis trustworthy and powerful.

4.4. Clustering

To analyze the corpus properly, we need to examine how its texts relate to each other through visualization. Our research uses hierarchical clustering from stylometry studies to organize texts that display similar writing patterns [47]. This approach forms groups of matching documents one step at a time until all texts appear as a tree diagram to show their connections easily. The clustering results improve when PCA reduces data complexity while maintaining essential style characteristics for analysis. Hierarchical clustering can be performed using either agglomerative ("bottom-up") or divisive ("top-down") approaches, facilitating a multi-level exploration of complex data structures. The final tree diagram displays how texts are grouped and show patterns related to language usage, sentence structure, author gender, and individual Shakespearean works. For example, our initial study grouped 19 Shakespeare play translations into clusters based on language since **Figure 1** shows this pattern while revealing smaller groupings related to the translator's gender and the characteristics of each play. Our findings demonstrate how different writing styles interact complexly and show how computational text analysis helps us better understand literature while encouraging more detailed research into writing characteristics.

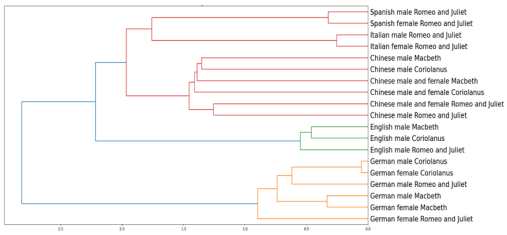

Figure 1. Hierarchical clustering analysis of *19 texts*.

4.5. Principal Component Analysis (PCA)

Through PCA, researchers can reduce data complexity and maintain important data patterns. The process involves five logical steps: First, we normalize continuous initial variables to create equal standards for feature comparison. Second, we calculate the covariance matrix to discover how variables relate. Third, the process identifies principal components by finding eigenvectors and eigenvalues in the covariance matrix, showing independent data variation dimensions. Fourth, we build a feature vector to pick the principal components with the highest variance explanation. Finally, the data transforms into an easy-to-understand ver-

sion when placed about the leading principal components. In stylometric analysis, we use Python's sklearn to perform PCA after converting text data into numerical values using TF-IDF vectorization. The PCA technique helps us find important writing patterns through a more straightforward representation. Showing principal components lets us discover what makes text different, which helps us understand writing styles better when grouping texts or assigning them to categories in literary research.

5. Experiments and Results

5.1. Stylometric Analysis of Word Frequency and Gender

The stylistic variation for the difference between male and female translators of Shakespeare's "*Romeo and Juliet*" in Italian is visually captivating in **Figure 2**, from PCA of 5-MFW and 100-MFW. When data points cluster tightly, they show similar writing styles, but when these dots mix freely, it demonstrates that the writing styles have minimal differences [48]. The principal components reveal that male and female translators display distinct styles, as shown by the two colored lines. The Euclidean distance analysis between data pairs confirms our findings. Our analysis of 5-MFW and 100-MFW shows minor but present differences in writing style between male and female translators with distances of 0.62 and 0.63. The standard deviation measurements of 0.51 show up in both data sets. The research shows that the book's Italian translations reveal distinct styles and techniques that male and female translators share.

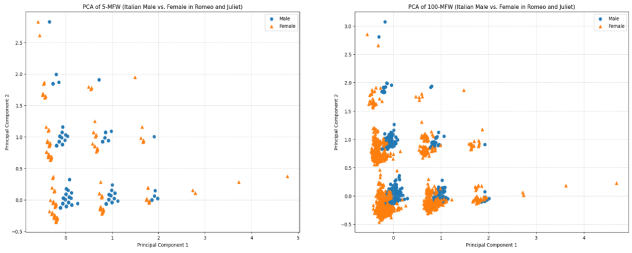

Figure 2. PCA of 5-MFW and 100-MFW of Italian male and female translators.

The research team increased their analysis to include more languages and more extensive selections of Shakespeare's plays in their stylistic study. The research team used two methods of word frequency analysis (5-MFW and 100-MFW) to measure the stylistic differences between male and female translators and their combined stylistic traits shown in **Table 2**. The study found that using 100 words to measure style showed more apparent differences than using five words, suggesting that longer word lists better capture style details. As a small example, in German, although the mean distance for "*Macbeth*" in the 5-MFW analysis was 0.63, this value grew to 0.69 in the 100-MFW analysis. Likewise, the mean distance of 5-MFW for "*Romeo and Juliet*" in Spanish increased slightly more than 0.05 (from 0.85 to 0.87) at 100-MFW.

Our analysis in **Figure 3** shows that using 100 MFW helps us see more detailed writing style patterns. The point plot emphasizes variances across plays and languages: While some subtle stylistic differences by gender exist in "Romeo and Juliet" in Italian, these are stronger in "Coriolanus" in German, particularly under a 100 MFW dataset. The distinct results imply that male and female translators have unique writing approaches. The degree of stylistic differences differs when we compare ranges of plays with their translations at specific word frequency levels. Our findings demonstrate that gender affects translation choices as translators interpret Shakespeare's works.

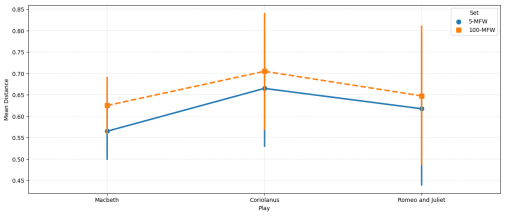

Figure 3. Comparison of mean distance for gender interactions across plays.

5.2. Stylometric Analysis of Character N-Grams and Gender

We advanced our research using sequences of five characters. These n-grams work with any language and offer simple ways to reveal distinct patterns in how authors write their text, according to Young et al. [49]. Author attribution research benefits from this technique because it remains effective when analyzing texts from different genres [50]. The graph shows that male and female translators, plus mixed teams, produce work that closely resembles Shakespeare's original text when using character 5-grams. Take, for example, the top 10 character 5-grams for the English Shakespeare source text: Most are made up of common English words and

Table 2. 5-MFW and 100-MFW pairwise Euclidean distances.

	Play	Gender	Language	Mean Distance	Standard Deviation	Min Distance	Max Distance
5-MFW	Macbeth	M vs. F	German	0.63	0.60	0.03	8.43
		M vs. M&F	Chinese	0.50	0.75	0.00	9.76
	Coriolanus	M vs. F	German	0.53	1.96	0.04	126.93
		M vs. M&F	Chinese	0.80	1.35	0.00	27.41
	Romeo and Juliet	M vs. F	German	0.62	0.51	0.05	4.8
		M vs. F	Italian	0.62	0.51	0.06	5.68
		M vs. F	Spanish	0.85	0.60	0.01	10.46
		M vs. M&F	Chinese	0.38	0.62	0.00	9.51
100-MFW	Macbeth	M vs. F	German	0.69	0.61	0.00	10.57
		M vs. M&F	Chinese	0.56	0.86	0.00	0.56
	Coriolanus	M vs. F	German	0.57	3.23	0.00	212.95
		M vs. M&F	Chinese	0.84	1.39	0.00	28.74
	Romeo and Juliet	M vs. F	German	0.66	0.49	0.00	4.68
		M vs. F	Italian	0.63	0.51	0.00	5.62
		M vs. F	Spanish	0.87	0.61	0.00	11.64
		M vs. M&F	Chinese	0.43	0.79	0.00	19.63

phrases, such as "the" and "and." Conversely, the male target text is filled with German words like "und" and "ich," and the female one with a mixture of maybe German and some Spanish or Italian words, for example, "que" and "und." The primary tokenization method uses character names, which creates translation challenges because the Chinese 5-grams hold cultural and linguistic meanings based on gender.

Pairwise Euclidean distances between Shakespeare's original works and their translations regarding genders and languages are quantitatively represented by **Table 3**. Interestingly, the mean distances for translation to German prove that male translators (0.70) are closer to "*Macbeth*" than their female counterparts (0.74). However, "*Coriolanus*" suggests that female German translators (0.62) are closer to Shakespeare's original than the men (0.64). Somewhat surprisingly, the Italian and Spanish translations of "*Romeo and Juliet*" show a trend that male translators tend to translate to a shorter mean distance from Shakespeare than do female translators, suggesting a closer stylistic alignment. While the gender difference is in Chinese translations, the "*Macbeth*" and "*Romeo and Juliet*" range to around 0.40–0.52 for males and both genders. The range of translation styles appears in the standard deviations and extreme values statistical data. **Figure 4** shows visually what these translation numbers tell us about how gender affects translation choices when working with Shakespeare's texts.

5.3. Stylometric Analysis of Lexical Richness and Gender

The Type-Token Ratio (TTR) serves as a vocabulary diversity measurement, according to Richards[51]. The value of this ratio indicates stylistic choices authors make: A better ratio shows a translator uses varied words, while a lower ratio shows they repeat words more often. A consistent pattern emerges: Female translators show better vocabulary diversity than male translators, and both genders fall behind in combined gender translation results. Macbeth translations show a TTR of 0.2149 for female translators and 0.1967 for male translators. Gender-based lexical variation becomes apparent when examining **Figure 5**'s point plots, which display vocabulary wealth for each play. The code can be found in the supplementary material named *Python Notebook*.

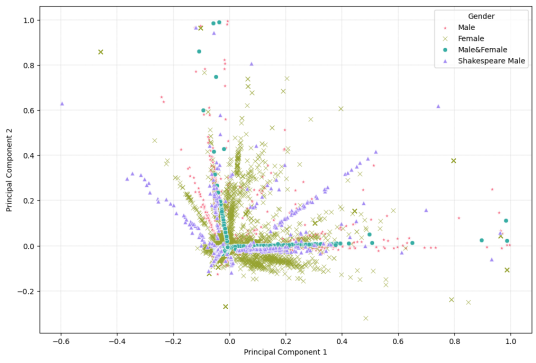

Figure 4. PCA of character 5-grams grouped by gender.

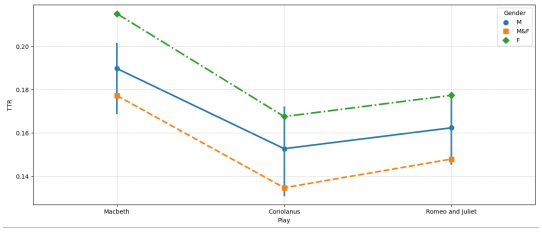

Figure 5. Point plot of TTR by play and gender.

Table 3. Character n-gram pairwise Euclidean distances between Shakespeare and translations.

Play	Gender	Language	Mean Distance	Standard Deviation	Min Distance	Max Distance
Macbeth	S vs. M	German	0.70	0.76	0.00	7.28
	S vs. F	German	0.74	0.90	0.00	8.96
	S vs. M	Chinese	0.43	0.67	0.00	6.97
	S vs. M&F	Chinese	0.52	0.62	0.00	6.24
Coriolanus	S vs. M	German	**0.64**	**4.29**	**0.00**	**250.58**
	S vs. F	German	**0.62**	**2.14**	**0.00**	**113.9**
	S vs. M	Chinese	0.46	2.32	0.00	112.47
	S vs. M&F	Chinese	0.71	2.15	0.00	112.65
Romeo and Juliet	S vs. M	German	0.57	0.66	0.00	7.84
	S vs. F	German	0.68	0.62	0.00	5.94
	S vs. M	Italian	0.43	0.62	0.00	6.53
	S vs. F	Italian	0.63	0.66	0.00	6.85
	S vs. M	Spanish	0.50	0.60	0.00	6.49
	S vs. F	Spanish	0.67	0.64	0.00	9.91
	S vs. M	Chinese	0.36	0.50	0.00	6.52
	S vs. M&F	Chinese	0.44	0.56	0.00	6.30

5.4. Sentence Length and Gender Differences

Figure 6 shows the frequency of sentences used by male and female translators and both genders for the German translation of Shakespeare's "Coriolanus." The results clearly show how the two groups differ. Female translators in the play write sentences that average 11.2 words and vary widely, with a standard deviation of 12.84. The wordiest female monology is 106 words. Male translators construct longer sentences averaging 29.87 words and show wider sentence variation with a standard deviation of 46.86. Male translators achieve the maximum sentence length of 468 words, showing that male translators in the German translation of 'Coriolanus' have an extended dialog of their female counterparts.

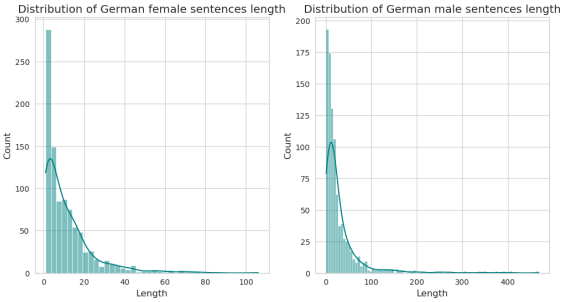

Figure 6. Sentence length distribution of German female and male translators.

Our analysis in **Table 4** reveals how translators' gender affects sentence length when reinterpreting Shakespeare's plays. Our data shows that translation sentence length differs between plays and languages, meaning the translator's gender and chosen language impact their translation style. For example, in the German translations of "*Macbeth*," female translators wrote sentences on average 23.09 words, and male translators 14.39 words, much less than the English source text average of 22.93. This trend of gender disparity is consistent across plays: Male translators in "*Coriolanus*" produced longer sentences in German than did females, and in "*Romeo and Juliet*," females produced shorter sentences in German, but less pronounced differences were observed between genders in Italian and Spanish.

5.5. Syntactic Trees and Gender Differences

Bird, Klein and Jakobson show that researchers can understand authorship better by studying word placement than they once believed [52, 53]. Our earlier discussion showed that modern natural language processing systems like space inspired the name trees because they help sentences become more detailed or direct representations. Looking at translation work helps us learn important things. We can discover how translators modify meaning from the original text based on their cultural views and gender beliefs when comparing texts with their translations. Examining sentence patterns by counting syllables lets us explore how people use language and verify translation quality, including gender representation evaluation. The code can be found in the supplementary material named *Python Notebook*.

Figure 7 shows the synthetic tree features distributed

Table 4. Sentence length across languages and translator gender.

Play	Gender	Language	Count	Mean	Standard Deviation
Macbeth	M	**English**	**789.00**	**22.93**	**20.94**
	F	German	755.00	23.09	20.17
	M	German	755.00	14.39	18.38
Coriolanus	M	**English**	**429.00**	**69.64**	**92.07**
	F	German	938.00	11.20	12.84
	M	German	938.00	29.87	46.86
Romeo and Juliet	M	**English**	**1604.00**	**15.87**	**23.46**
	F	German	1290.00	8.12	4.04
	M	German	1290.00	11.28	11.76
	F	Italian	822.00	37.45	36.00
	M	Italian	822.00	20.38	18.06
	F	Spanish	1115.00	21.31	18.55
	M	Spanish	1115.00	21.47	22.79

across the two main components in a scatter plot. The position of each point on the map shows us valuable information about sentence structure patterns. The translation work from Spanish, Italian, German, and Chinese shows that male and female translators follow similar sentence structures. The traditional Spanish translation landscape shows no clear separation between how men and women translate these three languages. The translated texts differ from the original English Shakespeare based on gender, highlighted in green. The unique gender patterns in translation reveal the deep complexity of this work and make the relationship between language and gender roles hard to understand. Different male and female translators of a single language show slight syntax variation in their translations. Our research reveals important style variations between Shakespeare's original work and its target texts.

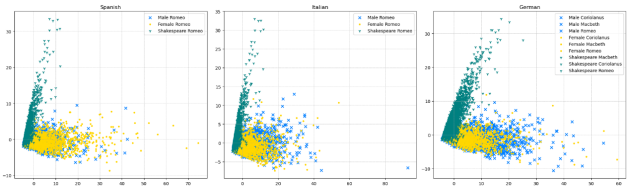

Figure 7. Syntactic tree features of source text vs. male and female translations.

5.6. Metrical Pattern and Gender Differences

Prologue to Shakespeare's "Romeo and Juliet," shown below in **Table 5**. In line with Shakespeare's sonnet style, the Chorus delivers this opening section. A Shakespearean sonnet consists of 14 lines arranged into iambic pentameter meter. The lines follow a regular ten-syllable pattern of unstressed and stressed syllables. According to **Table 5**, the poem uses the ABABCDCDEFEFGG rhyme sequence. Shakespeare's use of measured rhythm is illustrated in **Figure 8**, which provides a visual representation of the rhythmic patterns. To further enhance understanding, additional illustrations of rhythm representations, such as stress patterns and metrical variations, could be included to highlight the natural metrical sequences and rhythmic shifts that Shakespeare employs. These rhythmic changes, analyzed through text, amplify the emotional impact of the narrative, telling the tragic story of two forbidden lovers whose suffering stems from their families' longstanding feud. The Python package, utilizing its syllables library, assists in analyzing the poem's rhythm patterns and variations by counting syllables for each line, offering a deeper insight into Shakespeare's metrical craftsmanship.

Table 5. The sonnet at the beginning of Romeo and Juliet.

Sonnet	Rhyme
Two households, both alike in dignity,	A
In fair Verona, where we lay our scene,	B
From ancient grudge breaks to new mutiny,	A
Where civil blood makes civil hands unclean.	B
From forth the fatal loins of these two foes	C
A pair of star-crossed lovers take their life;	D
Whose misadventur'd piteous overthrows	C
Doth with their death bury their parents' strife.	D
The fearful passage of their death-mark'd love,	E
And the continuance of their parent's rage,	F
Which but their children's end naught could remove,	E
Is now the two hours traffic of our stage;	F
The which, if you with patient ears attend,	G
What here shall miss, our toil shall strive to mend.	G

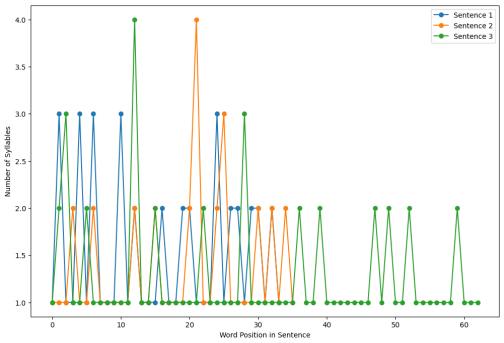

Figure 8. Metrical patterns in Shakespeare's sonnet.

This analysis shows the poetic structures of the Spanish translations for both male and female texts in **Figure 9**. Spanish poetry uses different syllable counts for each verse type, unlike English. The Spanish versions show more rhythmic variety than Shakespeare's original work. In order to view the differences between the male and female translations, we have to observe 'patterns.' The Spanish translations of the two versions match sound and beat but differ in rhythm stability between genders. These differences probably result from the translators' personal style choices. Spanish rhythmic tendencies and the translator's commitment to the original story explain these translation differences.

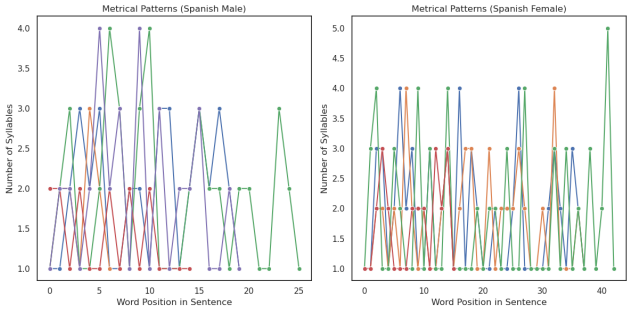

Figure 9. Metrical patterns in Spanish male and female translations.

Finally, our analysis reveals that Shakespeare's text and translations primarily use single syllables at 76.6%, with 16.5% double syllables and 5.6% triple syllables in **Figure 10**. The male translations show more syllable variation, using 64.3% single syllables, 22.0% double syllables, and 10.4% triple syllables. Female translations give similar results. The female translations distribute syllables into 62.9% single syllables, 22.7% double syllables, and 10.6% triple syllables. The rest of our discussion will be devoted to showing how this shift in metrical structure—which is especially evident in the translations—evokes the tensions of translating Shakespeare'sShakespeare's particular iambic pentameter into different languages, each of which has its rhythmic quirks. The slight differences between translations show that translators consciously selected their methods to reach metrical results. The differences show us how hard it is to translate because translators need to balance keeping the original content with adapting the language for new audiences.

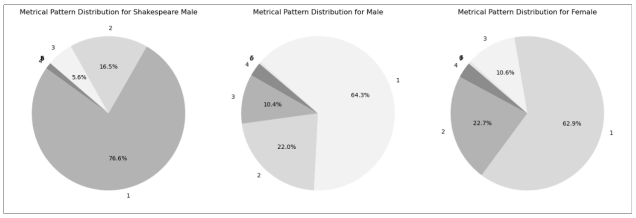

Figure 10. Metrical patterns across Shakespeare's original text and its translations.

6. Discussion

The findings of this study provide significant insights into the influence of gender on the stylistic choices made by translators of Shakespeare's works across multiple languages. This research used data analysis techniques to show how male and female translators have different stylistic approaches and reveal how gender affects translation work. This research shows how translation reflects gender differences and proves computational linguistics can detect subtle details traditional analysis methods miss.

6.1. RQ1: How Do Male and Female Translators Differ in Their Lexical, Syntactical, and Metrical Choices When Translating Shakespeare's Works?

The analysis reveals gender-based differences in lexical, syntactic, and metrical choices. Female translators consistently demonstrated more extraordinary lexical richness, as evidenced by higher TTRs, suggesting a preference for diverse vocabulary and creative language use. This aligns with previous research indicating that female translators often prioritize nuanced interpretations and detailed renderings of source texts[3, 8]. In contrast, male translators exhibited stronger fidelity to Shakespeare's iambic pentameter, particularly in Spanish translations, indicating a tendency to preserve the original metrical structure. This finding underscores the role of gender in shaping translation strategies,

with female translators leaning toward interpretive flexibility and male translators emphasizing structural fidelity. The syntactic analysis further highlighted gender differences, with male translators producing longer sentences in German translations of "Coriolanus," while female translators favored shorter, more varied sentence structures. This divergence may reflect differing approaches to narrative pacing and emphasis, with male translators potentially prioritizing rhetorical complexity and female translators focusing on clarity and accessibility. These findings align with broader trends in gender and translation studies, which suggest that gendered linguistic practices influence how translators reconstruct narratives and characters [2, 54].

6.2. RQ2: To What Extent Can Computational Methods, Such as PCA and Hierarchical Clustering, Uncover Gendered Stylistic Distinctions in Literary Translations?

The application of computational methods, notably PCA and hierarchical clustering, proved highly effective in uncovering gendered stylistic distinctions. PCA successfully differentiated between male and female translators based on lexical and syntactic features, as demonstrated by clustering data points in scatter plots. Hierarchical clustering further revealed that translations grouped primarily by language but exhibited sub-clusters based on translator gender, highlighting the interplay between linguistic and gendered stylistic choices. For instance, the analysis of MFW and character n-grams showed that male and female translators employed distinct stylistic signatures, with female translators often displaying more significant variability in word choice and sentence structure. These findings validate the utility of computational stylometry in identifying subtle, gender-based patterns in translation, offering a quantitative complement to traditional qualitative approaches. The success of these methods also underscores the potential for interdisciplinary research that bridges computational linguistics and literary studies.

6.3. RQ3: How Do These Gendered Choices Reflect Broader Translation Theory and Practice Trends?

The gendered stylistic choices observed in this study reflect broader trends in translation theory and practice, particularly the evolving role of gender as a lens for understanding translation dynamics. Female translators' emphasis on lexical diversity and creative interpretation aligns with feminist translation theory, which advocates for the visibility of the translator's voice and the subversion of patriarchal norms in source texts [8, 15]. Conversely, male translators' adherence to metrical structures and longer syntactic forms may reflect a more conservative approach, prioritizing fidelity to the source text's formal elements. These findings also resonate with contemporary debates on gender-inclusive translation practices, particularly in the context of Shakespeare's works, which often grapple with complex gender roles and relationships. The study highlights how translators' gendered perspectives can either reinforce or challenge traditional gender norms, offering new interpretations of classic texts for modern audiences.

6.4. Implications for Translation Studies and Beyond

This study has several implications for translation studies, gender studies, and computational linguistics. First, it demonstrates the value of interdisciplinary approaches in uncovering the nuanced interplay between gender and translation. By combining computational methods with literary analysis, this research provides a robust framework for exploring stylistic patterns in large text corpora. Second, the findings challenge traditional assumptions about translation as a neutral or objective process, highlighting the pervasive influence of translators' gendered perspectives. This has important implications for translation pedagogy and practice, encouraging greater awareness of how gender shapes linguistic choices and textual outcomes. Finally, the study contributes to ongoing discussions about the role of technology in the humanities. The successful application of PCA and hierarchical clustering in this research underscores the potential of computational tools to enhance our understanding of literary texts and their translations. As computational methods evolve, they offer exciting opportunities for future

research at the intersection of gender, language, and technology.

7. Conclusions

7.1. Major Findings

Studying the differences in style in how drama is translated shows us that men and women use language differently. The translation work of men and women shows distinct patterns in word group usage and vocabulary depth. Women who translate Shakespeare to German choose to write with shorter sentence structures. The sentence structure analysis shows that translation differences come from the translator's gender, cultural background, or language rules. Shakespeare's iambic pentameter posed translation challenges because adapting its poetic essence to another language's musical and rhythmic patterns proved difficult. These findings show how translating Shakespeare's work involves understanding gender roles and dealing with the complex relationship between language and literature.

7.2. Limitations

This study is limited in its many aspects, which should be considered. The dataset, though extensive, is limited to translations of three Shakespearean plays, which may not fully capture the diversity of gendered translation practices across different genres and historical periods. **Figure 11** shows that Chinese translations have a much different word count than the original text and other versions. Chinese translations included in the study are brief, with fewer characters, words of a smaller average length, and less variety in word choice than other languages. This might be because we did not have enough data or because the Chinese jieba package used for analysis does not effectively capture all the language's subtleties. Our research uses PCA and hierarchical clustering to study translation methods, yet these tools might miss the detailed human choices translators make, mainly when translating gender-related style nuances. The study's results have boundaries because it uses data from only a few translators rather than a broader range of translation styles across different genders and periods. The research focuses on word patterns and sentence structure at the expense of exploring cultural translation elements. The findings need cautious interpretation because these elements reduce usefulness in different situations.

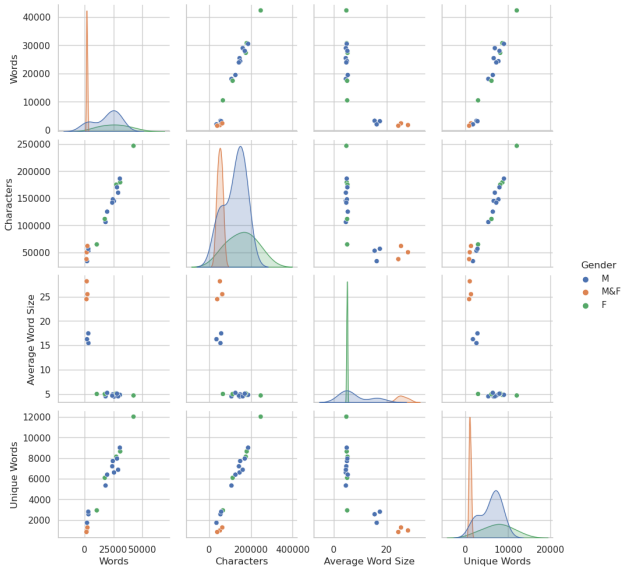

Figure 11. A pair plot for the dataset of all numeric features by gender.

7.3. Future Research Directions

Future studies should test if this study's results about translator gender influencing style choices are valid for different kinds of literature outside Shakespeare's plays. Studying more writers from different times and styles with this interdisciplinary approach will help us learn how translator gender affects translation style in many literary traditions. Testing additional languages and cultural backgrounds in our analysis will show if gendered translation habits exist everywhere or depend on specific language and cultural traits. By increasing translation samples in future research and using PCA plus clustering algorithms, researchers can gain better statistical insights into how gender affects translation methods worldwide.

Moreover, we can learn more about how translation methods are formed by studying how factors like gender theory education and translator beliefs mix with a translator's gender. The research shows how male and female translators differ in their work but leaves out how their personal and professional life may affect their translation style. Looking deeper into these elements can show us if translation styles linked to gender come naturally or are shaped by how translators learn, what they believe, or their cultural conversations about gender and translation. We need to find out if

readers react differently to texts when translators make clear decisions based on their gender.

Finally, by studying translation work over time, we can learn how translators adjust their gendered methods as society changes gender-role views. The study will review different translations of Shakespearean works made over time to see how translation methods used by male and female translators have developed. Since more people are now interested in fairness between genders in reading and translation work, this study can help us see how translating stories becomes a way to negotiate culture. It shows how both the original writing and its translation change with society's views on gender roles. Our study can help build gender-sensitive translation research that connects computer-based text analysis with social and cultural studies.

Supplementary Materials

Materials related to the article can be graciously accessed through the following link: https://osf.io/rdzpq/?view_only=97e955c6608d445c9d1c5608b6c15706

Funding

This work received no external funding.

Data Availability Statement

The author confirms that all data generated and analyzed in this study are included in this paper.

Acknowledgments

The author would like to extend sincere gratitude to anonymous reviewers, editors, and those who have supported this study.

Conflicts of Interest

There is no conflict of interest involved in this study.

References

[1] Boase-Beier, J., 2019. The translator's choices. Translation and Style, 2nd ed. Routledge: London, UK. 59.

[2] Hermans, T., 1996. The translator's voice in translated narrative. Target International Journal of Translation Studies. 8(1), 23-48. DOI: https://doi.org/10.1075/target.8.1.03her

[3] Baker, M., 2000. Towards a methodology for investigating the style of a literary translator. Target International Journal of Translation Studies. 12(2), 241-266. DOI: https://doi.org/10.1075/target.12.2.04bak

[4] Butler, J., 1990. Gender Trouble. Feminism and the subversion of identity. Routledge Press: New York, NY, USA.

[5] Gulnara, M., Akhmaral, K., Rita, B., et al., 2024. Influence of Gender-Marked Kazakh Proverbs on the Content of Cultural Stereotypes. Forum for Linguistic Studies. 6(5), 25-31. DOI: https://doi.org/10.30564/fls.v6i5.6833

[6] Alharahsheh, A.M.M., 2024. A Comparative Study of the Translated Arabic Versions of the Novel 'To the Lighthouse' under Feminist Translation Theory. Forum for Linguistic Studies. 6(5), 993-1001. DOI: https://doi.org/10.30564/fls.v6i5.6708

[7] Budilova, O., Volkova, M., Honsalies-Munis, S., et al., 2024. Gender Identity and Literary Expression: Examining the intersection of literature and sociocultural shifts. Traduction et Langues. 23(3), 111-134.

[8] Simon, S., 1996. Gender in translation, 1st ed. Routledge: London, UK. DOI: https://doi.org/10.4324/9780203202890

[9] Badenes, G.A., 2024. Model for queer drama translation analysis. Translation Studies. 17(3), 533-51. DOI: https://doi.org/10.1080/14781700.2024.2378317

[10] Bernardini, S., Onnivello, S., Lanfranchi, S., 2024. Italian normative data for the Unhelpful Thoughts and Beliefs about Stuttering (UTBAS) Scales for adults who stutter. Journal of Fluency Disorders, 81, 106074. DOI: https://doi.org/10.1016/j.jfludis.2024.106074

[11] Chai, Y.B., Li, X.D., 2024. Interpreting students' perceptions of interpreter status, market order, and educational support and their commitment to the profession: a survey in China. Onomazein. 64, 1-40. DOI: https://doi.org/10.7764/onomazein.64.01

[12] Mikros, G, Boumparis, D., 2024. Cross-linguistic authorship attribution and gender profiling. Machine translation as a method for bridging the language gap. Digital Scholarship in the Humanities. 39(3), 954-967. DOI: https://doi.org/10.1093/llc/fqae028

[13] Ishige, A., 2024. Inspired by Asian migrants: An adult English learner's imagined communities and study abroad trajectory. Linguistics and Education. 83, 101321. DOI: https://doi.org/10.1016/j.linged.2024.101321

[14] Misiou, V., Spoturno, M.L., 2024. Introduction: gender and/in drama translation. Translation Studies. 17(3), 439-444. DOI: https://doi.org/10.1080/14781700.2024.2390001

[15] Wang, H., Yu, Z.G., Wang, X.H., 2024. Expertise dif-

ferences in cognitive interpreting: A meta-analysis of eye tracking studies across four decades. Wiley Interdisciplinary Reviews-Cognitive Science. 15(1), e1667. DOI: https://doi.org/10.1002/wcs.1667

[16] Dastuyi, S.Z., Ahangar, A.A., Nourmohammadi, E., 2024. A gender-based study of interpersonal metadiscourse markers in the research papers of Iranian senior undergraduate students of translation studies. Onomazein. 65, 36-64. DOI: https://doi.org/10.7764/onomazein.65.03

[17] Stocco, M., 2024. Translation and gender in contemporary Mapuche dramaturgy. Translation Studies. 17(3), 481-497. DOI: https://doi.org/10.1080/14781700.2024.2374855

[18] López-Medel, M., 2023. The Directorate-General for Translation's gender-neutral English guidelines (1998-2022). Topics in Linguistics. 24(1), 1-17. DOI: https://doi.org/10.2478/topling-2023-0001

[19] Castro, C.G., Santos, S.L., 2024. Audiovisual translation, gender, and multilingualism: Godless or the return to a stereotyped feminist Western. Cadernos De Traducao. 44(1). DOI: https://doi.org/10.5007/2175-7968.2024.e95236

[20] van Dalen-Oskam, K., 2023. The riddle of literary quality: a computational approach. Amsterdam University Press: Amsterdam.

[21] Lee, C., 2021. How do machine translators measure up to human literary translators in stylometric tests? Digital Scholarship in the Humanities. 37(3), 813-829. DOI: https://doi.org/10.1093/llc/fqab091

[22] Duan, Y.J., Tian, J.W., 2022. 'Failed Feminism' Anne Tyler's Vinegar Girl in the Chinese Market. Critical Survey. 34(3), 68-81. DOI: https://doi.org/10.3167/cs.2022.340305

[23] Finn, K.M., 2016. Tragedy, transgression, and women's voices: The cases of Eleanor Cobham and Margaret of Anjou. Viator-Medieval and Renaissance Studies. 47(2), 277-303. DOI: https://doi.org/10.1484/J.VIATOR.5.111234

[24] Gallimore, D., 2023. Shakespearean comedy and Japanese (wo)men's Shakespeare: A refraction for the twenty-first century. Cahiers Elisabethains. 111(1), 42-57.

[25] Ingelbien, R., 2019. Single or return, ladies? The politics of translating and publishing Heine on Shakespeare. Comparative Critical Studies. 16(2-3), 181-200. DOI: https://doi.org/10.3366/ccs.2019.0326

[26] Kingery, S., 2014. Shakespeare's cross-dressing sister: The curious incident of translating gender from English to Spanish. Translation Review. 89(1), 59-72. DOI: https://doi.org/10.1080/07374836.2014.933150

[27] Sakib, N., Bhuiyan, A., Hossain, S., et al., 2022. Psychometric Validation of the Bangla Fear of COVID-19 Scale: Confirmatory Factor Analysis and Rasch Analysis. International Journal of Mental Health and Addiction. 20(5), 2623-2634. DOI: https://doi.org/10.1007/s11469-020-00289-x

[28] Zavrl, A., 2023. Who are the Addressees of Shakespeare's Sonnets (in Slovenian)? Primerjalna Knjizevnost. 46(1), 169-193.

[29] Zhu, S., 2013. Love Letter Collection of Zhus Shenghao. China Youth Press: China. 488.

[30] Liang, S., 2017. From a Cottager's Sketchbook. Yunnan People's Publishing House: China.

[31] Levith, M.J., 2004. Shakespeare in China. Bloomsbury Publishing: London, UK.

[32] Li, R., 2003. Shashibiya: Staging Shakespeare in China. Hong Kong University Press: Hong Kong.

[33] Zhang, X., 1996. Shakespeare in China: A Comparative Study of Two Traditions and Cultures. University of Delaware Press: Newark, DE, USA.

[34] Kennedy, D., Lan, Y.L., 2010. Shakespeare in Asia: Contemporary Performance. Cambridge University Press: Cambridge, UK.

[35] Shakespeare, W., 1989. William Shakespeare: the complete works. MetroBooks: New York, USA.

[36] McCarthy, J.A., 2018. Shakespeare as German Author: reception, translation theory, and cultural transfer. Brill: Leiden, the Netherlands.

[37] Pascal, R., Shakespeare, W., 1937. Shakespeare in Germany. Cambridge University Press: Cambridge, England.

[38] Williams, S., 2004. Shakespeare on the German Stage: Volume 1, 1586-1914. Cambridge University Press: Cambridge, England.

[39] Hortmann, W., Hamburger, M., 1998. Shakespeare on the German Stage: Volume 2, The Twentieth Century. Cambridge University Press: Cambridge, England.

[40] Höfele, A., 2016. No Hamlets: German Shakespeare from Nietzsche to Carl Schmitt. Oxford University Press: Oxford, England, United Kingdom.

[41] Jones, H.A., 1916. Shakespeare and Germany. C. Whittingham and Company: Chiswick, London, United Kingdom.

[42] Marrapodi, M., 2004. Shakespeare, Italy, and Intertextuality. Manchester University Press: Manchester, United Kingdom.

[43] Nulli, S.A., 1918. Shakespeare in Italia. Hoepli: Milan, Italy.

[44] Thomas, H., 1922. Shakespeare and Spain. Clarendon Press: Oxford, United Kingdom.

[45] Fischer, S.L., 2009. Boydell and Brewer. Boydell and Brewer: Woodbridge, Suffolk, United Kingdom.

[46] Hoover, D.L., 2004. Testing Burrows's Delta. Literary and Linguistic Computing. 19(4), 453-475. DOI: https://doi.org/10.1093/llc/19.4.453

[47] Eder, M., 2015. Visualization in stylometry: Cluster analysis using networks. Digital Scholarship in the Humanities. 32(1), 50-64. DOI: https://doi.org/10.1093/llc/fqv061

[48] Burrows, J., 2002. 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. Literary and Linguistic Computing. 17(3), 267-287. DOI: https://doi.org/10.1093/llc/17.3.267

[49] Young, J., Smith, P.J., Parsons, E., et al., 2019. VII Shakespeare. The Year's Work in English Studies. 98(1), 342-444. DOI: https://doi.org/10.1093/ywes/maz002

[50] Kanaris, I., Kanaris, K., Houvardas, I., et al., 2007. Words versus character N-grams for anti-spam filtering. International Journal on Artificial Intelligence Tools. 16(6), 1047-1067. DOI: https://doi.org/10.1142/S0218213007003692

[51] Richards, B., 1987. Type/Token Ratios: what do they really tell us? Journal of Child Language. 14(2), 201-209. DOI: https://doi.org/10.1017/S0305000900012885

[52] Bird, S., Klein, E., Loper, E., 2009. Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media: Sebastopol, California, United States.

[53] Jakobson, R., 1960. Linguistics and poetics. In: Sebeok, T. (ed.). Style in language. MIT Press: Cambridge, MA, USA. pp. 350-377.

[54] Wang, H., Wang, X.H., 2023. Sentiment analysis of tweets and government translations: Assessing China's post-COVID-19 landscape for signs of withering or booming. Global Media and China. 8(2), 213-233. DOI: https://doi.org/10.1177/20594364231181745