ISSN:2828-6235
Impact factor:13.77

International Journal of Literature and Language Studies(IJLLS)

Linguocultural and Pragmatic Challenges in English–Uzbek Machine Translation

Sevara Sultonova
Uzbekistan State University of World Languages,
Faculty of Translation Studies, Department of Practical English Translation, Tashkent, Uzbekistan
doi:https://doi.org/10.5281/zenodo.24090668

Abstract: This article examines the linguocultural and pragmatic challenges that arise when neural machine translation (NMT) systems are used to translate between English and Uzbek. Although modern NMT has improved fluency considerably, its output often fails at the level of meaning in context: culture-specific concepts are lost or replaced by foreign equivalents, politeness and address forms are chosen without regard to social relations, and the illocutionary force of utterances is flattened. Drawing on translation theory, pragmatics and the literature on low-resource machine translation, the article classifies the main problem areas, illustrates them with representative English–Uzbek examples, and proposes a practical framework for human evaluation and post-editing. The study argues that cultural and pragmatic adequacy should be treated as an independent quality dimension alongside accuracy and fluency, and that translators and teachers of translation remain essential in AI-assisted workflows.
Keywords: machine translation; English–Uzbek translation; linguoculture; pragmatics; cultural equivalence; politeness; low-resource languages; post-editing.

Annotatsiya: Ushbu maqolada ingliz va o‘zbek tillari o‘rtasida neyron mashina tarjimasi (NMT) tizimlaridan foydalanishda yuzaga keladigan lingvomadaniy va pragmatik muammolar tahlil qilinadi. Zamonaviy NMT tizimlari matnning ravonligini sezilarli darajada oshirgan bo‘lsa-da, kontekstdagi ma’noni uzatishda hamon xatolarga yo‘l qo‘yadi: madaniyatga xos tushunchalar yo‘qoladi yoki begona ekvivalentlar bilan almashtiriladi, murojaat va xushmuomalalik shakllari ijtimoiy munosabatlarga e’tibor berilmasdan tanlanadi, nutqiy aktlarning illokutiv kuchi susayadi. Maqolada tarjimashunoslik, pragmatika va resursi cheklangan tillar uchun mashina tarjimasi bo‘yicha adabiyotlarga tayanilgan holda asosiy muammoli sohalar tasniflanadi, ingliz–o‘zbek tillaridagi misollar bilan yoritiladi hamda inson tomonidan baholash va post-editing uchun amaliy yondashuv taklif etiladi. Muallif madaniy va pragmatik adekvatlik aniqlik va ravonlik bilan bir qatorda mustaqil sifat mezoni sifatida qaralishi lozimligini, sun’iy intellekt yordamidagi tarjima jarayonida tarjimon va tarjima o‘qituvchisining roli saqlanib qolishini asoslaydi.
Kalit so‘zlar: mashina tarjimasi, ingliz–o‘zbek tarjimasi, lingvomadaniyat, pragmatika, madaniy ekvivalentlik, xushmuomalalik, resursi cheklangan tillar, post-editing.

Аннотация: В статье рассматриваются лингвокультурные и прагматические проблемы, возникающие при использовании систем нейронного машинного перевода (НМП) между английским и узбекским языками. Несмотря на заметное повышение беглости текста,

современные системы по-прежнему допускают ошибки на уровне смысла в контексте: культурно-специфические понятия утрачиваются или заменяются чуждыми эквивалентами, формы обращения и вежливости выбираются без учёта социальных отношений, а иллокутивная сила высказываний ослабляется. На основе теории перевода, прагматики и литературы по машинному переводу для малоресурсных языков классифицируются основные проблемные зоны, приводятся английско-узбекские примеры и предлагается практический подход к экспертной оценке и постредактированию. Автор обосновывает, что культурную и прагматическую адекватность следует рассматривать как самостоятельный критерий качества наряду с точностью и беглостью, а роль переводчика и преподавателя перевода в процессах с участием искусственного интеллекта остаётся необходимой.

Ключевые слова: машинный перевод; англо-узбекский перевод; лингвокультура; прагматика; культурная эквивалентность; вежливость; малоресурсные языки; постредактирование.

1. Introduction

Machine translation (MT) has moved from rule-based and statistical systems to neural architectures, above all the Transformer (Vaswani et al., 2017), and general-purpose systems now produce English–Uzbek text that is often grammatical and readable. For many users this is sufficient for gisting. For translators, teachers and institutions, however, the central question is no longer whether a system can produce a fluent sentence but whether the sentence does what the source text does in its communicative situation.

This question belongs to linguoculture and pragmatics. Language is a carrier of culture (Kramsch, 1998; Vereshchagin & Kostomarov, 1990), and translation is a transfer not only between linguistic codes but between communicative norms (Hatim & Mason, 1990; Baker, 2018). English and Uzbek differ markedly in their typology (analytic with fixed SVO order versus agglutinative with SOV order), in their systems of address and politeness, and in the cultural domains that shape everyday discourse: family and kinship, hospitality, religion, community life and ceremonial practice. Uzbek is also a lower-resource language than the major European languages, which limits the amount of parallel data available for training (Haddow et al., 2022).

The aim of this article is to identify and classify the principal linguocultural and pragmatic challenges in English–Uzbek MT, to illustrate them, and to outline a framework for evaluating and correcting such errors. The article is conceptual and analytical: the examples are representative cases typical of the problem types discussed in the literature and of the author's classroom experience in teaching translation, not the results of a controlled corpus experiment. They should be verified against the output of the specific system in use, since outputs vary between systems and versions.

2. Theoretical background

2.1 Equivalence and culture in translation

The notion of equivalence has been central to translation studies since Nida (1964) distinguished formal and dynamic equivalence, and Newmark (1988) contrasted semantic and communicative translation. Vinay and Darbelnet (1995) described procedures such as borrowing, calque, adaptation and modulation, which remain the translator's main tools for culture-bound

material. In Uzbek scholarship, Salomov (1983) treated the adequacy of translation as the preservation of content, form and national colouring together. Venuti's (1995) opposition between domesticating and foreignizing strategies is also relevant, because MT systems tend, by the statistical nature of their training, to choose the most frequent rendering, which often means a neutral, domesticating one that erases cultural specificity.

2.2 Pragmatics and translation

Pragmatic meaning concerns what speakers do with language and how context shapes interpretation. Austin (1962) and Searle (1969) described speech acts and illocutionary force; Grice (1975) described conversational implicature; Brown and Levinson (1987) proposed a model of politeness based on face. Thomas (1983) introduced the concept of cross-cultural pragmatic failure, the inability to understand or convey what is meant by an utterance, which is directly applicable to MT: a system may translate every word correctly and still fail pragmatically. House (2015) places such functional-pragmatic fit at the core of translation quality assessment.

2.3 Neural MT and low-resource conditions

NMT systems learn mappings from large parallel corpora (Bahdanau et al., 2015; Koehn, 2020). Subword segmentation (Sennrich et al., 2016) helps with Uzbek's rich agglutinative morphology, and multilingual models (Johnson et al., 2017) allow transfer from better-resourced languages. Yet the survey by Haddow et al. (2022) shows that low-resource conditions still bring problems of data quality, domain mismatch and uneven evaluation. For Uzbek, an additional complication is the coexistence of Cyrillic and Latin scripts in available texts (see Salaev et al., 2022, on transliteration between the two), which affects data consistency. Standard automatic metrics such as BLEU (Papineni et al., 2002) measure surface overlap with reference translations and are largely insensitive to cultural and pragmatic appropriateness.

3. Linguocultural challenges

3.1 Culture-specific items

Culture-specific items (CSIs) are words and phrases that name realia of one culture with no ready equivalent in the other (Baker, 2018). Uzbek examples include mahalla (a self-governing neighbourhood community), hashar (collective voluntary work for the community), choyxona (teahouse as a social space), dasturxon (the spread table, with connotations of hospitality and blessing), to'y (a wedding or large family celebration with specific obligations) and osh/palov (a dish with ceremonial significance). In English-to-Uzbek translation, the reverse problem appears with items such as brunch, thanksgiving, prom, tenure or homeowners' association.

Typical MT behaviours at this level are:

- **Literal rendering.** The item is translated word by word (for example, mahalla committee becomes "neighbourhood committee"), losing the institutional and social meaning.
- **Over-generalisation.** A specific realia is replaced with a hypernym (choyxona becomes "cafe" or "restaurant").
- **Omission.** The item is dropped when the system cannot map it.
- **Inconsistent handling.** The same item is rendered differently within one text, which undermines terminological coherence.

A human translator would select a strategy (transliteration with explanation, descriptive equivalent, or cultural substitute) depending on text type, audience and purpose. Current MT does not make this decision explicitly.

3.2 Kinship and social terminology

Uzbek distinguishes kin and quasi-kin by relative age, side of the family and gender far more finely than English: aka and uka, opa and singil, amaki, tog‘a, xola, amma, kelin, qaynona, qaynota and others. These terms are also used for non-relatives as address forms (opa for an unrelated older woman). English uncle, aunt and cousin are underspecified in this regard. Translating from English into Uzbek therefore requires information that is not in the source (which uncle?), and MT systems usually pick the most frequent option without signalling the choice. In the other direction, opa may be rendered “sister” even when it is a polite address to a stranger, producing a misleading text.

3.3 Idioms, proverbs and figurative language

Idioms and proverbs encode culturally specific imagery. English it’s raining cats and dogs has no Uzbek counterpart with the same image, while Uzbek expressions built on the images of bread (non), salt (tuz), the hearth (o‘choq) and the guest (mehmon) carry connotations that a literal English rendering cannot convey. MT systems often produce literal translations for less frequent idioms and only handle well the high-frequency, well-documented ones. The error is particularly dangerous because the output remains grammatical, so non-expert readers cannot detect it.

3.4 Religious and ritual vocabulary

Phrases such as Xudo xohlasa, Alloh rozi bo‘lsin, Baraka toping or Omin function as blessings, thanks or hedges in Uzbek discourse. A literal English rendering (“If God wants”) may be read as a statement of belief rather than as a conventional expression of modesty or good wishes. In the opposite direction, English formulae such as God bless you or Good luck require functional rather than word-for-word equivalents.

4. Pragmatic challenges

4.1 Address forms and the T/V distinction

Uzbek distinguishes informal sen from polite siz, and often adds the plural -lar as an additional marker of respect. English you is unmarked. When translating from English, the system must infer the social relationship from context that is frequently absent at sentence level. The consequence is typically a default choice: siz in formal-looking text, sen in dialogue, or an inconsistent alternation within a single document. A rendering of a student’s email to a professor with sen would be a serious pragmatic error; the reverse, a message between close friends rendered with siz, would sound distant or ironic.

4.2 Politeness and face

Brown and Levinson’s (1987) model predicts that a request can be mitigated by indirectness, hedging, apology or positive politeness. English commonly uses modal and conditional mitigation (Could you possibly…, I was wondering if…). Uzbek uses different devices: honorific pronouns, particles and verb forms, formulae such as iltimos, marhamat, kechirasiz, and a lexicon of respectful verbs. MT tends to transfer English mitigation literally, resulting in unnatural or over-long Uzbek sentences, or to remove it, making a request sound like a command. Neither outcome preserves the original face-work.

4.3 Speech acts and illocutionary force

Illocutionary force is often carried by conventional formulae that are not interpretable compositionally. English How do you do? is a greeting, not a question. The Uzbek greeting set (Charchamang, Ishlaringiz yaxshimi?, Oilangiz yaxshimi?, Yo‘lingiz bo‘lsin) fulfils phatic and relational functions that depend on the situation: Charchamang (“Don’t be tired”) is said to someone at work as an expression of respect. A system that translates literally produces odd or even comic output; a system that translates only the propositional content strips away the social function. Compliments, refusals, invitations and condolences show similar patterns, with refusals in Uzbek usually softened and accompanied by an expression of gratitude.

4.4 Gender, reference and implicature

Uzbek has no grammatical gender and one third-person pronoun, u. In Uzbek-to-English translation, MT must choose he or she, and tends to reproduce statistical biases from training data, for example assigning he to a doctor or engineer and she to a nurse or teacher. In English-to-Uzbek translation, the gender information is simply lost, which may be acceptable or not, depending on whether the gender is relevant to the text. Implicature (Grice, 1975) poses further difficulties: sarcasm, understatement and indirect refusals depend on shared knowledge that is not marked in the text.

4.5 Register, genre and discourse

Administrative, legal, journalistic and academic discourse in Uzbek and English follow different conventions of sentence length, nominalisation, passive use and rhetorical structure. Because the same system translates all genres, it often applies a single “average” register. Official Uzbek letters, for example, require fixed opening and closing formulae and a specific level of formality that a general-purpose system may not apply unless prompted. Discourse-level phenomena such as cohesion, topic–comment structure and consistency of terminology also exceed the sentence-by-sentence translation that many systems still perform (Hatim & Mason, 1990).

5. Illustrative examples

The table summarises typical problem patterns. The “risk” column describes the kind of failure that has been reported for systems of this type, not a measurement of a specific system.

Source (English)	Context	Expected strategy in Uzbek	Typical MT risk
Could you possibly send me the report by Friday?	Email from student to professor	Polite siz form, softened request (iltimos, imkon bo‘lsa)	Literal “possibly” structure or direct imperative
My uncle is visiting us.	Mother’s brother, in a family narrative	Specify tog‘a if known; otherwise ask or flag	Unmarked default (amaki)
Thank God, we made it.	Informal conversation	Functional expression of relief (Xudoga shukr)	Literal “Thank God” without register adjustment
Uzbek:	Greeting to a	“Hello, thank you for your	“Don’t get tired”

Source (English)	Context	Expected strategy in Uzbek	Typical MT risk
Charchamang.	worker	work" or a situational greeting	
Uzbek: Mahalla qo'mitasi yig'ilishi	Local governance report	"Meeting of the mahalla committee (a community self-governance body)"	"Neighbourhood committee meeting"

6. A framework for evaluation and post-editing

Because BLEU-type metrics are insensitive to the phenomena above, a complementary human evaluation is needed. The following framework is proposed for use in translator training and in professional post-editing.

1. **Classify the text.** Identify genre, audience, purpose and the social relationship between the participants before translating.
2. **Mark culture-specific items.** Underline realia, kinship terms, idioms and religious formulae in the source and decide on a strategy (transliteration with gloss, descriptive equivalent, substitution, omission) in advance.
3. **Check address and politeness.** Verify consistent use of sen/siz, plural respect forms and request formulae across the whole text.
4. **Check illocutionary force.** For each speech act (greeting, request, refusal, thanks), ask whether the Uzbek version performs the same act with the same degree of politeness.
5. **Check gender and reference.** Verify pronoun choices and the absence of stereotype-driven assignments.
6. **Score separately.** Rate accuracy, fluency, and cultural–pragmatic adequacy on separate scales (for example, 1–5), so that a fluent but pragmatically inappropriate output does not receive a high overall score.
7. **Improve the input.** Where possible, give the system context (who speaks to whom, in what situation, in what register), and use terminology lists or glossaries for culture-specific items.

In the classroom, this checklist can be applied to comparisons of several systems' outputs, giving students a tool for critical analysis rather than passive acceptance of machine output.

The categories above show that the principal weaknesses of English–Uzbek MT are not mainly grammatical. They arise where meaning depends on knowledge of the culture and of the communicative situation, which is precisely the knowledge least represented in parallel corpora of a low-resource language. Improvements in model size and multilingual transfer will probably reduce some of these errors, especially for frequent items and fixed formulae. Others, such as choosing the right address form without social information, are underdetermined by the text itself and can be solved only by providing context or by human judgment.

This has three implications. First, translation quality evaluation should include cultural and pragmatic criteria explicitly. Second, the creation of high-quality, culturally annotated English–Uzbek parallel data, including dialogue and genre-diverse texts, would be a valuable contribution to the field. Third, translator training should include post-editing skills and critical reading of MT output, with emphasis on linguoculture and pragmatics, since these are the areas where the human translator contributes most.

The limitations of this study should be stated plainly. It does not report a systematic experiment, and the examples are illustrative. Future work should test specific systems on a

balanced set of source texts, with several raters and agreement statistics, to quantify the frequency of each error type.

8. Conclusion

English–Uzbek machine translation has reached a level of fluency that makes its errors harder to detect. The main remaining problems are linguocultural (culture-specific items, kinship terms, idioms, ritual vocabulary) and pragmatic (address forms, politeness, speech acts, gender and reference, register). These can be systematically identified and partly corrected with a structured evaluation and post-editing procedure. Treating cultural and pragmatic adequacy as a separate quality dimension, supporting data development for Uzbek, and training translators in critical post-editing are the most promising directions for making AI-assisted translation both efficient and trustworthy.

References

1. Austin, J. L. (1962). How to Do Things with Words. Oxford: Clarendon Press.
2. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. Proceedings of ICLR 2015.
3. Baker, M. (2018). In Other Words: A Coursebook on Translation (3rd ed.). London: Routledge.
4. Brown, P., & Levinson, S. C. (1987). Politeness: Some Universals in Language Usage. Cambridge: Cambridge University Press.
5. Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and Semantics, Vol. 3: Speech Acts (pp. 41–58). New York: Academic Press.
6. Haddow, B., Bawden, R., Miceli Barone, A. V., Helcl, J., & Birch, A. (2022). Survey of low-resource machine translation. Computational Linguistics, 48(3), 673–732.
7. Hatim, B., & Mason, I. (1990). Discourse and the Translator. London: Longman.
8. House, J. (2015). Translation Quality Assessment: Past and Present. London: Routledge.
9. Johnson, M., Schuster, M., Le, Q. V., et al. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. Transactions of the Association for Computational Linguistics, 5, 339–351.
10. Koehn, P. (2020). Neural Machine Translation. Cambridge: Cambridge University Press.
11. Kramsch, C. (1998). Language and Culture. Oxford: Oxford University Press.
12. Newmark, P. (1988). A Textbook of Translation. New York: Prentice Hall.
13. Nida, E. A. (1964). Toward a Science of Translating. Leiden: Brill.
14. Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the ACL, 311–318.
15. Salaev, U., Kuriyozov, E., & Gómez-Rodríguez, C. (2022). A machine transliteration tool between Uzbek alphabets. arXiv preprint arXiv:2205.09578.
16. Searle, J. R. (1969). Speech Acts: An Essay in the Philosophy of Language. Cambridge: Cambridge University Press.
17. Sennrich, R., Haddow, B., & Birch, A. (2016). Neural machine translation of rare words with subword units. Proceedings of the 54th Annual Meeting of the ACL, 1715–1725.

18. Thomas, J. (1983). Cross-cultural pragmatic failure. Applied Linguistics, 4(2), 91–112.
19. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
20. Venuti, L. (1995). The Translator's Invisibility: A History of Translation. London: Routledge.
21. Vinay, J.-P., & Darbelnet, J. (1995). Comparative Stylistics of French and English: A Methodology for Translation. Amsterdam: John Benjamins.
22. Salomov, G'. (1983). Tarjima nazariyasiga kirish. Toshkent: O'qituvchi.
23. Верещагин, Е. М., Костомаров, В. Г. (1990). Язык и культура. Москва: Русский язык.