ISSN:2828-6235
Impact factor:13.77

International Journal of Literature and Language Studies(IJLLS)

COMPARATIVE ANALYSIS OF GOOGLE TRANSLATE AND DEEPL IN ENGLISH–UZBEK LITERARY TRANSLATION

Ismatova Yulduz Nu'mon qizi

PhD, Senior Lecturer, Samarkand State Institute of Foreign Languages

https://orcid.org/0000-0001-8698-6250

Doi: https://doi.org/10.5281/zenodo.23375134

Abstract: This article presents a comparative analysis of two leading neural machine translation systems — Google Translate and DeepL — applied to English–Uzbek literary translation. A corpus of literary texts (short stories and novel excerpts, ~5,000 words) was translated with both systems, and the outputs were analysed using the five-group error typology. Results show that the systems exhibit different error profiles: Google Translate tends toward generalisation and omission, while DeepL tends toward preservation of source structures and transliteration. For literary translation, neither system reaches publishable quality; both require full post-editing, with DeepL output generally requiring somewhat less effort for morphological correction. The article analyses strengths and weaknesses of each system by error group, proposes a usage strategy (system selection and combination based on text characteristics), and discusses implications for post-editing workflow design. The conclusion argues that system comparison should be based on post-editing effort rather than raw output quality, and that for Uzbek, the choice between systems matters less than the post-editor's competence.

Keywords: Google Translate, DeepL, neural machine translation, literary translation, error analysis, post-editing, English–Uzbek.

I. INTRODUCTION

Practitioners of English–Uzbek translation face a practical question: which MT system should they use as the starting point for post-editing? Google Translate and DeepL are the two most accessible options, but their behaviour differs in ways that matter for post-editing effort. This article provides an empirical comparison based on the five-group error typology developed in Article 4 of this series. The comparison is not an abstract quality ranking but a workflow-oriented analysis: which system produces output that requires less post-editing effort for literary texts? (Koponen, 2016; Toral et al., 2018).

II. METHODOLOGY

A corpus of English literary texts (three short stories and two novel excerpts, ~5,000 words total) was selected. Selection criteria: presence of dialogue (to test register handling), descriptive passages (to test style preservation), phraseological units (at least 15 per text), cultural realia (at least 10 per text), and varied sentence structures (simple to complex). Each text was translated into Uzbek using Google Translate (web interface, 2025) and DeepL (web interface, 2025). Outputs were analysed segment by segment using the five-group typology: lexical-semantic, phraseological, morphological-syntactic, cultural-realie, artistic-aesthetic. Each error was classified by group and visibility (open/hidden). For a subsample, estimated post-editing time was recorded by two post-editors.

III. OVERALL ERROR PROFILES

Both systems produced substantial numbers of errors, confirming that neither is suitable for unedited literary publication. Total errors per 1,000 words: Google Translate 45–50; DeepL 42–48 — statistically similar overall, but the distribution across groups differs significantly (see table below).

Error group	Google Translate	DeepL	Dominant error type
Lexical-semantic	16–18 (35%)	14–16 (32%)	meaning selection
Phraseological	9–10 (20%)	8–9 (19%)	literal translation
Morphological-syntactic	12–14 (27%)	10–12 (24%)	form simplification
Cultural-realie	5–7 (12%)	6–8 (15%)	GT: generalisation; DL: transliteration
Artistic-aesthetic	3–4 (7%)	3–4 (7%)	style flattening

The overall similarity hides important differences. Google Translate tends to generalise: rare words are replaced with frequent ones, realia are replaced with generic terms (pound → money), complex structures are simplified aggressively. DeepL tends to preserve: source structures are kept closer to the original, realia are transliterated (pound → funt), complex structures are attempted but sometimes mangled. For post-editing, these tendencies have different implications: Google Translate output requires more 're-creation' (the generalisations must be undone), while DeepL output requires more 'correction' (the preservation attempts must be fixed). Post-editors reported that DeepL output felt 'closer to correct' but 'more awkward', while Google Translate output felt 'smoother but emptier'.

IV. GROUP-BY-GROUP ANALYSIS

Lexical-semantic group. Both systems select frequent meanings over contextually appropriate ones. Google Translate slightly more often chooses the most frequent dictionary equivalent; DeepL slightly more often preserves the source word's register. Example: 'She is cold today' (about a person's manner): GT → 'U bugun sovuq' (literal); DL → 'U bugun sovuq' (same). No system resolves this context-dependent meaning; post-editor correction is required in both cases. Phraseological group. Both systems translate idioms literally: 'to let the cat out of the bag': GT → 'mushukni xaltadan chiqarish'; DL → 'mushukni sumkadan chiqarish'. The noun choice differs (xalta vs sumka) but neither provides the communicative equivalent ('sirni ochib qo'yish'). Post-editing requirement is identical. Morphological-syntactic group. Here a difference emerges: DeepL produces correct Uzbek case and possessive suffixes more consistently; Google Translate occasionally drops suffixes (e.g., 'in their hands' → 'ular qo'lida' vs DL 'ularning qo'llarida'). Post-editors estimated 10–15% less correction time on this group for DeepL output. Cultural-realie group. Clear strategy difference: GT generalises (bank holiday → bayram), DL transliterates (bank holiday → bank bayrami). For literary post-editing, transliteration is generally preferable (it preserves the cultural signal), making DeepL output the better starting point here. Artistic-aesthetic group. Both systems flatten style equally; no difference in authorial voice preservation. Both require full artistic re-creation by the post-editor.

V. POST-EDITING EFFORT COMPARISON

Estimated post-editing time per 1,000 words (average of two post-editors): Google Translate output 55–65 minutes; DeepL output 50–60 minutes — a 5–10% difference favouring DeepL, driven primarily by the morphological-syntactic and cultural-realie groups. However, the

difference is smaller than the variation between post-editors (experienced vs novice: 40–50% difference). This confirms the finding of Article 12: post-editor competence matters more than system choice. The practical recommendation: use DeepL as the default starting point for literary texts (slightly lower effort, better realia handling), but verify its output more carefully for syntax (its preservation tendency can produce awkward structures). Google Translate may be preferable for texts with heavy simplification needs (technical content within literary works, e.g., letters quoted in novels). For maximum reliability, some post-editors report using both systems: translating with both, comparing outputs segment by segment, and post-editing the better one — a practice that adds 10 minutes per 1,000 words but reduces the risk of systematic error adoption from a single system.

VI. USAGE STRATEGY FOR PRACTICE

Based on the comparison, we propose a system selection strategy. For literary texts: DeepL as default (better morphology, realia preservation), with careful syntactic verification. For mixed texts (literary with embedded technical content): Google Translate may serve for technical segments. For maximum quality: dual-system translation with segment comparison. For all cases: full post-editing is mandatory; neither system output is publishable without it. The strategy should be formalised in post-editing guidelines (Article 13), so that post-editors do not waste time re-deciding system choice for each project. Training programmes (Article 9) should include system comparison exercises, so that post-editors develop intuition for each system's error patterns.

III-та. WHY THE SYSTEMS DIFFER: ARCHITECTURAL NOTES

The behavioural differences between Google Translate and DeepL reflect their different architectures and training data. Google Translate uses a massive multilingual transformer trained on the largest available web-crawled corpus; its scale gives it robustness but also a bias toward the 'most probable' output, which manifests as generalisation and simplification. Its training data for Uzbek includes significant web content, much of it official or informational, which biases its register toward the formal. DeepL, while smaller in scale, uses carefully curated training data and a proprietary architecture with particular attention to linguistic fidelity; its preservation tendencies reflect a design philosophy favouring accuracy over fluency. Neither system has substantial training data from Uzbek literary translations — this domain gap explains the identical artistic-aesthetic failures. Neither system has been specifically adapted for Uzbek morphology — hence the shared morphological errors, with DeepL slightly ahead due to its general linguistic precision. Understanding these architectural roots helps post-editors predict each system's behaviour: Google Translate will 'smooth away' unusual features, DeepL will 'preserve' them even when wrong.

IV-та. DETAILED EXAMPLES

The following examples illustrate the characteristic differences across error groups:

Source segment	Google Translate	DeepL	Post-edited
She is cold today. (manner)	U bugun sovuq.	U bugun sovuq.	U bugun sovuqqanlik qildi.
to let the cat out of the bag	mushukni xaltadan chiqarish	mushukni sumkadan chiqarish	sirni ochib qo'yish
bank holiday	bayram	bank bayrami	bank bayrami (rasmiy dam olish kuni)
in their hands	ular qo'lida	ularning qo'llarida	ularning qo'llarida
The long, winding	Uzoq, egri yoq?	Cho'ziq? aniqroq:	Uzoq, dahr? aniqroq:

road stretched into the distance.	aniqroq: Uzoq, burilma yo'l uzoqqa cho'zildi.	Uzun, aylanma yo'l uzoqlikka cho'zildi.	Uzun, qiyshiq yo'l ufq cho'qqisigacha cho'zilib ketdi.
He was under a cloud.	U bulut ostida edi.	U bulut ostida edi.	U gumondagi edi.

VI-та. BEYOND TWO SYSTEMS: LLM TRANSLATION OUTLOOK

The comparison between Google Translate and DeepL is only the beginning of the system landscape. Large language models (ChatGPT, Claude, Gemini) now offer translation capabilities that differ fundamentally from dedicated MT systems: they can follow stylistic instructions ('translate in the style of modernist prose'), preserve context across long passages, and explain their choices. Preliminary observations suggest that LLM translation of literary texts, with appropriate prompting, produces fewer gross errors than both Google Translate and DeepL, particularly in register and style preservation. However, LLMs have their own weaknesses: they may 'improve' the source text (adding interpretive flourishes), hallucinate content not in the source, and produce inconsistent translations across runs. For post-editing workflows, LLM output requires different checking: not error correction but fidelity verification (does the translation say what the source says?). This shifts the post-editor's task from 'detect and correct errors' to 'detect and correct distortions' — a subtle but important difference. As LLM translation matures, comparative studies should be extended to include these systems, and post-editing guidelines updated accordingly (Vieira, 2019).

VII. CONCLUSION

Google Translate and DeepL show different error profiles for English–Uzbek literary translation: Google Translate generalises and simplifies; DeepL preserves and transliterates. Overall error counts are similar, but DeepL output requires 5–10% less post-editing effort, driven by better morphological consistency and more useful realia handling. However, the post-editor competence factor is much larger than the system factor. The practical conclusion: system choice is a second-order decision; post-editor training is the first-order investment. For the developing Uzbek MT ecosystem, systematic comparison studies like this one should be repeated as systems update, and results shared through professional channels. Future research should extend the comparison to additional systems (Yandex, LLM-based translation), larger corpora, and other text genres; and should develop automated pre-editing pipelines optimised for each system's weaknesses.

REFERENCES

1. Koponen M. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort // The Journal of Specialised Translation. 2016. № 25. P. 131–148.
2. Toral A., Wieling M., Way A. Post-editing effort of a novel with statistical and neural machine translation // Frontiers in Digital Humanities. 2018. Vol. 5. Art. 9.
3. Moorkens J., Toral A., Castilho S., Way A. Translators' perceptions of literary post-editing using statistical and neural machine translation // Translation Spaces. 2018. Vol. 7, № 2. P. 240–262.
4. Богданова К. В. Постредактирование машинного перевода как студенческий проект (на материале художественного текста) // Исследования языка и современное гуманитарное знание. 2024. Т. 6, № 3. С. 172–183.

5. Vieira L. N. Post-editing of machine translation // The Routledge Handbook of Translation and Technology / ed. by M. O'Hagan. London: Routledge, 2019. P. 319–335.
6. Hamroyeva Sh., Matyakubova N. Mashina tarjimasida matnni moslashtirish usullari // Uzbekistan: Til va madaniyat. 2023. № 2. B. 6–19.
7. Yendirboyeva Z. Z. Mashina tarjimasi: til va texnologiya kesishmasidagi dolzarb masalalar // American Journal of Social Science. 2025. Vol. 3, № 9.