



IJTIMOYIY-GUMANITAR SOHADA ILMIY-INNOVATSION TADQIQOTLAR

ILMIY METODIK JURNALI

ISSN 3060-5059



VOL.3 № 4

2026

INGLIZ–O‘ZBEK TEMIRYO‘L MUHANDISLIGI TERMINOLOGIYASI UCHUN NEYRON MASHINA TARJIMASI TIZIMLARINI BAHOLASH: SEMANTIK ANIQLIK, STRUKTURAVIY IZCHILLIK VA MILLIY STANDARTLARGA MUVOFIQLIK (2021–2025)

Bekmurodova Sevara

Qarshi davlat universiteti, mustaqil izlanuvchi

Bekmurodov Jahongir

International Innovation University, assistent-o‘qituvchi

Annotatsiya

Ushbu empirik tadqiqot 2021–2025-yillarga oid 59 ta zamonaviy ingliz tilidagi temiryo‘l muhandisligi terminlarini o‘zbek tiliga tarjima qilishda Google Translate, DeepL va Microsoft Translator tizimlarini baholaydi. Tarjimalar uch mezon asosida tahlil qilindi: semantik aniqlik, strukturaviy izchillik va O‘zbekiston temir yo‘llari milliy standartlariga muvofiqlik. Natijalar shuni ko‘rsatdiki, Microsoft Translator semantik aniqlik (54,12%) va standartlarga muvofiqlik (93,22%) bo‘yicha eng yuqori natijaga erishgan, biroq strukturaviy izchillik barcha tizimlarda past darajada qolgan — terminlarning atigi 18,6% uchala platformada bir xil tarjima qilingan, 39,0% esa to‘liq nomuvofiqlikni namoyon etgan. Ushbu natijalar temiryo‘l hujjatlari, xavfsizlik kommunikatsiyasi va Markaziy Osiyoda temiryo‘l infratuzilmasi bo‘yicha xalqaro hamkorlik uchun muhim ahamiyatga ega.

Kalit so‘zlar: neyron mashina tarjimasini, temiryo‘l muhandisligi terminologiyasi, o‘zbek tili, tarjima sifatini baholash, texnik tarjima, O‘zbekiston temir yo‘llari.

ОЦЕНКА СИСТЕМ НЕЙРОННОГО МАШИННОГО ПЕРЕВОДА ДЛЯ АНГЛО- УЗБЕКСКОЙ ТЕРМИНОЛОГИИ ЖЕЛЕЗНОДОРОЖНОГО МАШИНОСТРОЕНИЯ: СЕМАНТИЧЕСКАЯ ТОЧНОСТЬ, СТРУКТУРНАЯ СОГЛАСОВАННОСТЬ И СООТВЕТСТВИЕ НАЦИОНАЛЬНЫМ СТАНДАРТАМ (2021–2025)

Бекмуродова Севара

Каршинский государственный университет, независимый исследователь

Бекмуродов Жахонгир

International Innovation University, ассистент-преподаватель

Аннотация

Данное эмпирическое исследование оценивает системы Google Translate, DeepL и Microsoft Translator при переводе 59 современных терминов железнодорожного машиностроения (2021–2025 гг.) с английского языка на узбекский. Переводы анализировались по трём параметрам: семантическая точность, структурная согласованность и соответствие национальным стандартам O‘zbekiston temir yo‘llari. Результаты показали, что Microsoft Translator достиг наивысшей семантической точности (54,12%) и соответствия стандартам (93,22%), однако структурная согласованность остаётся низкой во всех системах — только 18,6% терминов переведены одинаково всеми тремя платформами, а 39,0% демонстрируют полную несогласованность. Полученные данные имеют важное значение для железнодорожной документации, коммуникации по вопросам безопасности и международного сотрудничества в развитии железнодорожной инфраструктуры Центральной Азии.

Ключевые слова: нейронный машинный перевод, железнодорожная терминология, узбекский язык, оценка качества перевода, технический перевод, O‘zbekiston temir yo‘llari.

EVALUATING NEURAL MACHINE TRANSLATION SYSTEMS FOR ENGLISH– UZBEK RAILWAY ENGINEERING TERMINOLOGY: SEMANTIC ACCURACY, STRUCTURAL CONSISTENCY, AND NATIONAL STANDARD COMPLIANCE (2021– 2025)

Bekmurodova Sevara

Karshi State University, Independent Researcher

Bekmurodov Jahongir

International Innovation University, Assistant Lecturer

Abstract

This empirical study evaluates Google Translate, DeepL, and Microsoft Translator in translating 59 contemporary English railway engineering terms (2021–2025) into Uzbek. Translations were assessed across three dimensions: semantic accuracy, structural consistency, and compliance with national railway standards of O‘zbekiston temir yo‘llari. The results indicate that Microsoft Translator achieved the highest semantic accuracy (54.12%) and standard compliance (93.22%), while structural consistency remained low across all systems—only 18.6% of terms were translated identically by all three platforms, and 39.0% showed complete inconsistency. These findings have direct implications for railway documentation, safety communication, and international collaboration in Central Asian railway development.

Keywords: neural machine translation, railway engineering terminology, Uzbek language, translation quality assessment, technical translation, O‘zbekiston temir yo‘llari.

The digitalization of railway infrastructure between 2021 and 2025 has introduced a wave of specialized terminology—concepts such as *Digital Railway*, *FRMCS* (Future Railway Mobile Communication System), and *Battery Electric Multiple Units*—into international standards and technical documentation. For Uzbekistan, a country undergoing major railway modernization, the accurate translation of such terms is essential for technical documentation and specifications, safety protocols, international collaboration and technology transfer, training and capacity building, and compliance with international standards.

Neural machine translation (NMT) systems have become ubiquitous in technical translation workflows; however, their effectiveness in handling domain-specific, newly emerged terminology remains underexplored, especially for low-resource language pairs such as English–Uzbek.

1.2 Research Objectives

This study aims to:

1. systematically evaluate Google Translate, DeepL, and Microsoft Translator for the translation of English–Uzbek railway terminology;
2. assess semantic accuracy in preserving technical meaning;
3. analyze structural consistency across platforms;
4. evaluate compliance with Uzbek national railway terminology standards; and
5. provide evidence-based recommendations for railway engineering translation practice.

2. Literature Review

2.1 NMT for Technical Terminology

Recent advances in NMT have improved the handling of technical terminology, yet persistent challenges remain in achieving domain-specific fidelity [22, p. 13; 19, p. 284]. Studies emphasize that off-the-shelf NMT systems struggle in engineering, finance, and railway contexts, where mistranslations can cause operational errors [8]. Context-dependent terms require domain-adapted glossaries or integrated neural architectures [13, p. 12; 15, p. 15]. Large language models

show promise, but they still reveal gaps in low-resource languages and domain-specific nuances [22, p. 13].

2.2 Evaluation Metrics

Robust evaluation of NMT in technical domains combines lexical metrics (BLEU, METEOR, CHRF) with semantic scorers (BERTScore, COMET) and human assessment [21, p. 5]. Discrepancies often arise when automatic metrics undervalue outputs preferred by human evaluators [23, p. 11]. Accordingly, a single automatic score is insufficient; multidimensional assessment is essential for technical translation [24, p. 21].

2.3 Challenges in Domain Translation

Key NMT failure modes include morphological variation, errors in long multiword terms, and acronym disambiguation failures [20, p. 17; 17, p. 3]. Uzbek scholars have further highlighted structural disparities between English (prefix-compound formation) and Uzbek (agglutinative-suffixal formation) in term creation, and have advocated for a national standardization body to harmonize Uzbek railway lexicon with international norms [5, p. 4; 4, p. 230; 11, p. 93].

3. Methodology

3.1 Research Design

This study employs a quantitative empirical design emphasizing systematic, measurable evaluation across multiple dimensions. A comparative framework was used to assess three NMT systems by means of standardized scoring and consistency metrics. All translations were collected in November 2025 via publicly available web interfaces, thereby simulating real-world user conditions without API customizations.

3.2 Corpus Development

A corpus of 59 contemporary railway engineering terms was curated on the basis of four criteria:

1. temporal relevance (terms that emerged or gained prominence between 2021 and 2025);
2. technical specificity (domain-specific concepts only);
3. alignment with international standards (ERA, ISO, AREMA); and
4. practical significance for Uzbekistan's railway modernization agenda.

The terms were selected from a pool of over 200 candidates and organized into nine categories, as shown in Table 1.

Table 1. Corpus Categories and Term Distribution

Category	n	Example Terms	Description
Signaling & Control	7	ETCS, ERTMS, ATO, CBTC	Train signaling, automation, collision avoidance
Digital Railway & Automation	7	Digital Twin, Predictive Maintenance	Smart networks, data-driven maintenance
Infrastructure & Track	7	Maglev, Ballastless Track, Slab Track	Advanced track designs for durability
Electrification & Power	7	Pantograph, Regenerative Braking	Power delivery and energy recovery
Rolling Stock & Components	7	EMU, Bogie, Axle Box	Vehicles and mechanical components
Operations & Management	7	Traffic Management System	Scheduling and resource allocation
Safety & Security	6	Platform Screen Doors, Derailment Prevention	Risk mitigation measures
Sustainability & Environment	5	Carbon Footprint, Energy Efficiency	Eco-conscious transport practices
Maintenance & Inspection	6	Ultrasonic Testing, Drone Inspection	Proactive maintenance procedures

3.3 Evaluation Framework

Three evaluation dimensions were applied to all 177 translations (59 terms × 3 systems):

Semantic Accuracy (0–100 scale) assessed:

- completeness of term translation (0–5 points),
- use of standard Uzbek railway terminology (0–5 points),
- appropriate technical register (0–3 points), and
- domain consistency (0–2 points).

Thresholds were defined as follows:

- **High:** ≥ 70
- **Medium:** 40–69
- **Low:** < 40

Structural Consistency classified inter-system agreement as:

- **Highly Consistent** (all three systems match),
- **Moderately Consistent** (two systems match), or
- **Inconsistent** (all three differ),
using automated string comparison.

Standard Compliance evaluated adherence to five *O‘zbekiston temir yo‘llari* criteria:

- use of *temir yo‘l* terminology,
- correct *poezd* spelling,
- appropriate term structure,
- avoidance of unnecessary transliteration, and
- use of established technical vocabulary such as *boshqarish, nazorat, tizim, and himoya*.

The compliance threshold was set at $\geq 50\%$.

4. Results

4.1 Overall System Performance

Table 2. Overall Performance Summary

Metric	Google Translate	DeepL Translator	Microsoft Translator
Avg. Semantic Accuracy	50.28%	42.03%	54.12%
High Accuracy Terms	8/59 (13.6%)	3/59 (5.1%)	8/59 (13.6%)
Avg. Compliance Score	87.29%	84.18%	93.22%
Compliant Terms	46/59 (78.0%)	46/59 (78.0%)	51/59 (86.4%)

Microsoft Translator led on both semantic accuracy (54.12%) and standard compliance (93.22%), suggesting stronger domain adaptation for Uzbek technical content. DeepL scored lowest (42.03%), likely reflecting more limited investment in this low-resource language pair. Google Translate occupied an intermediate position (50.28%).

4.2 Category-Level Semantic Accuracy

Table 3. Category-Level Performance

Category	Best Performer	Avg. Score	Key Observation
Digital Railway & Automation	Microsoft	60.8%	Highest scores; Industry 4.0 terms handled relatively well
Operations & Management	Microsoft	57.1%	Abstract process terms translated effectively
Sustainability & Environment	Google	53.2%	Modern ecological concepts comparatively well covered
Electrification & Power	Microsoft	55.3%	Energy-related vocabulary handled relatively well

Safety & Security	Microsoft	51.8%	Consistent performance on protective barrier terms
Rolling Stock & Components	Google	52.4%	Google performed better on hardware vocabulary
Maintenance & Inspection	Microsoft	49.3%	Moderate performance; technical depth varied
Signaling & Control	Microsoft	48.9%	Acronym density contributed to errors
Infrastructure & Track	Microsoft	47.6%	Lowest performance; physical infrastructure jargon handled poorly

Conceptual and process-oriented terms scored markedly higher (>55%) than hardware-specific or acronym-heavy categories (<50%). *Digital Railway & Automation* (60.8%) and *Operations & Management* (57.1%) performed best, while *Signaling & Control* and *Infrastructure & Track* remained below 49%.

4.3 Structural Consistency

Table 4. Structural Consistency Distribution

Consistency Level	Terms	Percentage	Implication
Highly Consistent	11	18.6%	Rare consensus; mostly simple loanwords
Moderately Consistent	25	42.4%	Two systems agree; useful for cross-checking
Inconsistent	23	39.0%	Complete divergence; substantial standardization risk

A striking 39.0% of terms were translated entirely differently across all three platforms. For example, *Bogie* yielded **Telega** (Google), untranslated **Bogie** (DeepL), and **Vagon telekasi** (Microsoft). Similarly, *Rail Grinding* produced **Relslarni silliqlash**, **Temir yo‘l silliqlash**, and **Relslarni maydalash**, respectively—the last introducing an inappropriate “crushing” connotation. Only simple, established terms such as *Pantograph* (consistently rendered as **Pantograf**) achieved high structural consistency.

4.4 Standard Compliance

Table 5. Compliance Criteria by System

Criterion	Google	DeepL	Microsoft
Uses <i>temir yo‘l</i> terminology	89%	85%	95%
Standard <i>poezd</i> spelling	82%	88%	93%
Appropriate term structure	91%	78%	95%
Avoids unnecessary transliteration	86%	71%	92%
Uses established technical vocabulary	88%	86%	94%

All three systems exceeded the 84% compliance threshold overall. Microsoft’s superiority (93.22%) likely reflects greater exposure to Uzbek regulatory language during training. DeepL’s weakest criterion—avoidance of unnecessary transliteration (71%)—highlights a major gap in low-resource language support.

5. Discussion

5.1 Microsoft Translator’s Performance

Microsoft Translator’s consistent leadership across all metrics appears to be attributable to three main factors:

1. more comprehensive Uzbek-specific terminology databases (likely through Azure Cognitive Services),
2. better handling of Uzbek’s agglutinative morphology, and
3. training data more closely aligned with *O‘zbekiston temir yo‘llari* standards.

Its strategy of expanding acronyms into full Uzbek descriptions—for example, rendering *ETCS* as **Yevropa poezdlarini nazorat qilish tizimi**—enhances usability in technical manuals. However, with only 54.12% semantic accuracy and just 13.6% of terms classified as highly accurate, expert post-editing remains essential.

5.2 The Acronym Problem

Acronym translation diverges markedly across platforms. Google often appends a Uzbek parenthetical explanation, which can be helpful but occasionally non-standard. DeepL leaves most acronyms untranslated, which is least useful for non-specialist Uzbek readers. Microsoft typically provides canonical Uzbek expansions with the highest degree of standard alignment.

For acronym-intensive documents—such as signaling specifications and safety procedures—Microsoft and Google outputs require the least post-editing effort, whereas DeepL outputs may require near-complete revision.

5.3 Structural Inconsistency and Safety Risks

The 39.0% rate of complete inconsistency poses practical risks beyond mere inconvenience. Inconsistent renderings of safety-critical terms—such as *Emergency Braking* appearing as **Favqulodda tormoz** or **Shoshilinch tormoz**—could delay or confuse emergency response procedures. Maintenance procedures may likewise be misapplied when terminology such as *Rail Grinding* is rendered with a meaning closer to “crushing” than “smoothing.” In multi-document railway projects relying on different NMT tools, such fragmentation may undermine interoperability. A centralized English–Uzbek railway terminology database is therefore essential.

5.4 Comparison with Existing Literature

The results align with broader NMT research:

- multiword term error rates increase with n-gram length [17, p. 23];
- dedicated acronym modules improve outcomes [7, p. 26]; and
- morphological rule-based methods can outperform purely neural systems in inflection-heavy contexts [20, p. 63].

The primary contribution of this study lies in offering a pioneering systematic evaluation of English–Uzbek railway terminology that incorporates national standard compliance—a metric absent from prior comparative NMT studies for this language pair.

5.5 Practical Recommendations

For railway organizations:

- prioritize Microsoft Translator for initial draft translation;
- require expert post-editing for all safety-critical content; and
- begin constructing a centralized English–Uzbek–Russian railway terminology database, starting with the 59 terms evaluated in this study.

For technical translators:

- cross-check Microsoft and Google outputs;
- avoid relying on DeepL alone for Uzbek technical translation;
- maintain personal glossaries for frequently recurring terms; and
- apply heightened scrutiny to acronym-heavy signaling and infrastructure categories.

For international partners:

- budget for post-editing, since an estimated 40–50% of translations will require correction;
- define canonical translations for project-critical terms before work begins; and
- mandate human translation for all emergency and safety documentation.

6. Conclusions

This study evaluated three major NMT systems on 59 contemporary English railway engineering terms translated into Uzbek, assessing semantic accuracy, structural consistency, and national standard compliance.

The key findings are as follows.

Microsoft Translator achieved the highest performance in both semantic accuracy

(54.12%) and standard compliance (93.22%), making it the most suitable baseline tool for Uzbek railway translation among the systems examined. Nevertheless, all systems demonstrated only moderate semantic accuracy overall (42–54%), confirming that expert human review remains indispensable in technical railway contexts. The 39.0% rate of complete inconsistency across platforms underscores the urgent need for a standardized Uzbek railway terminology database. Compliance with *O‘zbekiston temir yo‘llari* standards was relatively strong across all systems (84–93%), suggesting that these systems do reflect some degree of exposure to national terminology norms. Category-specific patterns—where digital and operational terms outperformed infrastructure and signaling terms—should guide translation workflow decisions regarding where NMT assistance is most and least reliable. As Uzbekistan continues its railway modernization, investment in terminology standardization, NMT fine-tuning on Uzbek railway corpora, and translator training will yield cumulative benefits for safety, efficiency, and international interoperability.

REFERENCES

1. Alvarez-Vidal, S., & Oliver, A. (2023). Assessing MT with measures of post-editing effort. *Ampersand*, 10, 100125. <https://doi.org/10.1016/j.amper.2023.100125>
2. American Railway Engineering and Maintenance-of-Way Association. (2025). *AREMA 2025 manual for railway engineering*. Lanham, MD: AREMA.
3. Amrhein, C., Moghe, N., Sennrich, R., & Guillou, L. (2024). Machine translation meta-evaluation through challenge sets: ACES and Span-ACES. *Computational Linguistics*, 51(1), 73–125. https://doi.org/10.1162/coli_a_00530
4. Bekmurodova, S. (2025). Structural characteristics of railway terminology formation: A comparative study of English and Uzbek. *Confrencea*, 5(1), 229–243.
5. Bekmurodova, S., & Bekmurodov, J. (2025). Benefits of standardizing global railway terminology in Uzbekistan’s railway system. *Nordic Press*, 8(0008). <https://doi.org/10.5281/zenodo.15166668>
6. Bozorbekov, A. (2023). Railway transport terminology: From ancient layers to new layers. *Journal of Language and Linguistics*, 6(4). <https://doi.org/10.5281/zenodo.10091854>
7. Chopard, D., Corcoran, P., & Spasić, I. (2024). Word sense disambiguation of acronyms in clinical narratives. *Frontiers in Digital Health*, 6, 1282043. <https://doi.org/10.3389/fdgh.2024.1282043>
8. European Union Agency for Railways. (2024). *ERA railway terminology collection*. Retrieved from <https://www.era.europa.eu>
9. Garg, K. D., Shekhar, S., Kumar, A., Goyal, V., Sharma, B., Chengoden, R., & Srivastava, G. (2022). Framework for handling rare word problems in NMT using multi-word expressions. *Applied Sciences*, 12(21), 11038. <https://doi.org/10.3390/app122111038>
10. International Organization for Standardization. (2024). *ISO 9879:2024—Railway applications: Rolling stock terminology*. Geneva: ISO.
11. Kurbanova, I. S. (2025). Issues of equivalence and linguistic approaches in the translation of railway terms. *Ilmiy Anjumanlar*.
12. Lee, S., Lee, J., Moon, H., Park, C., Seo, J., Eo, S., Koo, S., & Lim, H. (2023). A survey on evaluation metrics for machine translation. *Mathematics*, 11(4), 1006. <https://doi.org/10.3390/math11041006>
13. Liu, Z., Chen, Y., & Zhang, J. (2023). Neural machine translation of electrical engineering based on integrated convolutional neural networks. *Electronics*, 12(17), 3604. <https://doi.org/10.3390/electronics12173604>
14. Odermatt, F., Egressy, B., & Wattenhofer, R. (2023). Cascaded beam search: Plug-and-play terminology-forcing for NMT. *arXiv*. <https://doi.org/10.48550/arXiv.2305.14538>
15. Oncevay, A., Smiley, C., & Liu, X. (2025). Domain-specific terminology and machine translation for finance in European languages. In *Proceedings of NAACL-HLT 2025* (pp. 1–15). <https://doi.org/10.18653/v1/2025.naacl-long.140>
16. O‘zbekiston temir yo‘llari. (2024). *Railway standards and terminology guidelines*. Tashkent: Uzbekistan Railways.
17. Shavarani, H. S., & Sarkar, A. (2021). Better NMT by extracting linguistic information from BERT. *arXiv*. Retrieved from <https://arxiv.org/abs/2104.02831>
18. Shavkatov, S. S., & Kurbanova, I. S. (2025). Comparative linguistic analysis of railway terms in English and Uzbek. *International Journal of Advance Scientific Research*, 4(8), 1–9.

19. Ulitkin, I., Filippova, I., Ivanova, N., et al. (2021). Automatic evaluation of the quality of machine translation of a scientific text. *E3S Web of Conferences*, 284. <https://doi.org/10.1051/e3sconf/202128408001>
20. Xu, W., & Carpuat, M. (2021). Rule-based morphological inflection improves neural terminology translation. *arXiv*. Retrieved from <https://arxiv.org/abs/2109.04620>
21. Yameng, P., Zhongchi, Z., & Lin, J. (2025). Translation quality assessment of mainstream NMT tools on multidimensional quality metrics. In *Proceedings of ICAIE 2025*. IEEE. <https://doi.org/10.1109/ICAIE64856.2025.11158658>
22. Zhang, Z., Syed Abdullah, S. N., & Abdullah, M. A. R. (2025). Translation quality in an evolving paradigm: NMT and LLMs in technical domains. *International Journal of English Language Education*, 13(2). <https://doi.org/10.5296/ijele.v13i2.23057>
23. Zhuo, T. Y., Xu, Q., He, X., et al. (2023). Rethinking round-trip translation for MT evaluation. In *Findings of ACL 2023* (pp. 1–15). <https://doi.org/10.18653/v1/2023.findings-acl.22>
24. Zouhar, V., Kloudová, V., Popel, M., et al. (2024). Evaluating optimal reference translations. *Natural Language Processing*. <https://doi.org/10.1017/nlp.2024.3>