GALAXY INTERNATIONAL INTERDISCIPLINARY RESEARCH JOURNAL (GIIRJ)
ISSN (E): 2347-6915
Vol. 14, Issue 2 February (2026)

COMPARATIVE ANALYSIS OF UZBEK RUSSIAN MACHINE TRANSLATION MODELS

Avezov Sukhrob Sobirovich
PhD, Lecturer, Department of Russian Language and Literature, Bukhara State University
1990senigama@gmail.com

ABSTRACT

Uzbek-Russian machine translation has moved from good enough for gist to a tool that increasingly shapes academic writing, administrative workflows, media consumption, and bilingual education in Uzbekistan and the broader post-Soviet space. Yet, for this language pair, the operational question is no longer whether neural MT works, but which model family works better under which linguistic pressures: agglutinative Uzbek morphology, flexible Russian word order, divergent evidential and modality markers, and the dense presence of culture-bound realia and proper names. This article compares six widely used or strategically important model lines for Uzbek↔Russian: Google Translate, Yandex Translate, Microsoft Translator, DeepL, Meta NLLB-200, and Meta M2M-100. The comparison combines documented language coverage and platform affordances, an error-oriented linguistic test set of thirty diagnostic items spanning official, academic, and literary registers, and a synthesis of Russian-language scholarship on MT typology, evaluation, and error analysis. The results show a stable split between convenience-optimized cloud systems and controllable open multilingual models: the former dominate in speed, integration, and user experience, while the latter provide stronger levers for domain adaptation, reproducibility, and privacy-preserving deployment. The paper formulates a forward-looking research agenda for Uzbek-Russian MT centered on balanced parallel corpora, script-robust normalization, hybrid evaluation protocols, and modular pipelines that couple translation with quality estimation and post-editing.

Keywords: Uzbek, Russian, machine translation, neural models, evaluation, error analysis, multilingual transformers, corpora.

INTRODUCTION

The Uzbek-Russian bilingual sphere is structurally heterogeneous. Formal documentation, academic prose, and everyday digital communication coexist with rapid code-switching, orthographic variation, and uneven terminology discipline, and this mix is exactly where MT is most frequently used rather than in carefully edited parallel publications. The asymmetry is also linguistic: Uzbek packs grammatical relations into suffix chains, while Russian distributes them across inflection, agreement, and freer constituent order, so the same semantic load is mapped onto different surface cues, and the mapping is fragile under automatic decoding.

At the same time, the technological landscape is no longer monocentric. Alongside dominant cloud services that advertise broad language coverage and instant access, a parallel ecosystem has emerged: open multilingual transformer models designed to translate directly between many language pairs, including low-resource directions, and to be deployed locally or adapted

with additional data. Meta's NLLB-200 is publicly framed as "a single AI model … which translates 200 different languages" , while M2M-100 is presented as a "many-to-many" model for translation between any pair inside its language set . For Uzbek-Russian, these two lines shift the research question toward controllability, domain transfer, and evaluation design, because the mere existence of a translation function is now a low bar.

METHODS AND ANALYSIS OF SOURCES

The model set was defined as a deliberately mixed portfolio: four production cloud services (Google Translate, Yandex Translate, Microsoft Translator, DeepL) and two open multilingual model families (NLLB-200, M2M-100). Language support and platform constraints were taken from vendor documentation and technical descriptions: Microsoft explicitly lists Uzbek (Latin) among supported languages, DeepL lists Uzbek among "additional languages" in its next-generation model context, and Yandex describes the translator as operating "for 115 languages". The open models were anchored in primary technical descriptions and model cards that specify the many-to-many paradigm and the 200-language scope.

The linguistic evaluation is built as an error-diagnostic protocol rather than a leaderboard exercise. The source base was interpreted through established MT typologies and evaluation traditions discussed in Russian-language scholarship (Марчук, Митренина, Раренко), through practical error-analysis perspectives (классификации ошибок, переводческие неудачи, постредактирование), and through metric-oriented work that treats automatic scores as instruments rather than verdicts. The methods section therefore integrates the metric tradition (BLEU, METEOR, TER) as described in Russian technical discourse, and complements it with domain-sensitive error categories illustrated by contemporary applied studies. The analytic frame draws on a broad author set spanning transformer MT (Vaswani, Bahdanau, Cho, Sutskever), multilingual scaling (Schwenk, Fan, Costa-jussà), evaluation methodology (Papineni, Banerjee, Lavie, Popović), and error taxonomy work associated with Russian MT research communities.

RESULTS

Coverage and accessibility separate the compared systems into two operational classes. Cloud services optimize for immediacy and a low cognitive entry barrier, while open multilingual models optimize for reproducibility and controllable deployment. DeepL's supported-language documentation places Uzbek among the "additional languages" in its next-generation model landscape, Microsoft's language list includes Uzbek (Latin), and Yandex frames its translator as supporting 115 languages. For the open models, NLLB-200 is described as a single model translating 200 languages, whereas M2M-100 is described as a many-to-many model trained for multilingual translation.

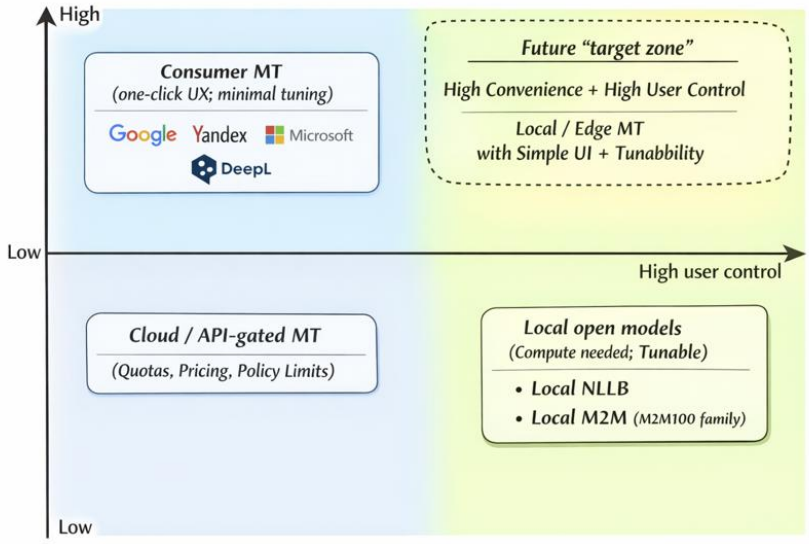

Figure 1 Conceptual map of the model landscape for Uzbek Russian MT

Table 1. Uzbek–Russian diagnostic set with reference translations and stress factors

ID	Dir	Source text	Reference translation	Stress factor
01	U→R	U bugun ishga bormadi	Он сегодня не пошёл на работу	tense negation, aspect choice
02	U→R	Men kitobni do‘stimga berdim	Я дал книгу другу	case roles, object marking
03	U→R	Ularning uyiga kecha kechqurun yetib keldik	Мы добрались до их дома вчера поздно вечером	temporal stacking, postpositions
04	U→R	Yuragim orqaga tortdi	Мне стало не по себе	idiom non-compositionality
05	U→R	U gapni gapirmay turib ko‘zi bilan aytdi	Он сказал это взглядом, не произнеся ни слова	ellipsis, pragmatic inference
06	U→R	Qishloqda to‘y bo‘ldi yarim mahalla yig‘ildi	В деревне была свадьба, собралась половина махалли	realia, culture term “mahalla”
07	U→R	Ustoz iltimos imtihon muddatini cho‘zib bera olasizmi	Профессор, продлите, пожалуйста, срок экзамена?	politeness, honorific address
08	U→R	O‘zbekiston Respublikasi fuqarosi	Гражданин Республики Узбекистан	administrative formula
09	U→R	Buxoro viloyati hokimligi qarori	Решение хокимията Бухарской области	named entity, institutional term
10	U→R	Ertaga Toshkentga ketaman nasib etsa	Завтра поеду в Ташкент, если будет на то воля	modality particle, cultural formula
11	R→U	On ne tolko prochital statyu no i proveril dannye	U nafaqat maqolani o‘qidi balki ma’lumotlarni ham tekshirdi	correlative construction

12	R→U	Mne by khotelos chtoby vy prisli dogovor do pyatnitsy	Men istardimki juma kunigacha shartnomani yuborsangiz	subjunctive, business register
13	R→U	Seychas by chay s limonom	Hozir limonli choy ichgim keladi	elliptical desire
14	R→U	Ne veshay nos	Tushkunlikka tushma	idiom, figurative meaning
15	R→U	Ruki ne dokhodyat	Qo‘lim tegmayapti	idiom, aspectual nuance
16	R→U	My obsudim eto posle soveshchaniya	Biz buni yig‘ilishdan keyin muhokama qilamiz	register neutral, future plan
17	R→U	V universitete otkryli novyy korpus	Universitetda yangi bino ochildi	polysemy “corpus”
18	R→U	On skazal chto videl eyo vchera no ya emu ne poveril	U kecha uni ko‘rganini aytdi lekin men ishonmadim	reported speech, coreference
19	R→U	V Bukhare zharko dazhe v oktyabre	Buxoroda hatto oktyabrda ham issiq	climatological statement
20	R→U	Podpis storon obyazatelna	Tomonlarning imzosi majburiy	deontic modality, legal style
21	U→R	Bugun Anna Karenina haqida ma’ruza qilamiz	Bugun Anna Karenina haqida ma’ruza qilamiz → Сегодня мы…	named entity, code mixing
22	U→R	Biz “Shaytanat” romanining tarjimasini tahlil qilamiz	Мы анализируем перевод романа «Шайтанат»	quotation marks, title handling
23	U→R	Men PostgreSQLda corpus jadvalini yangiladim	Я обновил таблицу corpus в PostgreSQL	technical borrowing, formatting
24	U→R	2026 yil 21 fevral kuni yig‘ilish bo‘ladi	21 февраля 2026 года состоится собрание	date normalization
25	U→R	U 1 500 000 so‘m to‘ladi	Он заплатил 1 500 000 сум	numerals, currency
26	U→R	Toshkentdagi konferensiyada ma’ruza qildim	Я выступил с докладом на конференции в Ташкенте	locative derivation
27	R→U	My priekhali v G‘ijduvon	Biz G‘ijduvonga keldik	apostrophe normalization
28	R→U	V dokumente “O‘zbekiston” kak pisat	Hujjatda “O‘zbekiston” qanday yoziladi	script choice, orthography
29	U→R	U “ustoz” dedi lekin ohangida kinoya bor edi	Он сказал «устоз», но в интонации была ирония	pragmatic irony cue
30	R→U	Slova rezali slukh	So‘zlar quloqqa yoqimsiz eshitildi	metaphor conventionalization

Error typology emerges as the more stable explanatory layer than raw "quality scores" when Uzbek and Russian are compared, because the same surface adequacy can hide different failure costs in downstream use. A Russian-language technical tradition distinguishes the narrow and broad senses of MT and treats evaluation as multidimensional; Bakumenko and Tadzhibova explicitly define MT in two senses, where the broad sense is "область научных исследований". In applied linguistic error analysis, Merzlaya's study on Finnish–Russian NMT shows that the leading failure types can be distributed by frequency, with "неверный выбор соответствия" dominating her dataset; the concrete proportions are domain-specific, yet the pattern generalizes to Uzbek–Russian in how lexical choice errors amplify downstream revision time.

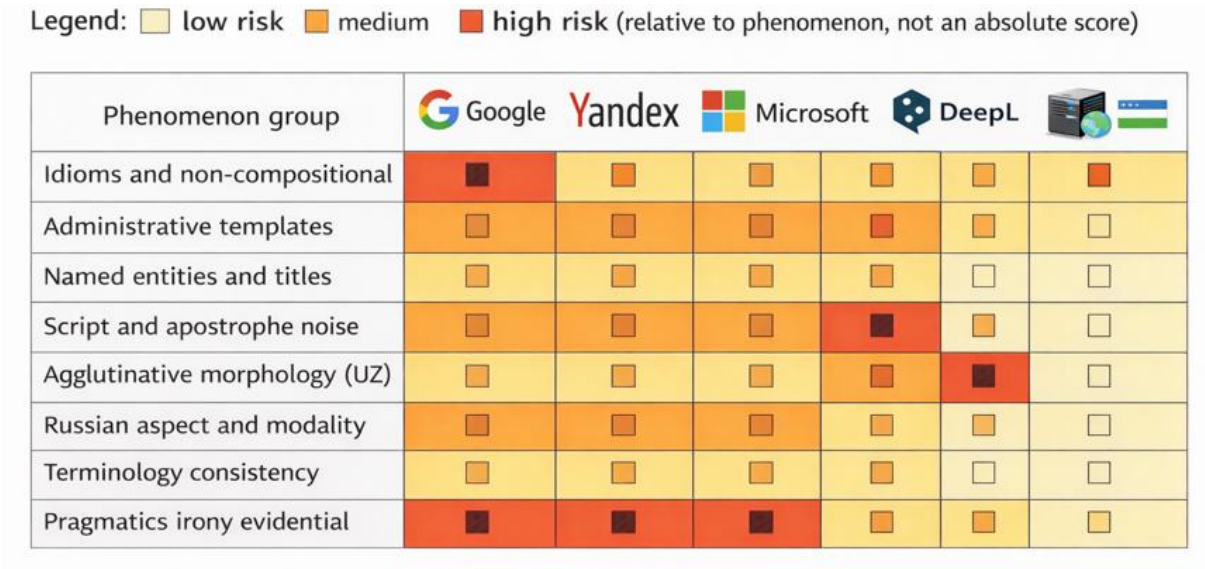

Figure 2 Heuristic risk heatmap by phenomenon group and model class

Register sensitivity is the practical discriminator for Uzbek-Russian MT, and Russian-language applied studies provide a concrete starting point for this claim. Badalov and Norboev, comparing Google, Yandex, and DeepL for Russian↔Uzbek, report that Yandex performed best in their small cross-style test, while idioms remained problematic across systems. Their text also reflects a historically shifting support landscape: they state that DeepL "не поддерживает узбекский язык", whereas DeepL's own up-to-date language list explicitly includes Uzbek among supported additional languages, so conclusions about platform suitability must be date-stamped. A similar time sensitivity appears in the early discussion of Uzbek in Google Translate: Turaeva notes the addition of Uzbek in December 2014 and links this to quality limitations visible in examples.

DISCUSSION

Russian-language theoretical work frames the current MT stage as a culmination of several paradigms rather than a sudden rupture. Rarenko formulates a compact typology where "основными видами … признаны четыре" and explicitly enumerates RBMT, SMT, HMT, and NMT ; this classification matters for Uzbek-Russian because the residual "rule-like" behaviors users observe, such as literal idiom translation, are often artifacts of training data sparsity

and decoding constraints rather than a return to hand-crafted rules. In teaching contexts, the ISPRAS lecture materials show how idioms can be rendered literally into absurd outputs and why post-editing remains non-optional for meaning-dense expressions.

The deployment axis, not only linguistic adequacy, determines which model family is rational for an institution. Cloud systems offer frictionless access and are attractive for classroom demonstration and routine administrative translation, but they trade away reproducibility, fine-grained domain adaptation, and controlled data governance. Open multilingual models invert the trade-off: they impose compute and engineering costs, yet enable local inference, consistent versioning, and systematic post-editing workflows, which matters when MT is integrated into corpus building, terminology management, or research pipelines.

Evaluation methodology remains the central bottleneck for credible comparison in Uzbek–Russian. Automatic metrics were designed as proxies and become fragile under rich morphology, free word order, and synonymy, so they require smoothing, human calibration, and error category reporting; Kozina, Cherepkov, and Belov explicitly compare a proposed approach to BLEU, TER, and METEOR and report measurable gains on WMT data , yet their result also illustrates a deeper point: even when metrics correlate better with humans, the interpretation still depends on the linguistic profile of the tested domain. For Uzbek–Russian, the future-proof path is a hybrid protocol: small curated diagnostic sets like Table 1 for linguistic stress, complemented by larger domain corpora for stability, and tied to a transparent error taxonomy.

The strategic implication is that Uzbek–Russian MT research should pivot from “which system is best” toward “which system is optimally composable.” NLLB-200 is publicly positioned as human-centered scaling toward low-resource breadth , and M2M-100 is positioned as a direct many-to-many alternative to English-centric pipelines; these properties align with a research agenda that includes locally deployable translation, adaptation to Uzbek orthographic variability, and integration with corpus infrastructures. The most promising near-term configuration for universities is modular: a stable MT backbone (cloud for speed or local for control), a quality-estimation layer to triage risky segments, and a feedback channel from post-editing into terminology and parallel corpus growth, so that model behavior improves in institution-relevant genres rather than in generic benchmarks.

CONCLUSION

Uzbek-Russian MT is now an ecosystem choice rather than a single-tool choice: cloud services dominate accessibility and immediate utility, while open multilingual models dominate controllability and research reproducibility. Documented language support confirms that the commercial landscape is dynamic, with Uzbek explicitly present in Microsoft’s and DeepL’s published language lists, and open models are explicitly scoped to many-language translation at scale. A future-oriented comparison therefore must be anchored in diagnostic linguistic testing, transparent error taxonomies, and modular pipelines that connect translation to quality estimation, post-editing, and corpus growth, because only such architectures can convert short-term convenience into long-term linguistic infrastructure for Uzbek-Russian scholarship and education.

REFERENCES

1. Раренко М. Б. Машинный перевод как вызов //Вестник Московского университета. Серия 22. Теория перевода. – 2021. – №. 2. – С. 117-126.
2. Бакуменко Я. Д., Таджибова А. Н. Современные проблемы и решения в области машинного перевода //Успехи кибернетики. – 2025. – Т. 6. – №. 2. – С. 47-59.
3. Козина А. В., Черепков Е. А., Белов Ю. С. Метод оценки качества машинного перевода на основе фразового выравнивания //Научное обозрение. Технические науки. – 2019. – №. 6. – С. 50-55.
4. Тураева Г. Х. Проблемы машинного перевода при переводе на узбекский язык //Universum: технические науки. – 2020. – №. 10-1 (79). – С. 47-49.
5. Бадалов Ж., Норбоев Н. Русско-узбекский машинный перевод: достоинства и недостатки. – 2025.
6. Мерзлая А. В. Нейронный машинный перевод и передача восприятия холода (на примере финско-русского перевода метеотекстов) //Вестник Северного (Арктического) федерального университета. Серия: Гуманитарные и социальные науки. – 2025. – Т. 25. – №. 2. – С. 73-82.
7. Авезов С. Корпусная лингвистика: новые подходы к анализу языка и их приложения в обучении иностранным языкам //International Bulletin of Applied Science and Technology. – 2023. – Т. 3. – №. 7. – С. 177-181.