Exploring Translator Style Using Word Alignments

Yunxiao Wang

School of English Studies, Shanghai International Studies University, Shanghai, China
Email: yxwang@shisu.edu.cn
Manuscript received November 2, 2023; revised December 10, 2023; accepted December 20, 2023; March 15, 2024

***Abstract*—How the source text is rendered in the target language reflects a translator's linguistics choices, which is informative of the translator's style. Leveraging computational techniques, the present research seeks to explore translator style through word alignments derived from an entire corpus. The text material is two Chinese translations of Virginia Woolf's novel, *Jacob's Room*, by translators Pu Long and Wang Jiaxiang. Using a Transformer-based model, alignments are automatically extracted from parallel texts. A Support Vector Machine classifier is trained to test whether the alignments are indicative of the translators' styles. Chi-square feature selection is then performed to identify the most distinguishing alignments for closer examination. Results indicate that Wang favours more explicit and literal translations, while Pu utilizes more concise, diverse, and idiomatic expressions. In addition, Wang's translation is closer to the original, while Pu's is more distant. This method enables us to provide a wide range of qualitative and quantitative evidence, and also observe differences not readily apparent when examining the target text alone.**

***Keywords*—translator style, word alignment, corpus methodologies, computational linguistics**

I. Introduction

Translator style has been a widely discussed topic in the field of translation studies. With the advent of corpus linguistics, scholars have been able to analyse this subject using quantitative methods. Baker [1] characterizes translators' styles as unique "fingerprint" in their translations. As a result, forensic linguistic patterns in the target text can be examined to study a translator's style. Alongside Baker's approach, which has been practised in various studies [2–4], some scholars employ a source-oriented view, arguing that style is also manifested in a translator's response to the original text, and that the relations between the source and target texts should be considered [5–7]. Winters [8–10], for instance, focuses on particular sets of words such as speech reporting words and investigate how they are rendered in the target language by different translators. Another example is Bosseaux [11], who examines the translations of free indirect speech to investigate the translator's discursive presence. The limited range of words investigated, however, could hinder researchers from uncovering more general patterns.

As an attempt to address this issue, the present research is formulated as a computational analysis of translator style utilizing alignments extracted from all words within an entire corpus. This enables a holistic view of the relations between the source and target text, allowing more general stylistic patterns to be uncovered. Our case study focuses on two translations of Virginia Woolf's novel, *Jacob's Room*, by Chinese translators Pu Long and Wang Jiaxiang. Our objective is to investigate whether and how the two translators differ stylistically in their renditions of the original text, by observing features in the word alignments generated from their translations.

II. Data and Methods

In order to facilitate efficient analysis of word alignments, two difficulties need to be addressed. First, it can be a tedious and laborious task to manually derive alignments. Second, analysing large sets of alignments can be a challenging task in itself. Fortunately, advancements in computational linguistics have provided promising solutions to both problems. For word alignment, Transformer-based models have exhibited state-of-the-art performance, capable of automatically deriving high quality alignments. In recent years, machine learning algorithms, such as Logistic Regression (LR) and Support Vector Machine (SVM), have been widely employed for stylometric analysis of translations [12–16]. These algorithms enable efficient and reliable analysis of large numbers of textual features, including most frequent words, word n-grams and part-of-speech n-grams. For the present study, by formalizing word alignments as features, the relations between source and target texts can be systematically investigated using machine learning. The remaining parts of this section will introduce the methods and procedures in greater detail.

A. Corpus Data

The present research utilizes the original English text of *Jacob's Room* and two translations by Pu Long and Wang Jiaxiang. The text data is manually cleaned and stored in plain text format, with both translations employing the same set of UTF-8 punctuation marks. We manually align each English sentence with its corresponding translations by both translators. For tokenization, we use spaCy's [17] tokenizer for English and the pkuseg package [18] for Chinese.

Table 1 presents basic statistics of our corpus, including token, type and sentence counts. Initial examination suggests that Pu employs a more diverse lexicon and a more succinct writing style, as evidenced by the smaller number of tokens but greater number of types. As the following sections will demonstrate, our method automates the extraction of relevant evidences for close analysis, allowing more detailed insights. It also enables us to quantify aspects that may not be apparent from simple statistics, such as how close or distant a translation is from the original text.

Table 1. Basic corpus statistics

Text	Tokens	Types	Sentences
English	67,751	7,327	2,996
Pu's translation	67,338	10,425	2,996
Wang's translation	70,723	8,873	2,996

B. Word Alignment

In the field of computational linguistics, there has been a continuous effort in developing word alignment tools and algorithms. In recent years, neural network-based models

doi: 10.18178/ijlll.2024.10.2.500

have shown significant improvements in alignment quality. After reviewing most of the available methods, we settle with the one proposed by Nagata *et al.* [19], which is based on Bidirectional Encoder Representations from Transformers (BERT) [20]. This supervised approach utilizes a pre-trained large language model and requires a relatively small training dataset to achieve decent performance. Furthermore, by formalizing word alignment as a general question-answering problem, this method is relatively easy to implement. Our study uses the BertForQuestionAnswering model from the Hugging Face Transformer library. The base model is the uncased, multilingual BERT-base model, trained on top of 102 languages using a masked language modelling objective.

Nagata *et al.* [19] utilize a set of queries to convert word alignments for a sentence from a token in the L1 sentence to a span in the L2 sentence. For more details on implementation, readers are referred to the original work. For training, the GALE Chinese-English Parallel Aligned Treebank [21] is used. To evaluate the model's performance on our literary texts, a small test set is hand-labelled from our corpus, consisting of 50 sentence pairs randomly selected and labelled for both translators. In this way, two sub test sets are constructed and evaluated separately to ensure similar alignment quality for both translators. Since a model can be trained from both directions, we can combine their results by taking their union or intersection, also known as symmetrization. The models' performances are reported in Table 2 using Precision, Recall and F1. The results demonstrate that the models achieve similar and acceptable performance for both Pu's and Wang's translations. Although the Chinese-to-English model generally optimizes F1, we have decided to utilize the Union model, which yields greater Recall, allowing more potential alignments to be detected without sacrificing too much Precision.

Table 2. Model evaluation results

Model	Test set for Pu			Test set for Wang		
	P	R	F1	P	R	F1
zh→en	0.79	0.77	**0.78**	0.80	0.74	**0.77**
en→zh	0.73	0.61	0.66	0.72	0.61	0.66
Union	0.69	**0.83**	0.75	0.70	**0.83**	0.76
Inter.	**0.88**	0.55	0.68	**0.88**	0.53	0.66

C. Feature Set

In this research, we distinguish between two different feature sets. The first set contains alignment-based features, which are essentially counts of alignments. For example, an alignment between the words "eyes" and "眼睛" is represented as "eyes→眼睛". If it occurs twice in a sample, this feature takes a value of 2. The second feature set includes document-level metrics intended to represent an overall statistical description of the alignments. Their descriptions are outlined in Table III. The first three features measure the degree of alignment. Since the use of semantically similar expressions will generally result in a greater number of alignments and fewer unaligned tokens, these features can quantify an important dimension of a translator's style, that is how closely they adhere to the original. Features 4, 5, and 6 help identify whether a translator favours longer or shorter expressions to render the same source word.

D. Classification and Features Selection

For the experiment, the English text is divided into segments of at least 200 tokens, with each segment beginning and ending with complete sentences. The length of each segment varies slightly accordingly. Each sample consists of one English segment and its corresponding translation by either Pu or Wang. This configuration yields a total of 311 samples for each translator. The task at hand is a binary classification problem that utilizes word alignments as features extracted from parallel texts. By evaluating the performance of the trained classifiers, we can determine how informative the features are of Wang and Pu's stylistic differences. Furthermore, we perform feature selection to extract the most discriminating alignments for close analysis.

In this study, the Support Vector Machine (SVM) is used as the classification algorithm, due to its reliable performance reported in past experiments. During training and testing, ten-fold cross-validation is used. For feature selection, we employ the chi-square metric, which helps to eliminate features that are likely to be independent of class and therefore irrelevant for classification.

III. Results and Discussions

In this section, we present the findings of the experiments and provide qualitative interpretations. Firstly, we report the classification results obtained using alignment-based features. Then, we employ chi-square feature selection to extract the most distinguishing features, and discuss what they might reveal about the translators' styles. In the second part, we experiment with the document-level features outlined in Table 3.

Table 3. Document-level features

ID	Feature	Description
1	Number of alignments	Total number of alignments detected in a sample
2	Ratio of unaligned TT tokens	Ratio of unaligned TT (Chinese) tokens to all TT tokens
3	Ratio of unaligned ST tokens	Ratio of unaligned ST (English) tokens to all ST tokens
4	Ratio of one-to-multi align	Ratio of alignments from one ST token to multiple TT tokens to all alignments
5	Ratio of multi-to-one align	Ratio of alignments from multiple ST tokens to one TT token to all alignments
6	Avg. character-to-token ratio	The average of the ratio of Chinese characters to English tokens in all alignments

A. Experiments Using Alignment-Based Features

In the initial experiment, we train a model using all alignment-based features. To address data sparsity and exclude possible alignment errors, we discard all alignments with a document frequency of less than 5. The SVM classifier using the remaining 1,118 features achieve an accuracy of 90%, significantly higher than the random baseline of 50%. This indicates clear differences between the translators. However, subsequent feature selection shows that a significant number of the most distinguishing features correspond to the translations of character names, for

instance the main character, Jacob. While translations of names tend to be informative for classification, they do not reveal much about stylistic choices.

In an additional experiment, alignments involving proper nouns are excluded to prevent the model from focusing on translations of character names. The resulting classification accuracy is 80%, which is 10% lower than the previous result. This, however, is still significantly higher than the random baseline, indicating that the remaining features do carry stylistic information. Following the same procedure, we extracted the top 100 features, which are presented in Table 4. These features are ranked based on their level of discriminability and are divided into two groups depending on whether they appear more frequently in Pu's or Wang's translation. Upon initial examination, it can be observed that the features are skewed towards Wang, as 63 of the top 100 features appear more frequently in Wang's translation. This may suggest that Wang's translation tends to use more equivalent expressions and/or fewer variations, but further evidence is necessary to support this hypothesis.

Upon closer analysis, we make the following discovery:

Table 4. Most discriminative alignments

Translator	Alignments
Pu	1. but→但, 2. but→然而, 3. oh→噢, 4. with→与, 5. could→能, 6. like→如同, 7. the_young_man→小伙子, 8. then→随后, 9. life→人生, 10. then→接着, 11. with→跟, 12. one→人们, 13. the_window→窗户, 14. yet→但, 15. a→一_位, 16. thought→想, 17. even→即便, 18. like→犹如, 19. more→更加, 20. the_terrace→露台, 21. young_men→小伙子, 22. a→一_种, 23. was→就是, 24. oh→哦, 25. said→说_着, 26. girl→女孩, 27. it→这种, 28. perhaps→或许, 29. presumably→也许, 30. the_moors→荒原, 31. to→跟, 32. young_man→小伙子, 33. or→或者, 34. but→不过, 35. this→这种, 36. there→那里, 37. where→那里
Wang	1. but→但是, 2. or→或, 3. but→可是, 4. oh→啊, 5. could→能够, 6. and→和, 7. with→和, 8. very→非常, 9. and→以及, 10. in→在, 11. and→而, 12. one→你, 13. green→绿色, 14. then→这时, 15. white→白色_的, 16. might→可能, 17. and→并, 18. thought→想_道, 19. down→沿, 20. never→没有, 21. well→哦, 22. great→巨大_的, 23. so_that→因此, 24. make→使, 25. young→年轻, 26. as→当, 27. sighed→叹_了, 28. this→这个, 29. cried→说道, 30. in_short→总之, 31. looked_at→看_着, 32. thought→心里_想, 33. when→时候, 34. windows→窗子, 35. words→词, 36. one→人, 37. to→到, 38. alone→独自, 39. along→沿, 40. could→可能, 41. those→那些, 42. blue→蓝色, 43. eyes→眼睛, 44. may→可能, 45. when→当, 46. is→是, 47. then→然后, 48. yellow→黄色, 49. word→字, 50. can→能够, 51. each→每_一个, 52. left→离开_了, 53. looked→看, 54. spires→尖顶, 55. there_was→有着, 56. all→所有, 57. life→生活, 58. as→时, 59. yet→然而, 60. nor→也, 61. about→关于, 62. and→并且, 63. lay→躺

1) Pu's rendering of the original text tends to be more concise than Wang's

This observation is manifested in two respects. Firstly, Pu shows a preference for the shorter option when semantically equivalent choices are available, whereas Wang prefers the longer option. For the word "but", Chinese offers several rough equivalents, with the shortest being "但". Longer options include "但是", "然而", and "可是". Wang's translation features 178 instances of explicit "but" translations, with 140 of them being the regular "但是" and only 12 the shorter "但". In contrast, Pu utilizes "但" much more frequently, accounting for 61 of his 144 explicit translations of "but". Another example involves colour terms, as four of Wang's most discriminative features relate to colours, such as "white→白色的", "blue→蓝色的", and "yellow→黄色的". For a colour term such as white, two semantically equivalent options are available: "白" and "白色", where "白" represents white and "色" translates to colour. In Wang's translation, she predominantly includes the omittable "色", whereas Pu tends to leave it out. The same tendency applies to other colours such as yellow and blue, explaining why their alignments prove to be discriminating.

Secondly, regarding certain functional words, Wang tends to provide explicit translations, while Pu tends to omit them. An exemplary illustration is the conjunction "and". In the 1,816 instances of its appearance, it is not aligned to any token in Pu's translation for 1,247 times, and for 1,130 times it is unaligned in Wang's translation. In Pu's translation, it is aligned to a comma for 131 times, and only for 108 times in Wang's. In Chinese, comma can serve as a shorthand for signalling parallel relationship. Overall, Wang prefers to explicitly translate "and" using words such as "和", "以及", "而", "并" (hence why these four alignments appear at high rank in Wang's list), while Pu prefers to leave it untranslated.

In subsection B, we will further examine this observation regarding conciseness with the help of document-level features.

2) While Wang tends to use more direct and literal renderings of the original text, Pu prefers more diverse, idiomatic Chinese expressions

This tendency can be observed from various alignments involving lexical words. For example, the alignment "eyes→眼睛" is listed as one of the most discriminative features for Wang. Throughout the novel, the word "eyes" appears 77 times. In Wang's translation, it is aligned to four different expressions, and the closest Chinese equivalent, "眼睛", is used a total of 61 times. Pu's translation, on the other hand, uses a total of 12 different expressions for "eyes". He translates "eyes" into "眼睛" only 35 times, and there are 21 instances where an alignment is not identified. The comparison shows that Pu's vocabulary for "eyes" is more diverse, while Wang tends to repeatedly use the lexically equivalent option. Their differences can be better understood by looking at specific examples.

Example 1:

ST. Shading her eyes, she looked along the road for Captain Barfoot…

Pu: 她手搭凉篷 (trans: She used her hands as an awning)，沿路眺望，看巴富特上尉来了没有……

Wang: 她用手挡在眼睛上 (trans: She used her hands to cover her eyes)，顺着路看巴富特上尉来没来……

In the example above, Wang opts for a more direct and literal translation of the expression "shading her eyes", while Pu employs a metaphorical and more idiomatic Chinese expression. As a result, the literal meaning of "eyes" is not explicitly conveyed in Pu's translation, which is why the model failed to identify an alignment for "eyes".

Example 2:

ST. Then her eyes went back to the sea.

Pu: 然后，她的目光又回到海上。(Trans: Then, her vision again went back to the sea)

Wang: 然后她的眼睛又回到了大海上。(Trans: Then her eyes again went back to the sea)

Similarly, in this example, Wang once again opts for a literal translation of "eyes", whereas Pu selects "目光", which literally refers to sight or vision. This choice highlights Pu's inclination towards conveying implied meanings rather than adhering strictly to the original form.

Another case in point is the translation of "word". In Chinese, the most direct equivalent term is "词" (word) or "字" (character). "Word" appears 15 times in the English text, and Wang translates it into "字" for 7 times out of 15. In Pu's translation, however, this alignment has never appeared: he always avoids using this direct translation and chooses alternative expressions instead.

Example 3:

ST. Well, not a word of this was ever told to Mrs. Flanders

Pu: 嗯，这样的事雅各对佛兰德斯太太绝口不提 (Trans: Well, about this Jacob has never told Mrs. Flanders anything)

Wang: 唉，这些一个字也没有告诉过弗兰德斯太太 (Trans: Well, about this not a word has been told to Mrs. Flanders)

In this example, Pu avoids the literal rendering of "word" and instead uses a four-character idiom "绝口不提", meaning to never mention or speak about a certain matter. By comparison, Wang's rendering is again more literal.

Example 4:

ST. …each word falling like a disc new cut…

Pu: …一言出口，声成金石 (Trans: once a word is spoken, it sounds as if made of Jinshi)

Wang: …每一个字出来都像新灌制的唱片 (Trans: each word comes out like newly pressed discs…)

In this instance, Wang's translation preserves the simile with the direct lexical conversion of "word" and "disc". Conversely, Pu opts for a Chinese idiom "声成金石", in which "金石" (Jinshi) refers to traditional Chinese percussion instruments like Zhong that can serve as a metaphor for stunning and beautiful sounds. This example clearly demonstrates the distinct translation strategies employed by the translators: Wang chooses a more foreignizing approach, preserving information from the source text, whereas Pu leans towards domestication by conforming the text to the target culture through the use of traditional Chinese expressions and by converting certain information from the source text, such as replacing the original simile with a Chinese metaphor.

To demonstrate that the above findings are not based on isolated incidents, we provide a quantitative overview using statistical features in the next subsection.

B. Experiments Using Document-Level Features

In the subsequent experiment, document-level features described in Table 3 are used. The features are standardized by removing the mean and scaling to unit variance before being processed by the classifier. A cross-validation procedure yields an accuracy of 65%, which is only moderately better than chance, suggesting that while document-level features do capture some differences between the translators, the effect is not as pronounced as that of alignment-based features. This can be attributed to the limited number of document-level features employed in this study.

To explore the discrepancy further, we utilize a bar plot in Fig. 1 to illustrate the distribution of feature values. As the features are approximately normally distributed, we perform an independent t-test for each feature. The levels of significance are marked in Fig. 1 as asterisks following feature names. All features exhibit highly significant differences between the two groups, with most p-values being smaller than 0.001.

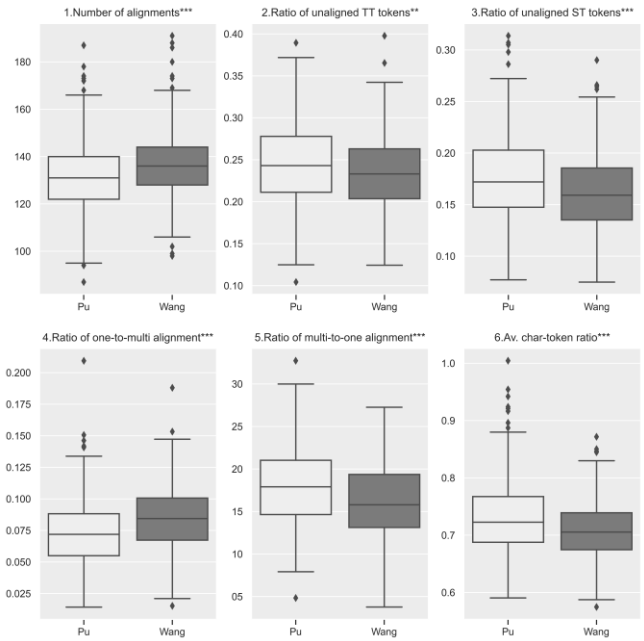

Fig. 1. Distribution of feature values in Pu's and Wang's translations. Asterisks (*) after feature names indicate levels of significance (*** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$).

Features 1, 2, and 3 each provide a perspective on the degree of alignment between the translation and the original text. In Pu's translation, a comparatively smaller number of alignments are detected than in Wang's. Additionally, with respect to Features 2 and 3, Pu's translation contains a significantly larger proportion of unaligned tokens in both the

source and target texts. These observations suggest that Wang's translation is better aligned with the original text, thereby preserving more information and remaining closer to the source. Conversely, Pu's translation appears to be, at least lexically, more distant. This conclusion supports the findings of the previous subsection, where specific examples of alignments were examined.

Feature 4, 5, and 6 provide insights into a translator's preference for concise or elaborate translations. Wang's translation displays a significantly higher proportion of alignments between one English token and multiple Chinese tokens, and conversely a smaller proportion of multi-to-one alignments. In addition, feature 6 shows that Pu's translation tends to use shorter and more contracted Chinese expressions, as evidenced by a significantly smaller average number of Chinese characters used for each English token compared to Wang's. These statistics suggest that Wang favours expanding upon source language items, while Pu prefers more succinct, contracted Chinese expressions, again confirming our previous observations about conciseness.

IV. Conclusion

In this study, we experimented with a novel approach to translator's style, using word alignments as features to model translators' different renderings of the original text. The results of our classification experiments confirm that word alignments are predicative of the translators' styles. Further analysis reveals that Wang tends to preserve the information in the original text by frequently selecting direct and lexically equivalent translations, whereas Pu often modifies the source text to conform to target culture, utilizing diverse vocabulary and idiomatic Chinese expressions. Additionally, when translating the same word, Wang prefers longer, more elaborate expressions, while Pu favours shorter and more concise ones. These observations are supported by both qualitative analysis of specific alignments and quantitative analysis based on document-level features. Notably, the difference in the level of deviation from the source text is not readily apparent when examining the target text alone, highlighting the necessity of taking into account the source text when analysing translator style.

Using word alignments as features, we are able to directly examine the translators' choices and decisions in the context of the source text. However, this method can so far only be applied in a parallel model, where translations of the same original texts are compared, since each feature carries information not only from the target, but also from the source text. In a comparable setting involving translations of different works, revealing source text information to a classifier can lead to trivial results, as identifying the source text is equivalent to identifying the translator. How this method can be modified for application in a comparable setting remains a potential direction for future research.

Conflict of Interest

The author declares no conflict of interest.

Author Contributions

Yunxiao Wang is the sole author of this paper, and is responsible for conducting the research and writing the article. All authors had approved the final version.

References

[1] M. Baker, "Towards a methodology for investigating the style of a literary translator," *International Journal of Translation Studies*, vol. 12, no. 2, pp. 241–266, 2000.

[2] M. Olohan, "How frequent are the contractions? A study of contracted forms in the translational English corpus," *International Journal of Translation Studies*, vol. 15, no. 1, pp. 59–89, 2003.

[3] G. Saldanha, "Translator style: Methodological considerations," *The Translator*, vol. 17, no. 1, pp. 25–50, 2011.

[4] D. Li, C. Zhang, and K. Liu, "Translation style and ideology: A corpus-assisted analysis of two English translations of Hongloumeng," *Literary and Linguistic Computing*, vol. 26, no. 2, pp. 153–166, 2011.

[5] J. Boase-Beier, *Stylistic Approaches to Translation*, Manchester: St: Jerome Publishing, 2006, p. 5.

[6] K. Malmkjær, "What happened to God and the angels: An exercise in translational stylistics," *Target. International Journal of Translation Studies*, vol. 15, no. 1, pp. 37–58, 2003.

[7] L. Huang and C. Chu, "Translator's style or translational style? A corpus-based study of style in translated Chinese novels," *Asia Pacific Translation and Intercultural Studies*, vol. 1, no. 2, pp. 122–141, 2014.

[8] M. Winters, "F. Scott Fitzgerald's *Die Schönen und Verdammten*: A corpus-based study of loan words and code switches as features of translators' style," *Language Matters: Studies in the Languages of Southern Africa*, vol. 35, no. 1, pp. 248–258, 2004.

[9] M. Winters, "F. Scott Fitzgerald's *Die Schönen und Verdammten*: A corpus-based study of speech-act report verbs as a feature of translators' style," *Meta*, vol. 52, no. 3, pp. 412–425, 2007.

[10] M. Winters, "Modal particles explained: How modal particles creep into translations and reveal translators' styles," *Target. International Journal of Translation Studies*, vol. 21, no. 1, pp. 74–97, 2009.

[11] C. Bosseaux, "Point of view in translation: A corpus-based study of French translations of *Virginia Woolf's to the Lighthouse*," *Across Languages and Cultures*, vol. 5, no. 1, pp. 107–122, 2004.

[12] J. Burrows, "The Englishing of Juvenal: Computational stylistics and translated texts," *Style*, vol. 36, no. 4, pp. 677–99, 2002.

[13] J. Rybicki, "The great mystery of the (almost) invisible translator," in *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*, M. P. Oakes and M. Ji, Eds. John Benjamins Publishing, 2012, pp. 231–248.

[14] H. El-Fiqi, E. Petraki, and H. A. Abbass, "A computational linguistic approach for the identification of translator stylometry using Arabic-English text," in *Proc. 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, June 2011, pp. 2039–2045.

[15] G. Lynch and C. Vogel, "The translator's visibility: Detecting translatorial fingerprints in contemporaneous parallel translations," *Computer Speech and Language*, vol. 52, pp. 79–104, 2018.

[16] C. Caballero, H. Calvo, and I. Batyrshin, "On explainable features for translatorship attribution: Unveiling the translator's style with causality," *IEEE Access*, vol. 9, pp. 93195–93208, 2021.

[17] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, *spaCy: Industrial-Strength Natural Language Processing in Python*, 2020.

[18] R. Luo, J. Xu, Y. Zhang, Z. Zhang, X. Ren, and X. Sun, "PKUSEG: A toolkit for multi-domain chinese word segmentation," arXiv preprint, arXiv:1906.11455, 2019.

[19] M. Nagata, C. Katsuki, and M. Nishino, "A supervised word alignment method based on cross-language span prediction using multilingual BERT," in *Proc. the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 555–565.

[20] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint, arXiv:1810.04805, 2018.

[21] X. Li, S. Grimes, S. Strassel, X. Ma, N. Xue, M. Marcus, and A. Taylor, *GALE Chinese-English Parallel Aligned Treebank—Training*, 2015.