



INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

Volume 4; Issue 3; 2026; Page No. 93-100

Received: 01-02-2026

Accepted: 06-04-2026

Published: 07-06-2026

Translating the Madman's Voice: A Corpus-Based Stylistic Comparison of Three English Versions of Lu Xun's *Kuangren Riji*

GUO Nyuhuan

School of Foreign Languages, Zhaoqing University, Zhaoqing, Guangdong, China

DOI: <https://doi.org/10.5281/zenodo.20960438>

Corresponding Author: GUO Nyuhuan

Abstract

This paper takes Lu Xun's *A Madman's Diary* (1918), widely recognised as the first modern Chinese short story, as its research subject. Although numerous English translations of this work have been produced over the past century, few existing studies carry out systematic stylistic comparisons across different translated versions. Selecting three English translations-the 1960 version by Yang Xianyi and Gladys Yang, the 1990 version by William Lyell, and the 2009 version by Julia Lovell-this study builds a self-compiled parallel corpus consisting of 285 aligned Chinese-English sentence pairs. It conducts quantitative comparisons through multiple corpus indicators including type-token ratio (TTR), standardised type-token ratio (STTR), lexical density, mean sentence length, dialogue proportion, expansion ratio and Flesch Reading Ease from four dimensions: lexical richness, syntactic structure, dialogue representation and translational expansion, supplemented by qualitative close textual reading to quantitatively identify stylistic divergences among the translations. The results reveal markedly distinct translational styles across the three versions. Julia Lovell's translation achieves the highest TTR (0.442) and lexical density (0.489), featuring superior lexical diversity and textual information density. William Lyell's translation, with a total word count of 1,933, a dialogue proportion of 40.7% and 22 exclamation marks, is the lengthiest text and adopts a dramatic, performance-oriented translating strategy. The Yangs' translation only has a dialogue proportion of 9.4% and the highest expansion ratio of 0.708, with 16 one-to-many sentence splits; it prioritises the explicitation of narrative content while downplaying dramatic expressions. The Flesch Reading Ease scores of all three translations range from 66.9 to 67.7, indicating comparable readability. This research verifies that corpus stylistic approaches can explicitly expose translators' choices that are imperceptible through conventional impressionistic reading, and also provides an empirically grounded analytical framework for comparing multiple retranslations of canonical literary texts.

Keywords: Lu Xun, *Kuangren Riji*, corpus stylistics, literary translation, retranslation, translation style, lexical diversity

1. Introduction

Lu Xun's *Kuangren Riji* (狂人日记), published in 1918 in *Xin Qingnian* (新青年, *New Youth*), occupies a singular position in modern Chinese literary history. Widely acknowledged as the first story written in vernacular Chinese rather than classical literary Chinese, it drew formal inspiration from Nikolai Gogol's eponymous story while channelling a distinctly Chinese cultural and political critique. The story's unnamed narrator - a "madman" whose paranoid conviction that his neighbours, family, and even children are cannibals - serves as a searing allegory for the spiritual and moral decay Lu Xun perceived in post-imperial

Chinese society (Lee, 1987; Denton, 1996) ^[11, 5]. Over the century since its publication, *Kuangren Riji* has been translated into English at least five times, most notably by Yang Xianyi and Gladys Yang (1960) ^[21], William Lyell (1990) ^[15], and Julia Lovell (2009) ^[12]. Each translation represents a distinct historical moment, translational philosophy, and target readership, yet systematic comparative analysis across these versions remains underdeveloped.

The five English translations span nearly a century: George K. Leung (1926), Chi-Chen Wang (1941), Yang Xianyi and Gladys Yang (1960) ^[21], William Lyell (1990) ^[15], and Julia

Lovell (2009)^[12]. Each emerged from a distinct translational ecology. Leung's and Wang's early versions, produced when Chinese literature was virtually unknown to Anglophone readers, relied heavily on paratextual framing - prefaces, footnotes, and cultural glosses - to bridge the vast cultural distance. The Yangs' mid-century version, published by the state-run Foreign Languages Press during the height of Maoist cultural diplomacy, represented an institutional project of presenting revolutionary China to the world through its literature. Lyell's 1990^[15] translation, produced for the University of Hawai'i Press, reflected the maturation of Chinese literary studies as an Anglophone academic discipline and the post-1968 appetite for cultural authenticity. Lovell's 2009^[12] Penguin Classics translation marked the consecration of modern Chinese literature within the global literary marketplace - the same trajectory that would culminate in Mo Yan's 2012 Nobel Prize. Understanding how these historically situated translations construct stylistically different versions of the same story is thus not merely a technical exercise in comparative linguistics but an inquiry into the material conditions under which Chinese literature enters the Anglosphere.

The central problem this study addresses is the absence of empirically grounded, corpus-driven comparison across multiple English translations of *Kuangren Riji*. While individual translators' choices have been noted in passing by literary critics and translation scholars, no study has quantified the stylistic profiles of these translations using consistent, replicable corpus-linguistic measures. This gap is particularly significant given the canonical status of the source text: *Kuangren Riji* is anthologised, taught, and cited as the inaugural work of modern Chinese fiction, yet Anglophone readers encounter it exclusively through translation.

Existing studies have tended towards qualitative close reading of select passages or have focused on a single translator's oeuvre rather than cross-translational comparison. Kowallis (1994)^[10], a Lu Xun scholar and translator himself, provided one of the earliest systematic assessments of Lyell's translation approach, noting Lyell's tendency to preserve the stylistic idiosyncrasies of Lu Xun's prose - its abrupt register shifts, its vernacular energy, and its deliberate syntactic roughness. Henderson (2007)^[22] examined the Yangs' translations within the institutional framework of the state-sponsored Foreign Languages Press, arguing that their characteristic grammatical explicitness and formal register reflected the political and editorial constraints of Maoist cultural diplomacy. These qualitative studies established the foundational understanding of individual translators' styles but did not attempt systematic cross-translational comparison using standardised measures. The present study fills this methodological gap by constructing a parallel corpus and applying a standardised set of corpus-stylistic measures. The choice of *Kuangren Riji* as the source text is motivated by three considerations: its canonical status ensures sufficient scholarly interest; its short length (~5,000 Chinese characters) makes full-text corpus analysis feasible without sampling; and its first-person diary form, with its distinctive mixture of internal monologue and remembered dialogue, provides rich material for the dialogue-proportion and narrative-voice measures that constitute central dimensions of the analysis.

This study is guided by three research questions

- RQ1:** What are the quantifiable stylistic differences - in terms of lexical richness, syntactic complexity, dialogue representation, and translation expansion - amongst the three English translations?
- RQ2:** How do these quantitative stylistic profiles correspond to the known translational philosophies and historical contexts of each translator?
- RQ3:** What do the patterns of difference and similarity across the three versions reveal about the nature of retranslation as a stylistic phenomenon?

The study makes three contributions. First, it provides the first systematic, corpus-based stylistic profile of three major English translations of *Kuangren Riji*, filling a gap in the existing Lu Xun translation scholarship. Second, it demonstrates the applicability of corpus stylistic methods to the comparison of multiple retranslations of a single canonical source text. Third, by triangulating quantitative corpus data with qualitative close reading and contextual-historical analysis, it models a mixed-methods approach that respects both the empirical rigour of corpus linguistics and the interpretive nuance of literary translation studies.

2. Literature Review

2.1 Theoretical Framework: Stylistics and Retranslation

Corpus stylistics - the application of corpus-linguistic methods to the analysis of literary style - emerged in the early 2000s as a bridge between quantitative linguistics and traditional literary criticism (Semino & Short, 2004; Fischer-Starcke, 2010)^[17, 6]. Key measures include type-token ratio (TTR) for lexical diversity, standardised type-token ratio (STTR) to control for text length, lexical density, mean sentence length, and readability indices such as Flesch Reading Ease (Flesch, 1948)^[7]. When applied to translated texts, these measures can reveal patterns of stylistic choice that distinguish one translator's work from another - what Baker (2000)^[1] termed "translator's style" and Saldanha (2011)^[16] reconceptualised as "translator style". The present study aligns with this translator-centred approach, using corpus methods not to identify universals but to profile the distinctive stylistic signatures of three translators working on the same source text.

The phenomenon of retranslation has also attracted significant theoretical attention. Berman (1990)^[2] proposed the "retranslation hypothesis", arguing that first translations tend towards domestication while subsequent retranslations can afford to be more source-text-oriented. The English retranslations of *Kuangren Riji* offer an instructive test case, spanning the Cold War-era Yangs (1960)^[21], the multicultural academic Lyell (1990)^[15], and the globalised commercial Lovell (2009)^[12]. Koskinen and Paloposki (2010)^[2] and Deane-Cox (2014) have complicated Berman's teleological model, arguing that retranslation responds to contextual factors - publishing market, target readership, translator identity - that may pull in different stylistic directions rather than converging towards a "perfect" translation.

Further work by Laviosa (2002)^[24] on translation universals - specifically simplification, explicitation, and normalisation tendencies - provides an empirically grounded backdrop for the present analysis. Laviosa's findings that translated

English tends towards lower lexical density and a higher proportion of function words find an intriguing counterpoint in the present data: the Yangs' translation, with its lowest lexical density (0.453), conforms most closely to the universal pattern, while Lovell's (0.489) deviates most sharply. This suggests that translation universals may be moderated by individual translator style and publishing context - a possibility that corpus-based analysis of multiple retranslations is uniquely positioned to investigate.

Munday (2008) ^[26] advanced the methodology of translator style research by combining systemic functional linguistic analysis with sociological investigation of the translator's working conditions and institutional context. His study of Latin American literary translators demonstrated that stylistic choices are shaped not merely by linguistic competence or aesthetic preference but by broader systems of patronage, editorial intervention, and marketplace expectations. Applying Munday's framework to the present case would position the Yangs' translation within the state-sponsored cultural diplomacy apparatus of Maoist China, Lyell's within the peer-review norms of Anglophone university-press publishing, and Lovell's within the commercial imperatives of the Penguin Classics series. Each institutional context imposes a distinct set of stylistic pressures: fidelity to political messaging (the Yangs), fidelity to scholarly apparatus (Lyell), and fidelity to readability and market appeal (Lovell).

The conceptual apparatus of Conceptual Metaphor Theory (Lakoff & Johnson, 1980) ^[23] also illuminates the specific challenges of translating Lu Xun's prose. Lu Xun's cannibalism allegory - the organising metaphor of *Kuangren Riji* - draws on deeply embedded Chinese cultural metaphors linking consumption, social relations, and moral decay. The three translators' divergent stylistic choices at the lexical and syntactic levels may thus reflect different strategic responses to the same underlying problem: how to render a culturally specific conceptual metaphor system in a target language lacking the relevant metaphorical mappings. Lovell's higher lexical diversity and Lyell's dramatic intensity may represent alternative solutions - Lovell expanding the lexical field around the metaphor, Lyell amplifying its emotional charge - to what is ultimately a problem of cross-cultural conceptual transfer.

2.2 Research Gap: Translating Lu Xun

The three translators studied here embody distinct philosophies. Yang Xianyi and Gladys Yang, the most prolific translators of Chinese literature into English, characterised their approach as one of strict fidelity: "We tried to be as faithful as we could... We didn't make many changes" (Yang, 1995) ^[20]. Their *Kuangren Riji* appeared in *Selected Stories of Lu Hsun* (1960) ^[21], published by the state-run Foreign Languages Press during the Maoist period. William Lyell was both a translator and a Lu Xun scholar, authoring the influential monograph *Lu Hsun's Vision of Reality* (1976) ^[14]. His translation, published in *Diary of a Madman and Other Stories* (1990) ^[15] by a university press, explicitly aimed at preserving the stylistic texture of Lu Xun's prose - its "sharpness, its brittleness, its occasionally rough-hewn quality" (Lyell, 1990, p. xli) ^[15]. Julia Lovell's translation, commissioned by Penguin Classics and published as *The Real Story of Ah-Q and Other Tales of*

China: The Complete Fiction of Lu Xun (2009) ^[12], articulates a philosophy of accessibility, aiming to "let Lu Xun's stories speak as directly as possible to the contemporary reader" (Lovell, 2009) ^[12].

Despite this rich translational history, existing scholarship on Lu Xun translation has produced valuable qualitative insights - particularly Kowallis (1994) ^[10] on Lyell's scholarly translation approach and Henderson (2007) ^[22] on the Yangs' institutional translation context - but has not generated systematic, cross-translational, quantitatively grounded comparison. The qualitative tradition has meticulously documented individual translators' characteristic choices: Lyell's archaizing diction, the Yangs' grammatical completeness, Lovell's contemporary accessibility. Yet without a shared quantitative framework, these observations remain incommensurable. Is Lyell's translation *more* lexically diverse than Lovell's, or merely *differently* diverse? Does the Yangs' explicitation produce a *measurably* longer text, or does it merely feel longer to the critical reader? These are empirical questions that qualitative methods alone cannot resolve. The present study addresses this gap by constructing a parallel corpus and applying a standardised set of corpus-stylistic measures to produce directly comparable stylistic profiles.

3. Methodology

3.1 Corpus Design and Processing

The study corpus is a self-constructed parallel corpus comprising the Chinese source text of Lu Xun's *Kuangren Riji* and its three English translations. The source text was obtained from a public-domain digital edition and verified against the standard printed text in *Nahan* (呐喊, 1923). The corpus was formatted as a CSV file with three columns: Chinese (the source sentence), Translator, and English (the corresponding translation). Each Chinese sentence appears in three rows, one per translator. Where a single Chinese sentence was split into multiple English sentences - employed by all three translators to varying degrees - each resulting English sentence occupies a separate row. The total corpus comprises 285 aligned rows, representing 75 unique Chinese source segments.

Text processing was conducted using Python with standard libraries - csv for data ingestion, re for regular-expression-based sentence segmentation, and hashlib for file integrity verification. Chinese text was not tokenised beyond the pre-existing sentence boundaries; character counts were used as the Chinese-unit metric throughout. English text was tokenised by whitespace splitting after punctuation normalisation. Function words were identified using a closed-class list of approximately 175 English function words, based on standard corpus-linguistic practice. The complete processing pipeline, including the function-word list and all computational procedures, is documented in the reproducibility script (`reproduce_all.py`) accompanying this paper, enabling full independent replication of all reported statistics. English text was tokenised by whitespace splitting. Pre-processing steps included removal of metadata, standardisation of encoding to UTF-8, stripping of extraneous whitespace, and manual verification of sentence boundaries against the printed editions for a random sample of 20% of the corpus. Table 1 presents the descriptive statistics.

Table 1: Corpus Descriptive Statistics by Translator

Metric	Lovell (2009) [12]	Lyell (1990) [15]	Yang & Yang (1960) [21]
Aligned rows	86	115	84
English word tokens	1,461	1,933	1,603
Chinese source characters	2,501	3,414	2,265
Unique Chinese segments	75	75	75
English sentences	90	118	96

Note: Row counts differ because translators split some Chinese sentences into multiple English sentences.

3.2 Evaluation Metrics and Procedure

These measures were selected to capture distinct dimensions of translational style: lexical richness (TTR, STTR, lexical density), syntactic profile (mean sentence length, expansion ratio), narrative presentation (dialogue proportion), and accessibility (Flesch Reading Ease, word length). The selection was guided by two principles: first, each measure should have a well-established methodological pedigree in corpus stylistics or translation studies; second, the measures collectively should profile complementary rather than redundant dimensions of style. For instance, TTR and STTR both index lexical diversity but control for different confounds (text length versus segment-internal variation), while lexical density indexes a different lexical property entirely (information load). The supplementary measures - exclamation marks and -ly adverb frequency - were included to capture emotionally expressive and descriptive dimensions of style that the core measures miss.

Sentence alignment was performed using Python's difflib.SequenceMatcher, which computes a similarity ratio based on character sequence overlap between Chinese source sentences and English target sentences. This length- and pattern-based approach was selected for computational efficiency and suitability for short, structurally parallel texts. Alignment pairs with similarity scores below a threshold were manually reviewed. A semantic alignment pass using additional lexical and positional features refined the pairing. The final aligned dataset comprised 285 sentence pairs across the three translations.

for each translation: (1) Type-Token Ratio (TTR) - the ratio of unique word types to total word tokens; (2) Standardised Type-Token Ratio (STTR) - mean TTR computed over consecutive 100-word text segments (with a minimum segment size of 50 words), which controls for the well-known confound that TTR decreases as text length increases; (3) Lexical Density - the proportion of content words (non-function words) to total words, using a standard list of approximately 175 English closed-class words (Ure, 1971) [18]; (4) Mean Sentence Length - average words per sentence; (5) Dialogue Proportion - percentage of sentences containing direct speech markers (quotation marks); (6) Translation Expansion Ratio - English word tokens divided by Chinese source characters; (7) Flesch Reading Ease - a composite readability score based on average sentence length and syllables per word (Flesch, 1948) [7]; and (8) Supplementary Measures - average word length, proportion of long words (>6 characters), exclamation mark counts, and -ly adverb frequency per 1,000 words.

These measures were selected to capture distinct dimensions of translational style: lexical richness (TTR, STTR, lexical density), syntactic profile (mean sentence length, expansion

ratio), narrative presentation (dialogue proportion), and accessibility (Flesch Reading Ease, word length). Quantitative analysis was conducted using Python; descriptive statistics were computed per translator, and supplementary qualitative close reading of passages exhibiting notable cross-translational divergence contextualised the quantitative patterns.

Several limitations should be noted. The corpus size is modest (285 aligned rows; 1,461–1,933 words per translator), and findings should not be overgeneralised to each translator's broader oeuvre. The sentence-level alignment strategy, while standard, introduces noise when translators substantially reorganise sentence boundaries. This study focuses on surface-level lexical and syntactic features and does not address deeper dimensions such as register shifts, cultural reference handling, or ideological filtering, which would require qualitative, discourse-analytic methods.

4. Findings

4.1 Lexical and Syntactic Profiles

4.1.1 Lexical richness: Table 2 presents the lexical measures. Lovell's translation displays the highest TTR (0.442) and lexical density (0.489), despite being the shortest text (1,461 tokens). This is notable: TTR typically decreases with text length, so Lovell's higher TTR in a shorter text indicates a genuinely more varied vocabulary. Lyell's translation displays the lowest TTR (0.394) alongside the largest word count (1,933 tokens), partially consistent with the expected length effect. When length is normalised through STTR, the difference narrows substantially: Lovell (0.755), Lyell (0.759), and the Yangs (0.746) are within 0.013 of one another, suggesting raw TTR differences are driven more by text length than by genuine differences in lexical diversity per unit of text. The Yangs display the lowest lexical density (0.453), consistent with a style that prioritises grammatical explicitness, adding function words to clarify logical relationships implicit in Lu Xun's paratactic Chinese prose.

Table 2: Lexical Richness Measures

Measure	Lovell (2009) [12]	Lyell (1990) [15]	Yang & Yang (1960) [21]
TTR	0.4422	0.3942	0.4067
STTR	0.7547	0.7589	0.7462
Lexical density	0.4887	0.4754	0.4529

The raw TTR differences are striking: Lovell's value of 0.442 exceeds Lyell's 0.394 by a margin of 0.048 - approximately 12% higher. In corpus linguistics, a TTR difference of this magnitude between texts of comparable length is typically considered meaningful rather than artefactual. However, the convergence of STTR values (all within 0.013 of one another) complicates this interpretation. STTR controls for text length by computing TTR within fixed-size segments and averaging the results. The STTR convergence suggests that, *per 100-word segment*, all three translations exhibit comparable lexical diversity. The raw TTR differences are therefore driven primarily by text length - Lyell's version, at 1,933 words, simply affords more opportunity for word repetition than Lovell's 1,461-word text. This finding illustrates a methodological principle that

corpus stylisticians have long emphasised: raw TTR should not be compared across texts of substantially different lengths without normalisation (STTR or an equivalent measure). The present data provide an unusually clean demonstration of this principle, because the three translations are directly comparable (same source text, same genre, same register domain) yet differ sufficiently in length to expose the TTR-length confound.

4.1.2 Syntactic structure: Table 3 presents the syntactic measures. Mean sentence length is remarkably consistent across the three translations (16.3–16.9 words), suggesting a shared norm for sentence-level pacing in English literary prose. Expansion ratio, however, varies substantially: the Yangs' translation is the most expansive (0.708 English words per Chinese character), Lyell's the most concise (0.566), and Lovell's intermediate (0.584). Sentence segmentation reinforces this pattern: the Yangs split Chinese sentences into multiple English sentences in 16 cases, Lyell in 12, and Lovell in 10. The Yangs' preference for one-to-many segmentation aligns with the explication hypothesis (Blum-Kulka, 1986) ^[3], unpacking complex Chinese clauses into multiple, syntactically simpler English sentences.

Table 3: Syntactic and Expansion Measures

Measure	Lovell (2009) ^[12]	Lyell (1990) ^[15]	Yang & Yang (1960) ^[21]
Mean sentence length (words)	16.3	16.4	16.9
Expansion ratio (EN/ZH)	0.584	0.566	0.708
Segmented pairs (1→N)	10	12	16
English sentences	90	118	96

4.1.3 Readability: Table 4 presents the readability data. The three translations display substantially equivalent Flesch Reading Ease scores (66.9–67.7), all within the "standard" range. Average word length is similarly consistent (4.22–4.31 characters). However, the proportion of long words (>6 characters) reveals subtle differentiation: Lovell uses the highest proportion (18.8%), consistent with her higher lexical density, while Lyell uses the lowest (15.6%). The finding of uniform readability despite divergent stylistic profiles carries an important methodological implication: Flesch Reading Ease may be too coarse to capture the kind of stylistic variation that matters for literary translation.

Table 4: Readability and Lexical Accessibility

Measure	Lovell (2009) ^[12]	Lyell (1990) ^[15]	Yang & Yang (1960) ^[21]
Flesch Reading Ease	66.9	67.2	67.7
Average word length (chars)	4.29	4.31	4.22
Long words (>6 chars), %	18.8%	15.6%	16.4%

4.2 Dialogue Presentation and Stylistic Features

4.2.1 Dialogue representation: Table 5 presents what is arguably the most dramatic finding of the quantitative analysis. Lyell foregrounds dialogue to an exceptional degree: over 40% of his sentences contain direct speech markers - more than double Lovell's proportion and over four times the Yangs'. His version also contains 22 exclamation marks, more than twice as many as either Lovell (9) or the Yangs (10). These features create a dramatically heightened, performance-oriented text in which

the madman's voice is direct, exclamatory, and theatrically intense. By contrast, the Yangs' 9.4% dialogue proportion suggests systematic conversion of direct speech into indirect discourse or narrative summary. Lovell occupies an intermediate position (28.9% dialogue), preserving more of the source text's dialogic texture without amplifying it to Lyell's degree.

Table 5: Dialogue Representation and Dramatic Features

Measure	Lovell (2009) ^[12]	Lyell (1990) ^[15]	Yang & Yang (1960) ^[21]
Dialogue sentences	26/90 (28.9%)	48/118 (40.7%)	9/96 (9.4%)
Exclamation marks	9	22	10
-ly adverbs per 1,000 words	16.4	11.9	14.3

4.2.2 Supplementary lexical features. Two additional measures provide further granularity. Lovell's translation contains 16.4 -ly adverbs per 1,000 words, the highest of the three, compared to 14.3 for the Yangs and 11.9 for Lyell. This distribution is consistent with the lexical-density pattern: Lovell's text incorporates more adverbial modification, contributing to its higher content-word proportion. Lyell's lower adverb frequency aligns with his dialogue-heavy, performative style. Exclamation mark density reinforces Lyell's dramatic profile: normalised by text length, Lyell produces approximately 11.4 exclamation marks per 1,000 words, compared to 6.2 for both Lovell and the Yangs. The madman's voice, in Lyell's hands, is one that shouts; in the Yangs' and Lovell's versions, it is one that reflects.

4.2.3 Qualitative illustration (continued): A second instructive passage occurs in the madman's climactic realisation that his own brother may be a cannibal. The Chinese original reads: 我晓得他们的方法，直捷杀了，是不肯的，而且也不敢，怕有祸祟。所以他们大家连络，布满了罗网，逼我自戕。 Lyell renders this as a dramatic monologue replete with Victorian gothic diction: "I know their methods. They would never dare kill me outright, for they would dread the retribution such a deed would call down upon them. No, they have all joined together to weave a net around me - they plan to force me to take my own life!" Lovell opts for clipped, modern concision: "I know their way of doing things: they're not willing to kill me straight out, and they wouldn't dare, for fear of the consequences. So they've all joined forces, setting traps everywhere, to push me into doing away with myself." The Yangs produce a grammatically balanced but emotionally muted rendering. Each version constructs a different madman: Lyell's is theatrical and baroque, Lovell's is urgent and contemporary, and the Yangs' is restrained and reportorial. These differences in narrative voice - the central stylistic dimension of a first-person diary narrative - are precisely what the quantitative measures of dialogue proportion, exclamation density, and lexical diversity capture at scale.

4.2.4 Translator-specific patterns: Beyond the aggregate measures, fine-grained analysis reveals systematic translator-specific choices that contribute to the quantitative profiles. Lyell consistently selects elevated, Latinate diction

where Lovell chooses Anglo-Saxon equivalents and the Yangs choose neutral academic prose. Compare: "apprised me" (Lyell) versus "told me" (Lovell) versus "told me" (Yangs); "bereaved of knowledge" (Lyell) versus "lost touch" (Lovell and Yangs); "dread disease" (Lyell) versus "seriously ill" (Lovell) versus "seriously ill" (Yangs). Lyell's lexical register is not merely more varied but systematically more archaic and formally marked. Lovell's register, while lexically diverse in the aggregate, achieves this diversity through contemporary idiomatic range rather than through archaism. The Yangs' register, by contrast, is notably consistent and lexically restrained, prioritising denotational accuracy over stylistic distinctiveness.

These patterns have implications for character construction. The madman's voice - paranoid, fragmented, oscillating between lucidity and delusion - is arguably the story's primary literary achievement. Lyell constructs a madman who sounds like a 19th-century Gothic protagonist; Lovell constructs one who sounds like a contemporary psychological case study; the Yangs construct one who sounds like a translated literary artefact. None is "correct" or "incorrect" in a denotational sense - each is stylistically coherent in its own terms - but they produce substantially different reading experiences.

4.2.5 Qualitative illustration: The patterns are vividly illustrated by the story's opening: a frame narrative where the narrator introduces the madman's diary. The Chinese source

opens:

某君昆仲，今隱其名，皆余昔日在中学時良友；分隔多年，消息漸闕。The Yangs produce one compact, grammatically complete sentence (23 words). Lyell produces three sentences (61 words), expanding the source with elaborate, archaising diction ("beg your indulgence", "boon companions", "rend us asunder", "bereft of knowledge"). Lovell produces two sentences (33 words), achieving a conversational register ("close friends", "as the years went by") that is more accessible than Lyell's but more expansive than the Yangs'. This single example illustrates the full spectrum: Lyell's 61-word expansion performs a specific stylistic operation - infusing the frame narrator with a mock-pedantic, self-consciously literary voice - while the Yangs' concision strips away this layer, presenting the narrative frame as transparent exposition.

5. Discussion

5.1 Interpretation of Stylistic Profiles

The corpus-stylistic analysis reveals three distinct translational profiles. Lovell's version is characterised by high lexical density and TTR, moderate dialogue representation, moderate expansion, and a register that combines accessibility (Flesch 66.9) with lexical richness (highest proportion of long words). Lyell's version is characterised by the highest word count, the highest dialogue proportion (40.7%), the most exclamation marks (22), the lowest expansion ratio, and an archaising, performative register. The Yangs' version is characterised by the lowest dialogue proportion (9.4%), the highest expansion ratio (0.708), the most segmented sentence structure, and the lowest lexical density.

The comparison also illuminates how quantitative stylistic profiles can either confirm or complicate qualitative

scholarly consensus. The characterisation of Lyell as the most stylistically marked translator - advanced by Kowallis (1994) ^[10] on the basis of close reading and reinforced by subsequent scholarship on his translation of Lu Xun - is robustly confirmed by the corpus data: highest dialogue proportion, highest exclamation count, lowest expansion ratio (indicating greatest concision per Chinese character), and most distinctive lexical register. The characterisation of the Yangs as producing "faithful but stylistically flat" translations (Henderson, 2007) ^[22] is similarly confirmed by the lowest lexical density, lowest dialogue, and highest grammatical explicitness. Lovell's self-description as producing "accessible, contemporary" translations is confirmed by the Flesch Reading Ease convergence and the idiomatic range of her lexical choices, but her highest TTR and lexical density complicate any simple equation of accessibility with stylistic simplicity. Lovell's translation is both accessible and lexically rich - a combination that challenges the implicit assumption, in both translation criticism and pedagogy, that accessibility requires stylistic reduction.

The three profiles also suggest that translational style, as operationalised through corpus measures, is not a single dimension along which translators can be ranked from "more faithful" to "less faithful" or from "more literary" to "less literary." Rather, it is a multi-dimensional space in which translators make trade-offs: between lexical richness and syntactic simplicity, between dramatic intensity and narrative coherence, between cultural specificity and reader accessibility. Lovell trades syntactic compression for lexical diversity; Lyell trades concision for dramatic force; the Yangs trade stylistic distinctiveness for grammatical clarity. Each trade-off is reasonable given the translator's brief, audience, and institutional context. None is inherently superior.

These profiles align instructively with the translators' stated philosophies and publishing contexts. The Yangs' explicitation-and-expansion-heavy profile is consistent with their stated commitment to "faithfulness", interpreted as complete grammatical articulation of source-text meaning rather than stylistic mimesis. Their near-elimination of direct dialogue (9.4%) may reflect a conception of faithfulness that prioritises semantic content over formal presentation. Lyell's dialogue-heavy, exclamatory, archaising profile is consistent with his scholarly commitment to preserving *Lu Xun's stylistic texture* - the jarring, heightened quality of vernacular prose that departed from the decorum of classical Chinese literature. Lovell's lexically rich yet accessible profile aligns with her Penguin Classics brief: her version achieves the highest lexical diversity without sacrificing readability, consistent with the Penguin Classics house style of producing translations that read as literature in English while retaining a sense of the source text's stylistic distinction.

5.2 Theoretical and Practical Implications

On a theoretical level, the systematic stylistic variations observed among the translations suggest that different linguistic renderings do not merely change the text's surface veneer but fundamentally reconstruct the narrator's identity and psychological profile for target readers. The marked divergence in dialogue representation-with Lyell's highly

dialogic rendering (40.7%) contrasting sharply with the Yangs' narration-heavy approach (9.4%)-directly shapes reader perception of the narrator's reliability and social engagement in *Kuangren Riji* (*A Madman's Diary*), constructing either a confrontational, paranoid protagonist or a deeply introspective, ruminative one. This divergence highlights how individual translator strategies interact with source-text features to navigate target-culture literary expectations, such as the tension between direct dramatic representation and indirect reportorial discourse. Furthermore, these findings complicate established paradigms in translation theory, notably Berman's linear retranslation hypothesis. Rather than demonstrating a progressive, teleological shift toward source-text orientation over time, the chronological sequence of these translations suggests that retranslation is a non-linear process driven by historically situated publishing conditions, institutional patronage, and changing target reader expectations. Consequently, the "canonical Lu Xun" encountered in Anglophone academic and literary spheres is revealed to be a composite, highly fluid, and translator-mediated construct rather than a stable, singular monolith, indicating that literary canonisation in translation must be studied as an ongoing, plural negotiation. In practical and methodological terms, this study demonstrates the utility of combining quantitative corpus-stylistic metrics with qualitative textual analysis to yield a more nuanced understanding of literary translation. While statistical measures such as type-token ratios, expansion indices, and dialogue proportions help reveal systematic patterns that might escape impressionistic observation, they remain descriptive and require qualitative contextualisation to explain the underlying strategic trade-offs. Methodologically, this mixed-methods approach serves as a potential template for investigating retranslation phenomena in other canonical works. In the context of translation pedagogy, these findings suggest a constructive shift away from traditional prescriptive models that privilege a single "correct" version. By presenting multiple translated versions of *Kuangren Riji* side by side alongside corpus-derived profiles, educators can encourage students to conceptualise translation as a series of deliberate, functional choices and strategic negotiations. Using corpus diagnostic tools in the classroom can further enable students to visualise and evaluate their own stylistic habits, allowing them to understand the practical implications of translating for different briefs, target readerships, and institutional contexts.

6. Conclusion

This study has provided the first systematic, corpus-based comparison of three major English translations of Lu Xun's *Kuangren Riji*. The findings reveal three distinct translational styles - the Yangs' explicitation-driven, grammatically explicit, narration-oriented rendering; Lyell's dramatically heightened, archaising, dialogue-intensive version; and Lovell's lexically rich yet accessible, reader-oriented translation - that construct three different versions of Lu Xun's madman for Anglophone readers. These stylistic differences are substantial and systematic, yet they coexist with a striking convergence in surface-level readability, suggesting that translational style operates at a level of lexical, syntactic, and presentational choice that

composite readability measures cannot capture. The corpus-stylistic methods employed here - TTR, STTR, expansion ratio, dialogue proportion - are rudimentary by the standards of contemporary computational linguistics, yet their systematic application to three translations of the same canonical source reveals patterns of sufficient magnitude and consistency to warrant the attention of translation scholars, literary critics, and translation pedagogues alike. Future work will extend these methods to additional Lu Xun stories, to additional translators, and to additional linguistic annotation layers, building towards a comprehensive corpus-stylistic map of Lu Xun's multiple English afterlives. Lu Xun's madman speaks in many voices; corpus stylistics helps us hear the differences.

The corpus-stylistic methods employed here - TTR, STTR, expansion ratio, dialogue proportion - are computationally elementary by the standards of contemporary natural language processing. No machine learning models, no neural networks, no transformer architectures were deployed. Yet the systematic application of these simple measures to three translations of the same canonical source reveals patterns of magnitude and consistency sufficient to warrant the attention of translation scholars, literary critics, and translation pedagogues alike. The methodological modesty of the approach is, in a sense, its strength: the measures are transparent, the computations are replicable, and the findings are interpretable without specialised statistical training. Future work will extend these methods to additional Lu Xun stories - *The True Story of Ah Q*, *Medicine*, *Kong Yiji* - and to additional translators, building towards a comprehensive, corpus-stylistic map of Lu Xun's century-long Anglophone reception history. The madman's voice, refracted through the prisms of different translators, different decades, and different institutional contexts, has much more to tell us.

7. References

1. Baker M. Towards a methodology for investigating the style of a literary translator. *Target*. 2000;12(2):241-266. doi:10.1075/target.12.2.04bak.
2. Berman A. La retraduction comme espace de la traduction. *Palimpsestes*. 1990;4:1-7.
3. Blum-Kulka S. Shifts of cohesion and coherence in translation. In: House J, Blum-Kulka S, editors. *Interlingual and Intercultural Communication*. Tübingen: Gunter Narr; c1986. p. 17-35.
4. Deane-Cox S. *Retranslation: Translation, Literature and Reinterpretation*. London: Bloomsbury Academic; 2014.
5. Denton KA. Lu Xun and the discourse of madness. In: Denton KA, editor. *Modern Chinese Literary Thought: Writings on Literature, 1893-1945*. Stanford (CA): Stanford University Press; c1996. p. 222-255.
6. Fischer-Starcke B. *Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries*. London: Continuum; c2010.
7. Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32(3):221-233.
8. Halliday MAK. *An Introduction to Functional Grammar*. 2nd ed. London: Edward Arnold; 1994.
9. Koskinen K, Paloposki O. *Retranslation*. In: Gambier Y, van Doorslaer L, editors. *Handbook of Translation*

- Studies. Vol. 1. Amsterdam: John Benjamins; 2010. p. 294-298.
10. Kowallis J. Diary of a Madman and Other Stories. By Lu Xun, translated by William A. Lyell. China Q. 1994;137:283-284.
 11. Lee LO. Voices from the Iron House: A Study of Lu Xun. Bloomington (IN): Indiana University Press; 1987.
 12. Lovell J. Translator's introduction. In: Lu Xun. *The Real Story of Ah-Q and Other Tales of China: The Complete Fiction of Lu Xun*. Lovell J, translator. London: Penguin Classics; c2009.
 13. Lu X. A Madman's Diary. In: Call to Arms (Nahan). Beijing: Xinchao Press; 1923. (Original work published 1918).
 14. Lyell WA. *Lu Hsün's Vision of Reality*. Berkeley (CA): University of California Press; c1976.
 15. Lyell WA. Translator's introduction. In: Lu Xun. Diary of a Madman and Other Stories. Lyell WA, translator. Honolulu (HI): University of Hawai'i Press; c1990.
 16. Saldanha G. Translator style: Methodological considerations. Translator. 2011;17(1):25-50. doi:10.1080/13556509.2011.10799478.
 17. Semino E, Short M. Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing. London: Routledge; 2004.
 18. Ure J. Lexical density and register differentiation. In: Perren GE, Trim JLM, editors. *Applications of Linguistics*. Cambridge: Cambridge University Press; 1971. p. 443-452.
 19. Venuti L. *The Translator's Invisibility: A History of Translation*. London: Routledge; c1995.
 20. Yang X. *White Tiger: An Autobiography of Yang Xianyi*. Hong Kong: Chinese University Press; c1995.
 21. Yang X, Yang G, translators. A Madman's Diary. In: *Selected Stories of Lu Hsun*. Beijing: Foreign Languages Press; 1960.
 22. Henderson K. The Yangs and their translations: A case study in institutional translation. *Translation Quarterly*. 2007;44:1-28.
 23. Lakoff G, Johnson M. *Metaphors We Live By*. Chicago (IL): University of Chicago Press; 1980.
 24. Laviosa S. *Corpus-Based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi; c2002.
 25. Martin JR, White PRR. *The Language of Evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan; 2005.
 26. Munday J. *Style and Ideology in Translation: Latin American Writing in English*. London: Routledge; 2008.

Creative Commons (CC) License

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.