

# English Linguistic Traces in Thai Translation: A Preliminary Stylometric Study of Textual Traceability in Human and LLM Translation

**Sanooch S. Nathalang\***

**Natthaphon Tripornchaisak**

Thammasat University, Thailand

\*Corresponding author's email: sanooch.s@arts.tu.ac.th

Received March 30, 2026; revised April 28, 2026; accepted April 28, 2026

## Abstract

This preliminary study examines stylometric patterns in a small corpus of English-to-Thai literary translations produced by six human translators and three LLM systems across six source passages. Human translations function both as evidence of stylometric patterns in literary translation and as reference points against which the three LLM systems are evaluated. The inclusion of multiple LLMs allows the study to examine how stylometric traces emerge in LLM translations, both in comparison with human translations and across the systems themselves. Using lexical diversity measures, sentence-based syntactic measures, Burrows' Delta, and BERTScore F1 as a semantic fidelity control, the study explores textual differences between the human translations and the outputs within this corpus. The most distinct contrasts appear in Maas values, sentence-length variance, and overall stylistic distance, while some measures show overlap across systems. Overall, the findings show that both human and LLM translations leave traceable stylometric patterns, though these are distributed differently across lexical richness, sentence-length patterning, pronominal reference, and overall stylistic distance. For English language studies, the findings offer a concrete illustration of how marked features of English literary prose are preserved, flattened, or shifted when rendered into a typologically distant language, which may be useful for translation pedagogy, post-editing, and the critical reading of AI-assisted translation.

**Keywords:** translator invisibility, linguistic traces, stylometry, LLM translation, human translation

For decades, Lawrence Venuti's critique of translator invisibility has shaped debates in translation studies. In his seminal work, *The Translator's Invisibility* (1995, 2008), he argues that the dominance of Anglophone norms prioritises fluent, domesticated translations, downplaying the translator's labour. Furthermore, successful translations are largely seen as those that do not "sound translated or foreign". These long-standing norms naturally invite what Venuti called an ideological illusion that meaning can be passed transparently from one language to another. Publishers, reviewers, and readers collectively prefer translations that do not foreground the intermediary work

of translators. In the present study, however, visibility is approached more narrowly through detectable textual traces in translated texts rather than as a direct measure of Venuti's broader socio-cultural concept. In literary translation, this matters because when the translator's mediating role is obscured, the labour of interpretation and stylistic decision-making can also disappear from view, making translation appear more direct and natural than the process actually is.

One important study that moves the discussion of translator visibility from a broad theory to observable features in the translated text is Baker (2000). She describes translator style as a textual "thumb-print" traceable through recurring linguistic patterns, including lexical items, syntactic patterns, cohesive devices, and even punctuation. Rybicki and Eder's (2010) stylometric work examines such textual patterning computationally through methods such as Burrows' Delta and most-frequent-word analysis, showing that translated texts can be investigated through measurable textual patterning. These two studies provide a methodological basis for the present study, which examines stylometric traces found in English-to-Thai literary translations by humans and LLMs.

Recent years have seen the emerging digital landscape dominated by large language models (LLMs). They have been acknowledged for many potential uses, ranging from complex tasks such as creating software applications to more fundamental tasks such as text generation, question answering, and translation. Systems such as ChatGPT, Claude, and Gemini are trained for pattern recognition and matching using an enormous number of datasets, which enables them to achieve a sophisticated understanding of language structure, meaning, and context. Many, therefore, have resorted to using LLMs as new and powerful tools for translation. This has led to growing concerns in the translation industry that AI tools may displace human translators (e.g., Kolirin, 2026). This growing use of LLMs in translation makes them an important subject of investigation.

This preliminary study investigates literary translation from English to Thai by means of a comparative stylometric analysis of six human translations and three sets of LLM translations. Using a corpus of six English source passages selected for their marked linguistic and stylistic features, together with the human and LLM translations of those passages, the study examines observable stylometric traces in the translated texts and compares them across the corpus. Given its exploratory nature, the study is guided by research questions rather than formal hypotheses. Specifically, the study seeks to answer three questions:

1. What stylometric traces can be observed in the human and LLM literary translations from English to Thai?
2. How do the stylometric traces of the human and LLM translations compare across the corpus?
3. What do these observed traces and patterns suggest about translator invisibility in literary translation from English to Thai?

## Literature Review

The theoretical and methodological background of this study can be organised into three strands: Venuti's (1995/2008) critique of invisibility as ideology, stylometric approaches to translator style, and emerging research on machine and LLM translationese.

### *Venuti's Critique of Translator Invisibility*

Venuti (1995/2008) observed that fluent translations are often praised precisely because they do not get flagged as translations. In the dominant Anglophone tradition, this fluency is frequently achieved through domestication: an "ethnocentric reduction of the foreign text to target-language cultural values" (Venuti, 1995, p. 20) that privileges transparency and readability over the preservation of linguistic and cultural difference. The resulting invisibility is ideological in two senses: it hides the fact that translation involves labour, interpretation, and negotiation rather than direct transfer, and it casts translators more as creative agents and less as professionals with limited recognition.

Venuti examines how invisibility is constructed through reviews, historical analysis of translation contracts and copyright law, among other things. He does not, however, 'measure' invisibility at the level of the translated texts. This leaves open the question of whether stylometric traces can be detected in translated texts and how far such traces may be relevant to broader discussions of translation visibility.

### *Stylometry and Translator Style*

Stylometry provides a set of quantitative tools for examining recurring linguistic patterns in texts. In translation research, it has been used to investigate whether translated texts carry traces that can be attributed more strongly to the original author, the translator, or both. One important line of research, associated with Rybicki and Eder (2010), Rybicki (2012), and Rybicki and Heydel (2013), applies methods such as Burrows' Delta and most-frequent-word analysis to translated literary corpora. These studies often find that translated texts cluster more strongly by the original author than by the translator, suggesting that stylometric analysis more readily picks up features of the original author's style than those linked to the translator's choices. At the same time, traces of the translator's own choices are not entirely absent from such analysis, though this is especially the case when the corpus is designed under tighter comparison conditions, especially when translations of a single author are examined.

Rybicki and Eder (2010) also show that Delta varies in performance across languages and is affected by linguistic differences such as the morphological properties of the target language. Because their study is based on European-language corpora, typologically more distant language pairs, such as English and Thai, remain largely underexplored. Rybicki and Heydel (2013) caution that stylometric findings should not be treated as self-sufficient proof of translator style, even when they help identify recurring textual traces.

Baker (2000) shifts attention to the translated text and asks whether recurring linguistic patterns may reveal a translator's own stylistic habits. In her exploratory corpus study, she points to preferences in lexical items, syntactic patterns, cohesive devices, and punctuation as possible places where such traces may emerge. Such orientation towards features provides a basis for examining stylistic traces in translation through observable linguistic patterning rather than through broad claims about invisibility alone. Baker's study, however, does not fully address how and under what conditions such recurring features can be counted as translator style. That line of inquiry is addressed by Saldanha (2011), who argues that translator style is not simply any repeated feature in a translation, but a recognisable and coherent pattern of choice that distinguishes one translator's work from another's across multiple texts and cannot be explained purely by the source text or linguistic constraints. The related literature indicates that the detectability of translator traces depends on corpus design and analytical focus.

The present study draws on these insights but is conducted in a narrower empirical frame. Rather than seeking to establish stable translator style in Saldanha's (2011) sense, we use a benchmark set of six human translations and three sets of LLM translations based on the same six English source passages. This makes it possible to compare whether detectable linguistic traces emerge across the corpus. Within this frame, stylometry is used here not as a definitive test of Venuti's broader concept of invisibility, but as a means of comparing textual traces in a small corpus of literary translations from English to Thai.

### ***From Corpus Studies to LLM Translation***

The stylometric and corpus-based research reviewed in the previous section focuses mainly on translators and their stylistic patterns. This raises a broader question: do translated texts differ systematically from texts originally written in the target language, and if so, what might this suggest about the nature of translation? Baker (1993) was among the first to identify tendencies that appear to characterise translated texts across language pairs, regardless of the source language. These include simplified lexical and syntactic choices, clearer sentence structures, and more conventional grammar usage. According to Baker, such tendencies are not individual translator choices but may reflect translation as a mediated communicative process.

Subsequent work lends further empirical support to this line of inquiry. Laviosa (1998), for example, examined a corpus of English narrative prose by pairing translated and non-translated texts of the same type and period. She discovered that translated texts consistently showed lower lexical density and a higher proportion of high-frequency words than texts originally written in English, labelled as "core patterns of lexical use". Bernardini (2006) addressed a similar question computationally by training a machine learning classifier to distinguish between original and translated Italian texts, which reported 86.7% accuracy. This provides further evidence that translationese can be detected statistically. Lembersky et al. (2012) further demonstrated

that language models trained on text translated into English were significantly better predictors of other translated sentences than those trained on texts originally written in English. This suggests that the translation process leaves statistical traces in translated text. It is, therefore, reasonable to ask whether different translation agents, including humans and LLM systems, also leave detectable traces distinct from one another

The previous research suggests that translated texts may exhibit recurring linguistic tendencies across corpora, leaving measurable traces in the translated text. With the advent of LLMs and the rapid shift toward adopting LLMs in the translation industry (Jiao et al., 2023; Karpinska & Iyyer, 2023), a related question is whether LLMs, as a new class of translation agent, show similar tendencies or produce a different set of traces. Early evaluations suggest that LLMs can produce fluent translated texts for language pairs with abundant training data, such as high-resource European languages. By contrast, translation quality drops considerably in language pairs with less training data or greater typological distance (Jiao et al., 2023). This is relevant to the present study, which focuses on English and Thai, a language pair that is typologically more distant and less well represented in available parallel data.

Early findings also suggest both the promise and the limitations of LLM translation. In Karpinska and Iyyer (2023), professional translators blindly annotated errors and judged sentence-level and paragraph-level translations produced by ChatGPT-3.5 across eighteen language pairs to determine which was better and why. They found that the paragraph-level translation produced fewer mistranslations, grammar errors, and stylistic inconsistencies than sentence-level translation. The most critical errors identified, however, were omissions, suggesting that human intervention remains important for preserving the original author's tone and register.

The present study adopts a stylometric approach to compare human and LLM translations, examining whether both agents leave detectable traces in the translated text and whether those traces are similar to, or distinct from, one another. Using English and Thai as the language pair in focus adds a further dimension to this comparison. Because English-Thai parallel data available for LLM training is somewhat limited, it remains to be seen how such constraints may shape LLM translations compared with human translations.

## **Research Design and Methodology**

This study uses comparative stylometry to examine textual traces in a small corpus of literary translation from English to Thai produced by human translators and LLM systems. The corpus consists of four sets of Thai translations based on the same six English source passages (see Table 2). This design allows comparison of human and LLM translations of the same six English source texts, and a closer examination of similarities and differences among the three LLM systems.

## The Source Texts

Six English literary texts were selected as source material: *Every Day* by David Levithan, *H is for Hawk* by Helen Macdonald, *The List* by Patricia Forde, *Please Stop Laughing at Me* by Jodee Blanco, *In Search of the English Eccentric* by Henry Hemming and *Bangkok Free Fall* by Richard Stevenson. The selection was purposive rather than representative as the texts were chosen to span multiple literary and narrative genres and to maximise variation in stylistically marked English features, including repetition, non-standard speech, figurative language, idiomaticity, register contrast, and embedded clause structure. High-quality human translations of these texts, validated through Master's thesis defence processes for the Translation M.A. program at Thammasat University, serve as benchmarks for comparison. The six texts span young-adult speculative romance, literary memoir/nature writing, dystopian children's fiction, personal narrative memoir, creative non-fiction, and social detective fiction. This generic spread was intended to increase the range of English linguistic features available for comparison in Thai translation. Such a variety of linguistic challenges creates the conditions necessary for detecting how human translators and LLMs handle English stylistic features differently when producing Thai translations. Table 1 summarises the key linguistic characteristics of each source text.

**Table 1**  
*Source Text Characteristics*

| Book Title                                                    | Genre                                        | Key Features                                                                                                                                                                                           |
|---------------------------------------------------------------|----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Every Day</i><br>By David Levithan                         | Young adult speculative romance              | <ul style="list-style-type: none"> <li>• Systematic repetition patterns</li> <li>• Intentional phonological devices (alliteration, assonance)</li> <li>• Syntactic parallelism structures</li> </ul>   |
| <i>H is for Hawk</i><br>by Helen Macdonald                    | Literary memoir/nature writing               | <ul style="list-style-type: none"> <li>• Figurative language</li> <li>• Technical vocabulary</li> <li>• Poetic intensity combined with factual precision</li> </ul>                                    |
| <i>The List</i><br>by Patricia Forde                          | Dystopian children's/<br>young adult fiction | <ul style="list-style-type: none"> <li>• Intentional non-standard grammar</li> <li>• Deliberate lexical restriction</li> <li>• Systematic linguistic deviance as thematic device</li> </ul>            |
| <i>Please Stop Laughing at Me</i><br>by Jodee Blanco          | Personal narrative memoir                    | <ul style="list-style-type: none"> <li>• Natural register variation (child vs. adult)</li> <li>• American cultural specificity</li> <li>• Emotional authenticity and introspective language</li> </ul> |
| <i>In Search of the English Eccentric</i><br>by Henry Hemming | Creative non-fiction                         | <ul style="list-style-type: none"> <li>• British cultural references</li> <li>• Tonal complexity (irony, sarcasm)</li> <li>• Idiomatic expressions and colloquialisms</li> </ul>                       |
| <i>Bangkok Free Fall</i><br>by Richard Stevenson              | Social detective fiction/literary memoir     | <ul style="list-style-type: none"> <li>• Sociolinguistic variation</li> <li>• Gender-marked language</li> <li>• Socially situated speech patterns</li> </ul>                                           |

### ***Human Translation Corpus***

The human corpus consists of six translations from English to Thai, each created by a different M.A. student in Translation Studies at Thammasat University as part of their thesis work. The translations were validated through a formal thesis defence. In practice, this means that each translation formed part of an M.A. thesis examined within the program's formal defence process, rather than being treated as unreviewed student output. The thesis translations were completed at different points across six years, from 2013 to 2019, predating the release of LLM systems such as ChatGPT (2022), Claude (2023), and Gemini (2023). This timing was a deliberate selection criterion to ensure that the human translations in this corpus were produced without access to the LLM tools against which they are compared. This removes a confounding factor that would otherwise make it impossible to treat the human corpus as unambiguously human-produced.

While all translations were produced within the same institutional programs, they were not produced concurrently or under standardised conditions. No controls were applied across translators for time constraints or the use of external resources. Background factors, such as prior training, language exposure, or translation experience, were not recorded. Because the present study compares translated texts rather than translator biographies or working processes, these factors were not analysed separately and remain uncontrolled variables in the dataset.

Each complete work contributed one continuous passage to the corpus, yielding six human translations of approximately 9,000 words in total. Because each translation comes from a different translator, the human corpus serves two functions in this study. First, it provides evidence of stylometric traces in human literary translation. Second, it functions as a reference set for comparing the three LLM outputs on the same passages, rather than for determining whether a single translator displays a recurring style across multiple texts. Since each human translation in this corpus is linked to a different translator and a different source text, the study cannot fully separate translator effects from source-text or genre effects. Observed differences between human and LLM translations may, therefore, reflect source-text characteristics, such as genre, register, or linguistic complexity, as much as they reflect translators' individual choices. Accordingly, the findings are interpreted as descriptive tendencies within this dataset and not as evidence of stable, translator-specific stylistic characteristics.

### ***LLM Translation Corpus***

The same six English source passages were translated by three LLMs: GPT-4o mini (OpenAI), Claude Sonnet 4 (Anthropic), and Gemini Pro 2.5 (Google). All translations were produced in July 2025 using premium subscription access, namely, ChatGPT Plus, Claude Pro, and Gemini Advanced, so that each passage could be processed in a single prompt.

All three LLM systems received the same prompt, so that differences in the translated outputs are less likely to reflect variations in prompt wording: “Translate the following English text into Thai, preserving its original tone, style, and register. The text is meant to inform, persuade, or narrate, so the Thai version should reflect that intention. Do not neutralize emotional tones or rhetorical emphasis. Produce only the Thai translation without explanations.”

Table 2 summarises the distribution of translation across the human corpus and the three LLM systems.

**Table 2**  
*Corpus Composition and Data Distribution*

| Source                        | Human  | GPT-4o mini | Claude Sonnet | Gemini Pro 2.5 | Total   |
|-------------------------------|--------|-------------|---------------|----------------|---------|
| Six English literary passages | 6      | 6           | 6             | 6              | 24      |
| Total Word Count              | ~9,000 | ~9,000      | ~9,000        | ~9,000         | ~36,000 |

### ***Sample Size Justification***

Each translation sample was limited to approximately 1,500 words. This length allowed a continuous passage to be taken from a single coherent scene or chapter while keeping the same source excerpt across the human and LLM translations. Ishihara (2017), for example, compares 500-, 1,000, 1,500-, and 2,500-word samples in stylometric text comparison. The study shows that 1500 words is a workable span, while larger samples yield stronger performance. All human and LLM translations were based on the same English excerpt as their source. In the present study, six such passages together form each translation set of approximately 9,000 words, so the analysis considers recurring patterns across the full set of translations rather than relying on any single passage. The findings are therefore presented as patterns within this corpus, not as general claims about human or LLM translation as a whole.

### ***Stylometric Analysis Framework***

The analysis combines measures of lexical diversity, syntactic patterning, overall stylistic distance, and semantic similarity. Lexical and sentence-based measures are used to identify recurring textual tendencies in the translations, while Burrows’ Delta is used to compare overall stylistic distance across texts. BERTScore F1 is included as a semantic check so that differences in stylometric patterning can be interpreted alongside a measure of semantic proximity to the human translations. As Thai is written without spaces between words, all texts were tokenised before any measures were applied. Each component is described in the sections that follow.

### ***Thai Language Processing***

Because Thai is written without spaces between words, all texts have to be split into individual words before any analysis can be conducted. The newmm tokeniser from the PyThaiNLP library (Phatthiyaphaibun et al., 2023) was used to identify word boundaries by matching text against a Thai dictionary. As different tokenisers can make different decisions about word boundaries, stylometric results must be linked directly to differences in translation style and not to tokeniser decisions. To check that the results were not dependent on this particular tool, Attacut, a second tokeniser, was used to retokenise the texts, and Burrows' Delta was then applied to confirm whether the same results emerged regardless of how the word boundaries were drawn. The two tokenisers produced the same overall pattern, suggesting that the human-LLM differences reported here are not simply an artifact of tokenisation.

### ***Lexical Trace Measures***

Four metrics were used to assess lexical diversity across the translated texts, with TTR and CTTR for vocabulary variety, and Hapax Legomena Percentage and the Maas Index for lexical richness. TTR was included as a conventional baseline measure of vocabulary variety because it remains widely used and easily interpretable. At the same time, because TTR is known to be sensitive to text length, it was interpreted alongside CTTR, Hapax percentage, and the Maas Index rather than used on its own (Koizumi & In'nami, 2012). This is also consistent with stylometric work showing that vocabulary-richness features are most useful when combined with other features rather than treated as sufficient by themselves (Ishihara, 2017).

The Type-Token Ratio (TTR) divides the number of unique words by the total number of words in a text. A text where most words appear only once tends to score higher, while a text where the same words recur more frequently tends to score lower. The Corrected Type-Token Ratio (CTTR) works on the same principle; however, it accounts for the fact that longer texts naturally receive lower TTR scores because of a higher chance of repetition. These two measures provide an initial indication of vocabulary variety across the corpus.

The Hapax Legomena Percentage counts the words that appear only once in a text. A higher Hapax percentage indicates a wider distribution of non-repeated vocabulary throughout the passage. A lower Hapax percentage, on the other hand, suggests greater reuse of the same lexical items. In this study, Hapax is treated as an indicator of lexical dispersion rather than as a direct measure of creativity. The Maas Index approaches lexical richness from a different angle. It tracks whether a translator's vocabulary diversifies or becomes more repetitive as the text progresses, as the translator begins to reuse words already introduced. A lower Maas value indicates that new words

continue to appear across the passage, while a higher value suggests that the vocabulary pool narrows early and the same words begin to recur.

### ***Syntactic Trace Measures***

Two measures were used to examine sentence-level patterning across the Thai translated texts: average sentence length and sentence-length variance. Average sentence length indicates whether a translation tends towards longer or shorter constructions. Sentence-length variance measures how much sentence length fluctuates around the average.

For each translated text, the length of every sentence was counted in words. The average sentence length was then computed as the mean of these values across the passage. Sentence-length variance was calculated to show how much individual sentence lengths deviated from the mean. Both measures were computed in *R* and compared across human and LLM translations. A high variance score indicates that the translator moves between short and long sentences across the passage. A low score shows that sentences remain roughly uniform in length throughout the passage.

### ***Stylistic Distance and Semantic Check***

Overall stylistic distance between texts was measured using Burrows' Delta in the *stylo* package in *R*. Burrows' Delta is a well-established measure in stylometry and has been widely used in authorship attribution (Burrows, 2002; Eder, 2015). The implementation in *stylo* follows the established stylometric practice of comparing texts through the frequencies of the most frequent words, and later work has examined in detail why Delta performs effectively under those conditions (Eder et al., 2016; Evert et al., 2017). Delta works by building a frequency profile for each text and recording how often these common words appear. It then compares those profiles between texts to produce a single distance score. Lower scores indicate greater stylistic similarity. In the present study, Delta is used to compare how closely the three LLM systems resemble one another and how far each stands from the human translations.

BERTScore F1 was used as a semantic check, with the human translations serving as the reference. In simple terms, it estimates how closely two texts match in meaning by comparing words in context rather than looking only for the exact same wording. Zhang et al. (2020) show that BERTScore correlates strongly with human judgements in machine translation evaluation, and they identify the F1 version as the most reliable variant across settings. In the present study, BERTScore F1 is used to complement the stylometric analysis by showing whether an LLM translation remains close in meaning to the human translation even when the wording or sentence structure differs.

## Statistical Analysis

For the primary comparison between human and the three LLM translations, the analysis used descriptive statistics and Cohen's *d* effect sizes to report how large the observed difference is within this dataset. For comparisons among the three LLM systems, one-way ANOVA was used for TTR, BERTScore F1, average sentence length, and sentence-length variance, followed by Tukey HSD post hoc tests, where appropriate. These analyses are used to identify patterns within the corpus rather than to support broad generalisations beyond this dataset. All analyses mentioned above were conducted in *R*. Table 3 summarises how each research question is addressed through the main analytical procedures used in the study.

**Table 3**

*Research Questions (RQs) and Analytical Procedures*

| RQ | Main data                                                                                                                                                                                         | Main tools                                                                                                                                                                                                                                       | Aim                                                                                 |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 1  | <ul style="list-style-type: none"> <li>• Human,</li> <li>• GPT-4o mini,</li> <li>• Claude Sonnet 4, and</li> <li>• Gemini Pro 2.5 translations of the same six English source passages</li> </ul> | <ul style="list-style-type: none"> <li>• TTR,</li> <li>• CTTR,</li> <li>• Hapax %,</li> <li>• Maas Index,</li> <li>• average sentence length,</li> <li>• sentence-length variance,</li> <li>• Burrows' Delta,</li> <li>• BERTScore F1</li> </ul> | Identify observable stylometric traces                                              |
| 2  | Comparison across the four translation sets                                                                                                                                                       | <ul style="list-style-type: none"> <li>• Descriptive statistics,</li> <li>• Cohen's <i>d</i>,</li> <li>• Burrows' Delta</li> <li>• ANOVA,</li> <li>• Tukey HSD</li> <li>• BERTScore F1</li> </ul>                                                | Compare similarities and differences across the corpus                              |
| 3  | Quantitative results plus selected textual examples                                                                                                                                               | Qualitative interpretation                                                                                                                                                                                                                       | Discuss what observed traces and patterns may suggest about translator invisibility |

## Results

### *Lexical and Syntactic Divergence*

Table 4 presents the stylometric results across all metrics grouped by vocabulary variety, lexical richness, and syntactic patterning.

**Table 4**

*Stylometric Trace Indicators: Human vs. LLM Translation*

| Trace Indicator                 | Human Mean | LLM Mean |        |        | Effect Size (d) vs. Human |        |        |
|---------------------------------|------------|----------|--------|--------|---------------------------|--------|--------|
|                                 |            | GPT-4o   | Claude | Gemini | GPT-4o                    | Claude | Gemini |
| Vocabulary variety              |            |          |        |        |                           |        |        |
| Lexical Diversity (TTR)         | 0.405      | 0.486    | 0.407  | 0.406  | -1.24                     | -0.04  | -0.02  |
| Corrected TTR (CTTR)            | 11.86      | 11.75    | 10.85  | 11.10  | 0.06                      | 0.60   | 0.45   |
| Lexical richness                |            |          |        |        |                           |        |        |
| Rare Word Usage (Hapax %)       | 0.265      | 0.324    | 0.254  | 0.260  | -1.03                     | 0.28   | 0.14   |
| Lexical Complexity (Maas Index) | 0.016      | 0.183    | 0.199  | 0.198  | -11.33                    | -19.41 | -19.49 |
| Syntactic patterning            |            |          |        |        |                           |        |        |
| Average Sentence Length         | 20.62      | 18.18    | 15.56  | 15.28  | 0.51                      | 1.31   | 1.45   |
| Sentence-Length Variance        | 287.39     | 227.82   | 154.33 | 133.17 | 0.29                      | 0.90   | 1.07   |

On vocabulary variety, human translations score 0.405 on TTR. GPT-4o scores 0.486, the highest of the three LLM systems. Claude and Gemini score 0.407 and 0.406, respectively, which are nearly identical to the human mean. The difference between GPT-4o and Claude and Gemini is statistically significant (Tukey HSD,  $p = 0.045$ , see Table 5). On CTTR, human translations score 11.86 above Claude at 10.85, and Gemini at 11.10, while GPT-4o at 11.75 is closer to the human mean. On TTR, the human mean aligns with Claude and Gemini rather than GPT-4o, but on CTTR, it aligns more with GPT-4o than Claude and Gemini.

On lexical richness, human translations score 0.265 on Hapax percentage, while GPT-4o, Claude, and Gemini score 0.324, 0.254, and 0.260, respectively. Human translations score lower than GPT-4o but higher than Claude and Gemini. On the Maas Index, human translations score 0.016, while the LLMs cluster together from 0.183 to 0.198. The human mean is considerably lower than all three LLMs despite the mixed alignment in the TTR results.

On syntactic patterning, the average sentence length of human translations is 20.62, while that of GPT-4o, Claude, and Gemini is 18.18, 15.56, and 15.28, respectively. Sentence-length variance of human translations is 287.39 while that of GPT-4o, Claude, and Gemini is 227.82, 154.33, and 133.17, respectively. Human translations show greater fluctuation in sentence lengths than all LLM translations, with Claude and Gemini showing the lowest sentence-length variance. ANOVA

results show no significant differences in average sentence length ( $F(2,15) = 1.16$ ,  $p = 0.339$ ) or sentence-length variance ( $F(2,15) = 0.77$ ,  $p = 0.486$ ) across the three LLM systems.

### ***Holistic Stylistic Distance***

Table 5 presents Burrows' Delta scores for two types of comparison: pairs of LLM translations and pairs of human and LLM translations.

**Table 5**

*Burrows' Delta Stylistic Distance Summary*

| Comparison Group                                | Average Delta Score | Range (Min – Max) |
|-------------------------------------------------|---------------------|-------------------|
| Between LLM Systems (e.g., Claude vs. Gemini)   | 0.70                | 0.29 – 1.13       |
| Between Human and LLMs (e.g., Claude vs. Human) | 0.82                | 0.46 – 1.20       |

The average Delta between LLM systems is 0.70 (range: 0.29–1.13), indicating relatively close stylistic similarity among the three systems. By contrast, the average Delta between the human translations and LLM translations is 0.82 (range: 0.46–1.20), suggesting that the human output is more stylistically distant from each LLM than the LLMs are from one another. The inter-LLM mean is derived from the three pairings GPT-4o/Claude, GPT-4o/Gemini, and Claude/Gemini, while the human-LLM mean is derived from Human/GPT-4o, Human/Claude, and Human/Gemini.

### ***Algorithmic Signatures Among LLM Systems***

Table 6 presents the Tukey HSD post-hoc results for inter-LLM comparisons on two metrics: TTR and BERTScore F1.

**Table 6**

*ANOVA and Tukey HSD Post-Hoc Results for Inter-LLM Comparisons*

| Metric                           | Comparison        | p-value (adjusted) |
|----------------------------------|-------------------|--------------------|
| Lexical Diversity (TTR)          | GPT-4o vs. Claude | 0.045*             |
|                                  | GPT-4o vs. Gemini | 0.08               |
|                                  | Claude vs. Gemini | 0.96               |
| Semantic Fidelity (BERTScore F1) | GPT-4o vs. Claude | 0.45               |
|                                  | GPT-4o vs. Gemini | 0.022*             |
|                                  | Claude vs. Gemini | 0.16               |

\*  $p < 0.05$  BERTScore computed using human translation as reference.

On TTR, GPT-4o scores significantly higher than Claude ( $p = 0.045$ ). The difference between GPT-4o and Gemini does not reach significance ( $p = 0.08$ ). Claude and Gemini show no significant difference from each other ( $p = 0.96$ ). On BERTScore

F1, no significant difference is found between GPT-4o and Claude ( $p = 0.45$ ). Gemini scores significantly higher than GPT-4o ( $p = 0.022$ ), while the difference between Claude and Gemini is not significant ( $p = 0.16$ ). These results suggest that GPT-4o shows higher lexical diversity, whereas Gemini shows the highest semantic similarity to the human translations.

### **Word-Level Traces: Frequency Distribution**

As a supplementary analysis following the broader lexical diversity analysis above, Table 7 presents the 15 most frequent words in the human and LLM corpora. Thai items are presented in script only because the analysis at this stage concerns frequency and grammatical function rather than pronunciation. While this analysis has been used in stylometric research (e.g., Rybicki & Eder, 2010), it is used here only to provide a closer view of high-frequency lexical patterning in the translated texts.

**Table 7**

*Top 15 Most Frequent Words in Human and LLM Translations*

| Rank | Human Translation |             |             | LLM Translation |             |             |
|------|-------------------|-------------|-------------|-----------------|-------------|-------------|
|      | Word              | Word Class  | % of tokens | Word            | Word Class  | % of tokens |
| 1    | ที่               | Conjunction | 6.2%        | ที่             | Conjunction | 6.3%        |
| 2    | ของ               | Adposition  | 4.2%        | ของ             | Adposition  | 4.8%        |
| 3    | เป็น              | Verb        | 3.7%        | และ             | Conjunction | 4.7%        |
| 4    | ไม่               | Particle    | 3.6%        | ฉัน             | Pronoun     | 4.3%        |
| 5    | ใน                | Adposition  | 3.4%        | ไม่             | Particle    | 4.2%        |
| 6    | และ               | Conjunction | 3.3%        | ใน              | Adposition  | 4.1%        |
| 7    | ว่า               | Conjunction | 2.8%        | เขา             | Pronoun     | 3.2%        |
| 8    | ไป                | Auxiliary   | 2.6%        | ว่า             | Conjunction | 2.8%        |
| 9    | มี                | Verb        | 2.6%        | เธอ             | Pronoun     | 2.8%        |
| 10   | ผม                | Pronoun     | 2.6%        | เป็น            | Verb        | 2.7%        |
| 11   | ก็                | Particle    | 2.5%        | มี              | Verb        | 2.5%        |
| 12   | จะ                | Auxiliary   | 2.3%        | ผม              | Pronoun     | 2.4%        |
| 13   | เขา               | Pronoun     | 2.1%        | จะ              | Auxiliary   | 2.3%        |
| 14   | ฉัน               | Pronoun     | 2.0%        | มัน             | Pronoun     | 1.9%        |
| 15   | เธอ               | Pronoun     | 1.9%        | ไป              | Auxiliary   | 1.8%        |

Note: Percentage calculated as proportion of total tokens within each corpus (Human: 4,160 tokens; LLM: 10,373 tokens). Word class labels follow Universal Dependencies taxonomy (Nivre et al., 2020).

The top three items appear in both groups at comparable percentages, which likely reflects the grammatical structure of Thai rather than differences in translation approaches. This pattern diverges from rank 4 onwards. In the LLM corpus, ฉัน (*I/me - neutral register*) appears at rank 4 with 4.3% of all tokens. In the human corpus, the

first pronoun does not appear until rank 10, where *ผม* (*I/me - male register*) accounts for 2.6%. Four pronouns appear in the human top 15 (ranks 10-15). Five pronouns appear in the LLM top 15. *มัน* (*It - non-human/inanimate register*) appears at rank 14 in LLM translations but does not enter the human top 15 at all.

Two further observations are worth noting. The discourse particle *ก็* appears at rank 11 in human translations but is absent from the LLM top 15. This suggests somewhat broader grammatical variation in the human translations. *เป็น* (verb: *to be/become*) ranks third in human translations but falls to tenth in LLM translations. This suggests that the human translations rely more heavily on this high-frequency verb in this corpus, while the LLMs rely less on this verb and more on other high-frequency items.

Across the 15 ranks, LLM translations show a higher concentration of pronouns at high- frequency positions. In human translations, pronouns appear lower in the frequency ranks, while verbs and particles occupy the higher positions.

### **Lexical Traces**

The frequency pattern in Table 7 shows that in the human corpus, *ผม* and *ฉัน* mark different speakers (male vs. gender-neutral first person). *เธอ* and *เขา* distinguish female and male third-person referents, respectively. In the LLM corpus, *ฉัน* is used more frequently in first-person reference regardless of speaker identity, and *มัน* appears as a more general third-person pronoun for both human and non-human referents. *คุณ*, a generic second-person pronoun, is absent from human translations, but appears across all three LLM systems.

These pronoun choices are visible in the source text examples. In the rendering of child speech (“We ready now,” the younger one had said, smiling), the human translation uses orthographic variants that cue non-standard pronunciation: *เลา* /lāw/ for a standard *เรา* /rāw/ (“we”) and *พร้อม* /phō:m/ for a standard *พร้อม* /phrō:m/ (“ready”). These choices mark age, informality, and orality simultaneously. The LLMs render the line as standard Thai (e.g. *เราพร้อมแล้ว* “we are ready”), remaining within conventional spelling and reducing the phonological play found in the human translation.

A second contrast involves the handling of emphatic repetition in “Justin listens to loud and obnoxious music on a loud and obnoxious station where loud and obnoxious DJs make loud and obnoxious jokes...”. The human translator used four different Thai expressions for “*loud and obnoxious*”: *เสียงดังน่ารำคาญ* (“annoyingly loud”), *หนักหู* (“ear-splitting”), *พูดเป็นต๋อยหอย* (an idiom comparing noisy talk to hammering oysters), and *ตลกโปกฮาไม่น่าฟัง* (“slapstick jokes that are unpleasant to hear”). The LLMs repeated *ดัง* (“loud”) and *น่ารำคาญ* (“annoying”) across all four instances. The expression *ต๋อยหอย* alone contains more lexical variation than the repeated sequence used in some LLM versions.

A third example includes verb choices describing movement through a dark field (“We had walked away, hurriedly, feet stumbling...”). The human translator used compound verbs that connote physical and emotional state: อารามรีบ (“a hectic rush”), and หอบสังขาร (“dragging one’s body along with exhaustion”). The LLMs render it with more generic phrasing as รีบถอยหนี (“hurriedly retreated”) or เดินหนีไปอย่างรวดเร็ว (“walked away quickly”), which, while accurate, are less specific about the physical and emotional details.

The frequency pattern in Table 7 shows that in the human corpus, ผม and ฉัน mark different speakers (male vs. gender-neutral first person). เธอ and เขา distinguish female and male third-person referents, respectively. In the LLM corpus, ฉัน is used more frequently in first-person reference regardless of speaker identity, and มัน appears as a more general third-person pronoun for both human and non-human referents. คุณ, a generic second-person pronoun, is absent from human translations, but appears across all three LLM systems.

These pronoun choices are visible in the source text examples. In the rendering of child speech (“We ready now,” the younger one had said, smiling), the human translation uses orthographic variants that cue non-standard pronunciation: เลา /lāw/ for a standard เรา /rāw/ (“we”) and พร้อม /phō:m/ for a standard พร้อม /phrō:m/ (“ready”). These choices mark age, informality, and orality simultaneously. The LLMs render the line as standard Thai (e.g. เราพร้อมแล้ว “we are ready”), remaining within conventional spelling and reducing phonological play found in the human translation.

A second contrast involves the handling of emphatic repetition in “Justin listens to loud and obnoxious music on a loud and obnoxious station where loud and obnoxious DJs make loud and obnoxious jokes...”. The human translator used four different Thai expressions for “loud and obnoxious”: เสียงดังน่ารำคาญ (“annoyingly loud”), หนักหู (“ear-splitting”), พูดเป็นต๋อยหอย (an idiom comparing noisy talk to hammering oysters), and ตลกโปกฮาไม่น่าฟัง (“slapstick jokes that are unpleasant to hear”). The LLMs repeated ดัง (“loud”) and น่ารำคาญ (“annoying”) across all four instances. The expression ต๋อยหอย alone contains more lexical variation than the repeated sequence used in some LLM versions.

A third example includes verb choices describing movement through a dark field (“We had walked away, hurriedly, feet stumbling...”). The human translator used compound verbs that connote physical and emotional state: อารามรีบ (“a hectic rush”), and หอบสังขาร (“dragging one’s body along with exhaustion”). The LLMs render it with more generic phrasing as รีบถอยหนี (“hurriedly retreated”) or เดินหนีไปอย่างรวดเร็ว (“walked away quickly”), which, while accurate, are less specific about the physical and emotional details.

## Discussion

### *Lexical and Syntactic Traces of Human and LLM Translators*

At first glance, the human translations do not appear more lexically varied than Claude and Gemini (TTR scores of 0.405 and 0.406, respectively), while GPT-4o shows a more varied vocabulary (TTR score of 0.486). The Maas Index, however, points to a different pattern (human: 0.016 vs LLMs: 0.183-0.198). In LLM translations, vocabulary tends to narrow as the passage progresses (See Table 4). In human translations, however, new and varied lexical items continue to appear throughout. What distinguishes the human translations in this corpus is not how many unique words (TTR) appear in total, but that the vocabulary does not narrow as the passage progresses, at least within this corpus. This pattern is also reflected in the four unique translations of “loud and obnoxious,” the child speech, and verb choice examples discussed in the previous section.

A similar contrast appears in sentence-level patterning. The three LLM systems do not differ significantly from one another on average sentence length or sentence-length variance (ANOVA,  $p = 0.339$  and  $p = 0.486$ , respectively), compared with human translations, which show substantially higher sentence-length variance (See Table 4). As elaborated previously using Blanco’s *Please Stop Laughing at Me*, the human translator included details, content, and emotional states within one extended sentence. GPT-4o, Claude, and Gemini each break the same content into shorter, separate sentences, using a more generic verb and a two-word noun phrase. In this corpus, the human translations are, therefore, associated with fluctuation in sentence length, while the LLM outputs show more uniform sentence patterning.

These findings address the first research question by showing that both human and LLM translations leave observable lexical and syntactic traces. They also suggest that those traces are distributed differently across the corpus. The human translations display sustained lexical variation and fluctuate more in sentence length, whereas the LLM outputs tend towards a narrower lexical range and more uniform sentence patterning.

### *Stylistic Distance between Human and LLM Translations*

The second research question asks how the stylometric traces of the human and LLM translations compare across the corpus. Human translations are more different from each LLM system (average Delta = 0.82) than the LLM systems are from each other (average Delta = 0.70). This suggests that the human translations are more clearly separated from the LLM translations than the LLM translations are from one another in this dataset. The pronoun patterns (See Table 7) are a case in point. Across the corpus, the human translations use *ကျွန်* (male first person), *ကျွန်* (gender-neutral first person), *သူ* (male third person), and *မိမိ* (female third person) to mark different speakers and registers, while the three LLMs rely more heavily on *ကျွန်* and *မိမိ* (“it” or a more generic third-person form) as defaults for the first- and the third-

person pronouns. One likely reason มั่น does not enter the human top 15 is that the human translations differentiate referents more by person, gender, and register, which reduces reliance on a single default third-person form. At the other end of the frequency continuum, orthographic variants like เลา (“we,” non-standard spelling) and พร้อม (“ready,” non-standard spelling) are rare and would not be detected by Delta at all. Yet, they still contribute to the distinctiveness of the human translations in the qualitative examples. The Delta gap (0.12), therefore, captures only part of what separates human from LLM translations. Measuring this distance through a broader feature set to yield a more complete picture is an area for further study. BERTScore F1 results show that the LLM outputs remain semantically close to the human translations, even where their lexical and syntactic patterning differs.

### ***Translator Invisibility Revisited***

The third research question asks what these observed traces and patterns may suggest about translator invisibility in literary translation from English to Thai. Venuti (1995/2008) argued that translator invisibility is built through norms and institutions that reward fluency and discourage foreignisation. The present findings do not measure that broader socio-cultural process directly. They do show, however, that both human and LLM translations leave detectable stylometric traces in the translated texts. Across the corpus, these traces are observable in lexical richness, sentence-length patterning, pronominal reference, and overall stylistic distance. In this sense, the study approaches translator visibility at the textual level by examining how translation leaves measurable traces in the translated text.

Baker (2000) described translator style as a recurring pattern of linguistic behaviour, while Saldanha (2011) sets out stricter methodological conditions for identifying translator-specific style across a translator’s body of work. The present findings do not meet that stronger standard, since each human participant contributed only one passage. Even so, the human translations in this corpus display stylistic traces at the level of lexical choice, sentence patterning, and local textual effect. The examples of เลา (a non-standard spelling of “we”) and พร้อม (a non-standard spelling of “ready”) in rendering a child’s voice, as well as the four different expressions used for “loud and obnoxious”, suggest that these traces are not random variation alone, but form part of how the human translations respond to tone, register, and local textual effect. The contrast is especially clear in how repeated wording is handled. As Zhang et al. (2020) note, evaluation should attend to semantic equivalence rather than surface-form similarity only. In this example of ‘loud and obnoxious’, the LLM preserves the repeated surface wording more literally, whereas the human translation responds with a wider range of Thai lexical choices. This suggests that surface repetition and lexical diversity capture different aspects of translation. Repeating the same wording may preserve the source more literally at the surface level, while varied lexical choices may preserve tone, emphasis, or local textual effect in a less literal way.

The three LLM systems also show recurring, albeit different, stylometric patterns. Across the corpus, they tend towards a narrower lexical range, more regular pronoun use, and shorter or more uniform sentence structures. These tendencies are in line with Jiao et al. (2023), who observed that translation quality drops for language pairs with less training data or greater typological distance, such as the pair of English and Thai. Related limitations were also noticed in Karpinska and Iyyer's (2023) findings, leading them to conclude that even at the paragraph level, LLMs still show limitations in preserving the original author's tone and register. Their study focused on human evaluation rather than stylometric analysis, but the results of the present study are compatible with their broader observation.

These results should not be interpreted as a full empirical test of Venuti's argument. Both the human translations and the LLM outputs leave detectable traces in our corpus. The difference lies in the kinds of traces they leave. Within this corpus, the human translations show comparatively broader lexical variation, greater fluctuation in sentence length, and more differentiated pronoun use. The LLM outputs, on the other hand, cluster more closely around a narrower vocabulary range and more uniform sentence patterns. The stylometric differences found describe the translations analysed here and should not be regarded as patterns that every human translator or every LLM would follow universally.

From the perspective of English language studies, the findings also show how stylistic features of the English source texts are mediated differently in Thai translation. Across the corpus, features such as emphatic repetition, non-standard speech, pronominal contrast, and long embedded sentence structures are more often preserved or elaborated in the human translations. By contrast, the LLM outputs tend to standardise, simplify, or restructure them into shorter, more uniform structures. The contribution of this study, therefore, is not only about distinguishing human and LLM translation patterns but also about showing which aspects of English textual patterning remain visible, become flattened, or are restructured in Thai literary translation.

### ***Implications for Human-in-the-loop Translation in the AI Age***

The stylometric differences between the human and LLM corpora suggest that humans and LLMs may be better suited to different roles in the translation process. LLM translations in this study are fluent and close to the source meaning, but they draw on a narrower lexical range, produce less varied sentence lengths, and adopt a more uniform style. Human translations, however, maintain a wider vocabulary and more flexible sentence patterning throughout. This suggests that LLMs may be well-suited for producing first drafts and handling routine segments. Human translators, on the other hand, are especially important when the focus is on specific lexical meanings, tone, register, cultural reference, and narrative effects. In post-editing workflows, this means prioritising human attention on points where LLMs are less effective, including preserving lexical contrasts, choosing more precise alternatives, and modifying sentence patterns to reflect shifts in mood and emphasis.

This human-in-the-loop workflow also raises copyright questions. In many legal contexts, copyright protection is tied to human authorship, resulting in the status of purely AI-generated outputs being uncertain or limited. In the changing landscape of translation practice, hybrid workflows are increasingly common, and they complicate the matter even further because the contributions of humans and algorithms are tightly interwoven. As Nam (2023) argues, such a nuanced workflow invites a re-evaluation of how to distinguish between a reproduction of the original work and a translation in its own right.

### ***Implications for English/ Translation Classrooms***

The first implication from the human and LLM corpora concerns how English could be taught to students in a translation study program. The human translations in this study show a broader lexical variation and a greater sentence-length fluctuation than the LLM outputs. This suggests that literary translation classes should focus on more than core vocabulary and grammar so that students develop the ability to reach beyond familiar, high-frequency items. As for sentence patterning, classroom activities could focus on syntactic variety and rhythm. For example, students can be asked to rewrite short English passages into Thai with deliberate changes in sentence lengths, alternating between short and long sentences, and analyse how their choices affect information flow and tone.

Another practical implication is glossary-building as an active process rather than a record-keeping task. When students work with a text containing clusters of near-synonymous words, technical terms, or recurring motifs, they can be asked to map each English item to a range of Thai options and discuss which choices best preserve nuance, register, and voice in different contexts. Over time, this kind of work can make the range of available choices visible and reduce students' reliance on a small set of safe, generic equivalents.

### **Conclusion**

The study set out to examine whether stylometric analysis could detect differences between human and LLM translations in a small corpus of English-Thai literary translations, focusing on three research questions, namely, what traces are observable, how they compare, and what they might suggest about translator invisibility. The findings, though mixed, suggest a general pattern in that both human and LLM translations leave detectable traces, but that these are distributed differently. The human translations maintained a wider and more varied vocabulary throughout each passage and fluctuated more noticeably between longer and shorter sentences. The LLM outputs, by contrast, drew on a narrower lexical range, favoured more generic phrasing, and produced more uniform sentence structures across the passage. Despite these stylometric differences, the LLM outputs remained semantically close to the human translations even when their stylistic patterning differed.

These results do not constitute a test of Venuti's invisibility argument in its socio-cultural sense. They do, however, suggest that traces of translation are detectable at the textual level, and that the two groups of translations examined here left different kinds of traces. For English language studies, the corpus offers an illustration of these detectable traces through examples of emphatic repetition, child speech orthography, pronominal differentiation, and embedded clause structure, which are features that were more often carried across in the human translations but flattened or broken into shorter units in the LLM outputs. This study should, therefore, be read as a focused case study with a specific corpus, a specific language pair, and a limited feature set, and is therefore not generalisable to human or LLM translation in a broader sense.

Future research would benefit from a more controlled design. This may involve asking the same translators to translate multiple source texts or collecting several human translations of the same passage, so that translator effects can be disentangled from source-text effects more cleanly. Expanding the analytical scope to include other linguistic features, such as discourse markers or figurative language, and testing additional language pairs and different genres would help establish whether the tendencies observed here are specific to English-Thai literary translation or reflect broader patterns. Combining stylometric analysis with reader-response data would help clarify whether the textual traces detected here are ones that readers actually notice or find meaningful.

### ***Limitations***

As a preliminary investigation, this study is intentionally small in scope. The corpus is restricted to six passages from English literary texts, each translated into Thai by six human translators and three LLM systems. The most important limitation lies in the structure of the human corpus. Each human participant contributed one passage only, so the study compares a benchmark set of human translations with three LLM sets rather than identifying recurring styles across multiple texts by the same translator. The patterns reported here, therefore, characterise this particular set of texts, translators, and models and should not be generalised beyond it.

The study also focused on a small set of stylometric features, including lexical diversity metrics, sentence-length patterns, and a supplementary most-frequent-word analysis. As a result, while the study identifies textual differences between human and LLM translations, it does not show how those differences are perceived or valued by readers. Within these constraints, the main contribution is methodological. The analysis demonstrates one way of examining translator visibility at the textual level through stylometric traces and of comparing human and LLM translations in a small English-Thai corpus.

## Author Biographies

Sanooch S. Nathalang, PhD. is a lecturer at Thammasat University. As a former researcher at NECTEC, she merges technology and education across her research interests in technology-enhanced learning, AI, translation and interpreting applications, and their pedagogical implications. She is a key contributor to a national policy-level AI in education initiative. <https://orcid.org/0009-0008-7392-8404>

Natthaphon Tripornchaisak, PhD is a lecturer at Thammasat University, specialising in interpreting pedagogy and intercultural communication. An evaluator for the Thai Parliament and interpreter for the IMF and UN, he has completed AIIC Training of Trainers seminars. A former visiting scholar, he bridges academic theory with high-level conference interpreting standards. <https://orcid.org/0000-0001-9540-462X>

## References

- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–250). John Benjamins
- Baker, M. (2000). Towards a methodology for investigating the style of a literary translator. *Target*, 12(2), 241–266. <https://doi.org/10.1075/target.12.2.04bak>
- Baroni, M., & Bernardini, S. (2006). A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, 21(3), 259–274. <https://doi.org/10.1093/lc/fqi039>
- Burrows, J. (2002). ‘Delta’: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3), 267–287. <https://doi.org/10.1093/lc/17.3.267>
- Eder, M. (2015). Does size matter? Authorship attribution, small samples, big problem. *Digital Scholarship in the Humanities*, 30(2), 167–182. <https://doi.org/10.1093/lc/fqt066>
- Eder, M., Rybicki, J., & Kestemont, M. (2016). Stylometry with R: A package for computational text analysis. *The R Journal*, 8(1), 107–121. <https://doi.org/10.32614/RJ-2016-007>
- Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C., & Vitt, T. (2017). *Understanding and explaining Delta measures for authorship attribution*. *Digital Scholarship in the Humanities*, 32(suppl\_2), ii4–ii16. <https://doi.org/10.1093/lc/fqx023>
- Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). *Is ChatGPT a good translator? A preliminary study*. arXiv preprint arXiv:2301.08745. <https://arxiv.org/abs/2301.08745>

- Karpinska, M., & Iyyer, M. (2023). *Large language models effectively leverage document-level context for literary translation*. arXiv preprint arXiv:2304.02210. <https://doi.org/10.48550/arXiv.2304.03245>
- Kolirin, L. (2026, January 23). Like digging ‘your own professional grave’: The translators grappling with losing work to AI. CNN Business. <https://edition.cnn.com/2026/01/23/tech/translation-language-jobs-ai-automation-intl>
- Koizumi, R., & In’nami, Y. (2012). *Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens*. *System*, 40(4), 554–564. <https://doi.org/10.1016/j.system.2012.10.012>
- Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, 43(4), 557–570. <https://doi.org/10.7202/003425>
- Lembersky, G., Ordan, N., & Wintner, S. (2012). Language models for machine translation: Original vs. translated texts. *Computational Linguistics*, 38(4), 799–825. [https://doi.org/10.1162/COLI\\_a\\_00111](https://doi.org/10.1162/COLI_a_00111)
- Nam, H. D. (2023). Copyright Law Issues Regarding AI-based Literature Translation: Déjà-vu of the Tower of Babel? *Jeo’jag’gweon*, 36(4), 33–88. <https://doi.org/10.30582/kdps.2023.36.4.33>
- Nivre, J., de Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C., Pyysalo, S., Schuster, S., Tyers, F., & Zeman, D. (2020). Universal dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)* (pp. 4034–4043). European Language Resources Association. <https://aclanthology.org/2020.lrec-1.497/>
- Phatthiyaphaibun, W., Chaovavanich, K., Polpanumas, C., Suriyawongkul, A., Lowphansirikul, L., Chormai, P., Limkonchotiwat, P., Suntornitip, T., & Udomcharoenchaikit, C. (2023). PyThaiNLP: Thai natural language processing in Python. In L. Tan, D. Milajevs, G. Chauhan, J. Gwinnup, & E. Rippeth (Eds.), *Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023)* (pp. 25–36). Association for Computational Linguistics. <https://aclanthology.org/2023.nlposs-1.4/>
- Rybicki, J. (2012). The great mystery of the (almost) invisible translator: Stylometry in translation. In M. Oakes & M. Ji (Eds.), *Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research* (pp. 231–248). John Benjamins.
- Rybicki, J., & Eder, M. (2010). Deeper Delta across genres and languages: Do we really need the most frequent words? *Literary and Linguistic Computing*, 26(3), 315–321. <https://doi.org/10.1093/lilc/fqr031>

- Rybicki, J., & Heydel, M. (2013). The stylistics and stylometry of collaborative translation: Woolf's Night and Day in Polish. *Literary and Linguistic Computing*, 28(4), 708–717. <https://doi.org/10.1093/LLC/FQT027>
- Saldanha, G. (2011). Translator style: Methodological considerations. *The Translator*, 17(1), 25–50. <https://doi.org/10.1080/13556509.2011.10799478>
- Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.
- Venuti, L. (2008). *The translator's invisibility: A history of translation* (2nd ed.). Routledge. (Original work published 1995). <https://doi.org/10.4324/9780203360064>
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT. In *Proceedings of the International Conference on Learning Representations (ICLR 2020)*. <https://openreview.net/forum?id=SkeHuCVFDr>