DOI: https://doi.org/10.37547/supsci-ojp-06-06-51 Pages: 475-486

Oriental Journal of Philology

ORIENTAL JOURNAL OF PHILOLOGY

journal homepage:
http://www.supportscience.uz/index.php/ojp/about

PRAGMATIC AND CULTURAL CHALLENGES IN AI-BASED TRANSLATION OF IDIOMATIC EXPRESSIONS

Iroda Bakhodirovna Utayeva
Dean of the Faculty of Languages, Samarkand State Pedagogical Institute
Doctor of Philosophy (PhD) in Pedagogical Sciences, Acting Professor
Samarkand, Uzbekistan

ABOUT ARTICLE

Key words: artificial intelligence, idiomatic expressions, machine translation, pragmatic equivalence, cultural appropriateness, semantic accuracy, English–Uzbek translation, translation quality assessment.

Received: 25.10.26
Accepted: 26.10.26
Published: 27.10.26

Abstract: The article examines the pragmatic and cultural challenges of translating English idiomatic expressions into Uzbek with artificial intelligence-based tools. Idiom translation is treated as a three-dimensional problem of semantic accuracy, pragmatic equivalence and cultural appropriateness. Emotional, interpersonal and culture-bound idioms are analysed in context and compared with literal renderings and functional Uzbek equivalents, showing how word-for-word transfer can obscure figurative meaning, shift speaker stance or register, and produce images that Uzbek does not recognise as idiomatic. The author proposes the Pragmatic-Cultural Translation Adequacy Index (PCTAI), which combines anchored five-point scales, explicit weights, a semantic gating rule, decision bands and a coding scheme for translation strategies and error types, together with a reproducible protocol for the expert evaluation of Google Translate, DeepL and ChatGPT outputs. The examples are analyst-constructed rather than recorded platform outputs, and no comparative platform scores are claimed. The framework provides a transparent basis for future empirical comparison and for culturally sensitive assessment of machine translation.

SUN'IY INTELLEKTGA ASOSLANGAN TARJIMADA FRAZEOLOGIK IBORALARNING PRAGMATIK VA MADANIY MUAMMOLARI

Iroda Baxodirovna Utayeva
Samarqand davlat pedagogika instituti Tillar fakulteti dekani
pedagogika fanlari bo'yicha falsafa doktori (PhD), professor v.b.

MAQOLA HAQIDA

Kalit so‘zlar: sun’iy intellekt, frazeologik iboralar, mashina tarjimasi, pragmatik ekvivalentlik, madaniy muvofiqlik, semantik aniqlik, inglizcha-o‘zbekcha tarjima, tarjima sifatini baholash.

Annotatsiya: Maqolada ingliz tilidagi frazeologik iboralarni sun’iy intellektga asoslangan vositalar yordamida o‘zbek tiliga tarjima qilishdagi pragmatik va madaniy muammolar o‘rganiladi. Iboralar tarjimasi semantik aniqlik, pragmatik ekvivalentlik va madaniy muvofiqlikdan iborat uch o‘lchovli muammo sifatida talqin etiladi. Hissiy holat, shaxslararo munosabat va madaniy obrazlarga oid iboralar kontekstda tahlil qilinib, so‘zma-so‘z talqinlar hamda o‘zbekcha funksional muqobillar bilan qiyoslanadi. Tahlil so‘zma-so‘z tarjima ko‘chma ma’noni yo‘qotishi, so‘zlovchi munosabati va uslubiy qatlamni o‘zgartirishi hamda o‘zbek tilida ibora sifatida qabul qilinmaydigan obrazlarni yuzaga keltirishi mumkinligini ko‘rsatadi. Muallif langarli besh ballik shkalalar, aniq vaznlar, semantik cheklov qoidasi, qaror chegaralari hamda tarjima strategiyalari va xato turlarini kodlash tizimini birlashtirgan pragmatik-madaniy tarjima adekvatligi indeksini (PCTAI), shuningdek Google Translate, DeepL va ChatGPT natijalarini ekspertlar yordamida baholashning takrorlanadigan protokolini taklif etadi. Keltirilgan misollar muallif tomonidan tuzilgan tahliliy namunalar bo‘lib, platformalardan yozib olingan natijalar emas; platformalarning qiyosiy ballari keltirilmaydi.

ПРАГМАТИЧЕСКИЕ И КУЛЬТУРНЫЕ ПРОБЛЕМЫ ПЕРЕВОДА ФРАЗЕОЛОГИЧЕСКИХ ВЫРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Ирода Баходировна Утаева
Декан факультета языков Самаркандского государственного педагогического института
доктор философии (PhD) по педагогическим наукам, и. о. профессора
Самарканд, Узбекистан

О СТАТЬЕ

Ключевые слова: искусственный интеллект, идиоматические выражения, машинный перевод, прагматическая эквивалентность, культурная адекватность, семантическая точность, англо-узбекский перевод, оценка качества перевода.

Аннотация: В статье рассматриваются прагматические и культурные проблемы перевода английских идиоматических выражений на узбекский язык с помощью систем на основе искусственного интеллекта. Перевод идиом

трактуется как трёхмерная задача, включающая семантическую точность, прагматическую эквивалентность и культурную адекватность. Идиомы, выражающие эмоции, межличностные отношения и культурно обусловленные образы, анализируются в контексте и сопоставляются с буквальными вариантами и функциональными узбекскими эквивалентами. Показано, что пословный перевод может утрачивать переносный смысл, смещать позицию говорящего и стилистический регистр, а также порождать образы, не воспринимаемые в узбекском языке как идиоматичные. Автор предлагает индекс прагматико-культурной адекватности перевода (PCTAI), объединяющий шкалы с описанными уровнями, явные веса, правило семантического порога, пороговые зоны решений и систему кодирования стратегий перевода и типов ошибок, а также воспроизводимый протокол экспертной оценки переводов Google Translate, DeepL и ChatGPT. Приведённые варианты являются аналитическими примерами автора, а не зафиксированными результатами платформ; сравнительные оценки платформ не приводятся.

Introduction. The rapid development of artificial intelligence has transformed multilingual communication and translation practice. Neural machine translation systems and large language models now offer instant translation support to millions of users, including students, teachers and professional translators working between English and Uzbek. Yet the translation of idiomatic expressions remains one of the most persistent weaknesses of these systems. Idioms convey meanings that cannot be derived from their individual words; their interpretation depends on convention, discourse setting and shared cultural knowledge. Grammatical fluency is therefore not a sufficient measure of success: a sentence can be perfectly well-formed in Uzbek and still say something the source author never meant.

Idiomatic expressions combine figurative semantics with social and emotional functions. Expressions such as *break the ice*, *spill the beans* and *on cloud nine* are used, respectively, to ease interpersonal tension, to report the disclosure of a secret and to express intense happiness. A literal transfer of their components into Uzbek may yield utterances that are formally comprehensible but

pragmatically empty or even misleading. A functional translation must account for what the speaker means and what the utterance does in a particular context.

English and Uzbek differ in grammatical structure, phraseological inventory and culturally familiar imagery. The Uzbek idiom *boshi osmonga yetdi* naturally conveys the meaning of *on cloud nine*, although its imagery is different; conversely, *tarvuzi qo'ltig'idan tushdi* expresses sudden disappointment through an image that is opaque to English speakers. Such pairs show that semantic accuracy, preservation of metaphorical form and target-language naturalness are separate properties of a translation and must be evaluated separately.

Classical translation scholarship treats idioms as a problem of equivalence above the word level [1], while research on idiomaticity emphasises conventionalised meanings and discourse functions [2]. Computational studies demonstrate that neural models often process idioms too compositionally and produce overly literal output [3; 4], and recent evaluations of large language models show that their translation quality drops markedly for lower-resource language pairs [5]. English–Uzbek is such a pair, yet idiom-specific evaluation for it remains largely absent.

The aim of this study is to develop and illustrate a comparative pragmatic-cultural framework for evaluating English–Uzbek idiom translation produced by Google Translate, DeepL and ChatGPT. The research questions are: (1) how figurative meaning, communicative purpose and cultural suitability can be assessed in a transparent and reproducible way; (2) which categories of idioms are most likely to require contextual adaptation; and (3) how the quality of a translation can be summarised without concealing the type of failure. The contribution of the study is methodological and analytical: it offers an operational index, a coding scheme and a set of diagnostic examples. It does not claim to have collected or statistically evaluated outputs from the three platforms.

Literature review. M. Baker [1] places idioms within the problem of equivalence above the word level and identifies four main strategies for dealing with non-equivalence: using an idiom of similar meaning and form, using an idiom of similar meaning but dissimilar form, translating by paraphrase and, where unavoidable, translating by omission. This typology is directly relevant to English–Uzbek translation, where comparable communicative functions are frequently expressed through different metaphorical images. It also underlies the principle of dynamic equivalence formulated by E. Nida and C. Taber [6], according to which the response of the target reader, rather than the formal correspondence of words, is the decisive criterion of a translation.

C. Fernando [2] describes idiomaticity as a graded phenomenon: expressions vary in transparency, fixedness and conventionality, and idioms perform ideational, interpersonal and relational functions in discourse – they evaluate, signal stance and organise text. An assessment confined to dictionary meaning can therefore accept a fluent translation that has in fact weakened

humour, softened criticism or altered speaker attitude. For Uzbek, the descriptive basis for identifying conventional equivalents is provided by Sh. Rahmatullayev's explanatory phraseological dictionary [7], which records the meanings, stylistic marking and usage of Uzbek idioms and thus serves as an external check on whether a proposed equivalent is actually established in the language.

P. Koehn and R. Knowles [8] identify domain mismatch, rare words and limited training data among the central challenges of neural machine translation. All three affect idioms, whose frequency and distribution vary sharply across genres. V. Dankers, C. Lucas and I. Titov [3] showed that Transformer models tend to process idioms compositionally and that literal translations are linked to how the model internally represents idiom components. C. Baziotis, P. Mathur and E. Hasler [4] developed automatic measures of literal idiom translation errors and showed that broad quality metrics can miss them, while context and monolingual pretraining reduce such errors. Studies of large language models confirm both their potential and their limits: GPT models are competitive for high-resource pairs but weaker for lower-resource ones [5], and supplying explicit idiom meanings from a knowledge base improves idiomatic translation by language models [9].

Automatic metrics such as BLEU [10] measure n-gram overlap with reference translations and cannot by themselves establish whether figurative meaning, interpersonal stance or cultural imagery has been preserved. J. House's functional-pragmatic model of translation quality assessment [11] argues that a translation must be judged against the function of the source text in its context. Large-scale studies of human evaluation show that expert raters using explicit error categories produce more reliable judgements than crowd ratings of general adequacy [12], and the Multidimensional Quality Metrics framework demonstrates how quality dimensions and error types can be declared in advance and applied systematically [13].

Together, these studies reveal a clear research gap. Theoretical work describes idiomatic meaning and equivalence strategies; computational work documents literal-translation failures and their causes; quality-assessment research shows how to make expert judgement systematic. What is missing for English–Uzbek is a transparent protocol that links concrete idiom examples, context-sensitive expert judgement and a reproducible comparison of named AI systems. The present study proposes such a protocol without presenting hypothetical platform scores as observed findings.

Methodology. The study uses comparative pragmatic-cultural analysis, combining semantic interpretation of the source idiom, functional analysis of its role in context and verification of candidate Uzbek equivalents against the phraseological dictionary [7]. The proposed evaluation corpus comprises 60 English idioms selected purposively from contemporary usage, with 20 expressions in each of three categories: emotional states, interpersonal relations and

culture-bound imagery. Selection criteria are: frequent occurrence in present-day English; a figurative meaning that is not recoverable from the components; and, for half of the items in each category, the existence of an established Uzbek idiom with a comparable function. Every idiom is embedded in a complete English sentence, and twelve items are additionally presented in a literal context (for example, *The ship broke the ice*) to test whether systems detect idiomaticity at all rather than translating every occurrence figuratively.

Data collection protocol. Identical source sentences are submitted to Google Translate, DeepL and ChatGPT within one week. For ChatGPT, a fixed zero-shot instruction ("Translate the following sentence into Uzbek (Latin script).") is used in a new session for each item, and each item is run three times to capture output variability. The full output is retained without post-editing; the date of access, model or system identifier, interface and settings are recorded. Support for the English–Uzbek pair is checked at the time of collection; if a platform cannot translate a given item under the specified conditions, the output is recorded as unavailable rather than reconstructed.

The PCTAI index. Two independent bilingual evaluators rate every translation on three dimensions, each measured on a five-point scale with verbal anchors (Table 1). Semantic accuracy (S) captures whether the figurative sense is preserved; pragmatic equivalence (P) captures whether the communicative function, speaker stance, emotional force and register are preserved; cultural appropriateness (C) captures whether the expression is natural, conventional and culturally acceptable in Uzbek. A contextual reference interpretation prepared in advance by the experts informs the ratings but does not require identical wording.

Table 1. Anchored rating scale of the PCTAI dimensions

Score	Semantic accuracy (S)	Pragmatic equivalence (P)	Cultural appropriateness (C)
5	Figurative meaning fully preserved	Function, stance and register fully preserved	Established, natural Uzbek expression
4	Meaning preserved with minor loss of nuance	Function preserved; slight shift in intensity or register	Natural, though not the most conventional choice
3	Core meaning recoverable only with effort	Function partly preserved; noticeable stance or register shift	Understandable but marked or stylistically unusual
2	Meaning partly distorted	Function largely lost	Unnatural; calque perceived as foreign
1	Literal or wrong meaning	Function lost or reversed	Incomprehensible or culturally inappropriate

Because meaning is a precondition for the other two dimensions, the dimensions are weighted rather than averaged equally. The weighted score is calculated as $W = 0.40 \cdot S + 0.35 \cdot P + 0.25 \cdot C$ and is normalised to a 0–100 scale: $PCTAI = 25 \cdot (W - 1)$. Two decision rules complete the index. First, a semantic gate: if $S \leq 2$, the item is classified as inadequate regardless of its PCTAI value, since cultural naturalness cannot compensate for a wrong meaning. Second, decision bands: $PCTAI \geq 85$ – adequate; 60–84 – acceptable after post-editing; below 60 – inadequate. As a sensitivity check, the unweighted mean $(S + P + C)/3$ is also reported, and conclusions are considered robust only if both versions lead to the same ranking. The weights are a theoretically motivated proposal and are to be re-examined once empirical data are available.

Qualitative coding. Each output is also coded for the strategy used – E1: equivalent Uzbek idiom with a similar image; E2: Uzbek idiom with a different image; PR: paraphrase; L: literal rendering; B: borrowing with explanation; O: omission – and, where relevant, for the error type: loss of figurative meaning, false idiomatic reading of a literal context, stance shift, register shift, or culturally inappropriate image. This coding preserves information that a single number would hide.

Reliability and comparison. Inter-rater agreement is calculated with quadratically weighted Cohen's kappa [14]; a value of at least 0.61 is required, and items on which raters differ by more than one point are resolved by discussion with a third expert. When all three systems produce translations for the same items, differences in PCTAI are tested with the Friedman test [15], followed by Bonferroni-adjusted pairwise comparisons where justified. No outputs of the platforms have been collected for this report; all numerical values below illustrate the procedure and must not be interpreted as observed comparative results.

Results. The analysis below illustrates recurrent semantic, pragmatic and cultural difficulties through English idioms, their literal Uzbek renderings and functionally appropriate Uzbek equivalents. All renderings are analyst-produced examples, not authenticated outputs of Google Translate, DeepL or ChatGPT. Table 2 summarises the diagnostic set.

Table 2. Diagnostic idioms: literal and functional Uzbek renderings

Category	English idiom	Literal Uzbek rendering	Functional Uzbek equivalent	Strategy
Emotional	on cloud nine	to‘qqizinchi bulut ustida	boshi osmonga yetmoq	E2
Emotional	over the moon	oy ustida	do‘ppisini osmonga otmoq	E2
Emotional	butterflies in one’s stomach	qornida kapalaklar	hayajondan yuragi gupillab urmoq	PR

Category	English idiom	Literal Uzbek rendering	Functional Uzbek equivalent	Strategy
Emotional	feel blue	ko‘k his qilmoq	ko‘ngli g‘ash bo‘lmoq	E2
Interpersonal	break the ice	muzni sindirmoq	suhbatdagi noqulaylikni yo‘qotmoq	PR
Interpersonal	spill the beans	loviyani to‘kib yubormoq	gullab qo‘ymoq; sirni ochib qo‘ymoq	E2
Interpersonal	give someone the cold shoulder	sovuq yelka ko‘rsatmoq	yuz o‘girmoq; sovuq muomala qilmoq	E2
Interpersonal	pull someone’s leg	oyog‘idan tortmoq	hazillashmoq; laqillatmoq	E2
Culture-bound	when pigs fly	cho‘chqalar uchganda	tuyaning dumi yerga tekkanda	E2
Culture-bound	kill two birds with one stone	bir tosh bilan ikki qushni o‘ldirmoq	bir o‘q bilan ikki quyonni urmoq	E2
Culture-bound	add fuel to the fire	olovga yoqilg‘i qo‘shmoq	olovga moy sepmoq	E1
Culture-bound	kick the bucket	chelakni tepmoq	oyog‘ini uzatmoq (colloquial)	E2

Emotional idioms primarily test the recognition of figurative affect and its intensity. In *After receiving the scholarship, she was on cloud nine*, the expression describes not a location but an extreme degree of happiness. The rendering *Grantni qo‘lga kiritgach, uning boshi osmonga yetdi* preserves both the emotional force and idiomatic naturalness, whereas *to‘qqizinchi bulut ustida edi* copies the source imagery without forming an Uzbek idiom; an Uzbek reader is likely to search for a literal meaning that does not exist. *Over the moon* allows an even stronger match: *do‘ppisini osmonga otdi* conveys exuberant joy through a culturally specific gesture. In *I had butterflies in my stomach before the interview*, the idiom signals nervous excitement. *Suhbat oldidan hayajondan yuragim gupillab urdi* retains this function, while *qornimda kapalaklar bor edi* may even suggest a physical complaint. The same English idiom can, however, express pleasant anticipation in a romantic context, so the equivalent must be chosen from the context, not from a fixed glossary.

Interpersonal idioms perform social actions rather than describe states. In *The teacher told a funny story to break the ice*, the idiom names the purpose of the action – relieving initial awkwardness. *O‘qituvchi suhbatdagi noqulaylikni yo‘qotish uchun kulgili voqea aytib berdi* is a paraphrase (strategy PR): it preserves meaning and function, but loses figurativeness, which a strict

evaluation records as a minor cost to C rather than as an error. The decisive diagnostic is the literal context: *The ship broke the ice* must be rendered *Kema muzni yorib o‘tdi*; a system that "correctly" paraphrases the idiom but also paraphrases this sentence has not recognised idiomaticity but merely memorised a pattern. *Spill the beans* has a precise colloquial counterpart, *gullab qo‘ymoq*, which conveys not only disclosure but also the speaker's mild disapproval of indiscretion; the neutral *sirni ochib qo‘ydi* is semantically correct but pragmatically flatter. In *Don't worry, I'm just pulling your leg*, the pragmatic core is the playful, non-serious stance: *Xavotir olma, shunchaki hazillashyapman* preserves it, whereas *Seni aldadim* ("I deceived you") keeps part of the propositional meaning but reverses the interpersonal effect.

The example *After their disagreement, Sarah gave him the cold shoulder – Kelishmovchilikdan keyin Sara undan yuz o‘girdi* – is instructive because both languages use coldness as a metaphor of interpersonal distance (*sovuq muomala*), but the conventional expressions differ. Pragmatic equivalence therefore cannot be predicted from metaphorical similarity alone: shared conceptual metaphors do not guarantee shared idioms.

Culture-bound idioms test whether a translation is conventionally acceptable and retains the intended emphasis. *He will apologise when pigs fly* expresses an event regarded as impossible; *U tuyaning dumi yerga tekkanda kechirim so‘raydi* achieves the same effect with an established Uzbek image, while the literal *cho‘chqalar uchganda* introduces an image that is both unidiomatic and culturally marked in a predominantly Muslim speech community. *By attending the conference, she killed two birds with one stone* is naturally rendered *Konferensiyada qatnashib, u bir o‘q bilan ikki quyonni urdi*. *Kick the bucket* shows that register is part of meaning: the neutral *vafot etdi* and the elevated *olamdan o‘tdi* are semantically correct, but only the colloquial *oyog‘ini uzatdi* reproduces the irreverent tone of the source; in formal text, of course, such a rendering would itself be inappropriate, which again shows that evaluation must be tied to context.

Table 3 demonstrates how the PCTAI distinguishes between translations that a single adequacy rating would treat alike. The scores are analyst-assigned illustrations, not data from evaluators.

Table 3. Illustration of PCTAI calculation (hypothetical ratings, not empirical data)

Rendering	S / P / C	PCTAI	Decision	Code
on cloud nine → to‘qqizinchi bulut ustida	1 / 1 / 1	0	Inadequate (gate)	L; loss of figurative meaning
on cloud nine → boshi osmonga yetdi	5 / 5 / 5	100	Adequate	E2

Rendering	S / P / C	PCTAI	Decision	Code
break the ice → muzni sindirmoq	1 / 1 / 2	6.3	Inadequate (gate)	L; loss of figurative meaning
break the ice → suhbatdagi noqulaylikni yo‘qotmoq	5 / 4 / 4	85	Adequate	PR
pull someone’s leg → seni aldadim	3 / 2 / 4	47.5	Inadequate	PR; stance shift
spill the beans → sirni ochib qo‘ydi	5 / 4 / 5	91.3	Adequate	PR; minor stance loss

The calculation is straightforward. For *break the ice → muzni sindirmoq*, W = 0.40·1 + 0.35·1 + 0.25·2 = 1.25, which gives PCTAI = 25·0.25 = 6.3; moreover, S = 1 triggers the semantic gate. For the paraphrase of the same idiom, W = 2.00 + 1.40 + 1.00 = 4.40 and PCTAI = 85, which places it at the lower boundary of the adequate band and correctly reflects the loss of figurativeness. The rendering *seni aldadim* passes the semantic gate (S = 3) and is natural in Uzbek (C = 4), yet its PCTAI of 47.5 classifies it as inadequate because the stance has been reversed – precisely the kind of failure that general adequacy ratings and overlap metrics tend to miss.

Discussion. The analysis suggests a hierarchy of difficulty that can be tested empirically. The first and most basic level is the detection of idiomaticity: a system must decide whether a sequence is used figuratively or literally, and literal-context items are the only way to verify that this decision is made rather than guessed. The second level is semantic interpretation, where literal calques such as *to‘qqizinchi bulut ustida* fail. The third level is pragmatic calibration – preservation of stance, intensity and register – where even semantically correct outputs such as *seni aldadim* or *vafot etdi* can fail. The fourth level is cultural selection, where the choice between an equivalent idiom, a different-image idiom and a paraphrase depends on genre, audience and the literary weight of the source image. Computational findings on over-compositional processing [3; 4] explain failures at the first two levels; the third and fourth levels require the human-centred, function-oriented evaluation advocated by translation quality assessment theory [11].

The analysis also shows that neither of the two common intuitions about idiom translation is reliable. Literal translation is not invariably wrong: where both languages share an established image, as in *add fuel to the fire – olovga moy sepmoq*, near-literal transfer is the optimal strategy. Conversely, free paraphrase is not invariably safe: it usually preserves meaning, but at the cost of figurative style and, sometimes, of stance. Cultural substitution, finally, should be applied with judgement: when a source metaphor carries specific literary, historical or culture-specific value,

preserving it with a brief explanation may serve the reader better than replacing it with a domestic image. This is why the three dimensions must be recorded separately and the strategy coded alongside the score.

The framework has practical implications for three groups of users. For translator training, the diagnostic set and the anchored scale can serve as material for teaching students to check machine output against function rather than form. For developers and users of language models, the results of studies on knowledge-augmented translation [9] suggest that supplying the meaning of an idiom in the instruction may reduce literal errors; the proposed protocol can test whether this holds for Uzbek. For lexicographers, systematic comparison of machine outputs with dictionary data [7] can identify Uzbek idioms whose absence from digital resources leads to recurrent calques. The limitations of the present study are explicit: the examples are illustrative, the weights are a reasoned proposal rather than empirically estimated parameters, and platform behaviour changes over time, so any future comparison must be dated and versioned.

Conclusion. The study has treated English–Uzbek idiom translation by artificial intelligence as a problem involving three interdependent dimensions: semantic accuracy, pragmatic equivalence and cultural appropriateness. The illustrative analysis shows that word-for-word rendering can obscure figurative meaning and interpersonal intention, that semantically correct paraphrases can still shift stance or register, and that established Uzbek idioms – often with a different image – frequently communicate the intended effect most naturally.

The main contribution of the study is the Pragmatic-Cultural Translation Adequacy Index, which combines anchored scales, explicit weights, a semantic gate and decision bands with a qualitative coding of strategies and errors, together with a reproducible data collection protocol that includes literal-context control items. These are methodological proposals, not validated comparative findings. Systematic collection of platform outputs, independent bilingual assessment with verified inter-rater agreement and transparent statistical reporting are necessary before ranking translation systems.

Future research should apply the protocol to the full 60-item corpus, examine the effect of context length and prompting on large language models, estimate the index weights empirically and extend the approach to Uzbek–English translation and to other Turkic languages. Such work would make the evaluation of machine translation of idioms more sensitive to what speakers actually mean and to how meaning is expressed in Uzbek culture.

References:

1. Baker M. In Other Words: A Coursebook on Translation. – London; New York: Routledge, 1992. – 304 p.

2. Fernando C. Idioms and Idiomaticity. – Oxford: Oxford University Press, 1996.

3. Dankers V., Lucas C., Titov I. Can Transformer Be Too Compositional? Analysing Idiom Processing in Neural Machine Translation // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – Dublin, 2022. – P. 3608–3626. – DOI: 10.18653/v1/2022.acl-long.252.

4. Baziotis C., Mathur P., Hasler E. Automatic Evaluation and Analysis of Idioms in Neural Machine Translation // Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. – Dubrovnik, 2023. – P. 3682–3700. – DOI: 10.18653/v1/2023.eacl-main.267.

5. Hendy A., Abdelrehim M., Sharaf A., Raunak V., Gabr M., Matsushita H., Kim Y.J., Afify M., Awadalla H.H. How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation // arXiv preprint. – 2023. – arXiv:2302.09210. – DOI: 10.48550/arXiv.2302.09210.

6. Nida E.A., Taber C.R. The Theory and Practice of Translation. – Leiden: E.J. Brill, 1969.

7. Rahmatullayev Sh. O‘zbek tilining izohli frazeologik lug‘ati. – Toshkent: O‘qituvchi, 1978.

8. Koehn P., Knowles R. Six Challenges for Neural Machine Translation // Proceedings of the First Workshop on Neural Machine Translation. – Vancouver, 2017. – P. 28–39. – DOI: 10.18653/v1/W17-3204.

9. Li S., Chen J., Yuan S., Wu X., Yang H., Tao S., Xiao Y. Translate Meanings, Not Just Words: IdiomKB’s Role in Optimizing Idiomatic Translation with Language Models // Proceedings of the AAAI Conference on Artificial Intelligence. – 2024. – Vol. 38, № 17. – P. 18554–18563. – DOI: 10.1609/aaai.v38i17.29817.

10. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: A Method for Automatic Evaluation of Machine Translation // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. – Philadelphia, 2002. – P. 311–318. – DOI: 10.3115/1073083.1073135.

11. House J. Translation Quality Assessment: Past and Present. – London; New York: Routledge, 2015.

12. Freitag M., Foster G., Grangier D., Ratnakar V., Tan Q., Macherey W. Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation // Transactions of the Association for Computational Linguistics. – 2021. – Vol. 9. – P. 1460–1474. – DOI: 10.1162/tacl_a_00437.

13. Lommel A., Uszkoreit H., Burchardt A. Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics // Tradumàtica. – 2014. – № 12. – P. 455–463. – DOI: 10.5565/rev/tradumatica.77.

14. Cohen J. Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit // Psychological Bulletin. – 1968. – Vol. 70, № 4. – P. 213–220. – DOI: 10.1037/h0026256.

15. Friedman M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance // Journal of the American Statistical Association. – 1937. – Vol. 32, № 200. – P. 675–701. – DOI: 10.1080/01621459.1937.10503522.