BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg

THE LIMITS OF ARTIFICIAL INTELLIGENCE IN TRANSLATING CULTURAL NUANCES: A COMPARATIVE STUDY OF HUMAN AND MACHINE TRANSLATION

Xasanova Durdona Adxam qizi
UzSWLU, theory and practice of translation:Chinese, TXIT-2527
doi:https://doi.org/10.5281/zenodo.23970494

Abstract. This article investigates the limits of artificial intelligence (AI) in rendering cultural nuances through a comparative study of human and machine translation (MT). Its aim is to identify the categories of culturally loaded meaning in which contemporary neural machine translation and large language models most systematically fall short of human translators, and to explain why. Methodologically, the study adopts a qualitative, illustrative comparative design: a purposive sample of culturally loaded source items — idioms, proverbs, culture-specific realia, honorifics and forms of address, kinship terms and pragmatic markers — is analysed by comparing typical MT output with professional human renderings against a cultural-adequacy framework drawn from Newmark and House. The findings show that MT performs well on denotative, context-independent content but falters wherever meaning depends on shared cultural knowledge, pragmatics, affect and register, while human translators retain a decisive advantage as cultural mediators. The article argues for a complementary, human-in-the-loop model rather than full automation, and draws implications for translator training and for lower-resource languages such as Uzbek.

Keywords: machine translation, artificial intelligence, neural machine translation, cultural nuance, human translation, equivalence, pragmatics, untranslatability, Uzbek.

1. Introduction

Advances in artificial intelligence have transformed translation. Neural machine translation (NMT) and, more recently, large language models (LLMs) produce fluent output that, on general-purpose texts, can approach human quality [8]. This progress has prompted claims that professional translation will soon be fully automated. Yet such claims tend to rest on content that is denotative and culturally neutral — the very texts on which MT performs best.

The trajectory of the technology frames the problem. Early rule-based systems encoded linguistic rules by hand; statistical machine translation then learned from large bilingual corpora; and neural models, which translate whole sequences through learned vector representations, brought a marked gain in fluency [7; 8]. Each generation improved surface quality, yet each inherited the same ceiling: the systems learn correlations in text, not the cultural experience that gives words their force. Understanding where this ceiling lies is the aim of the comparison that follows.

Translation theory, however, has long held that translation is not a word-for-word substitution but an act of cultural mediation. Nida distinguished formal from dynamic equivalence, stressing the effect on the receiver rather than surface form [1]; Newmark classified culture-specific categories that resist straightforward transfer [3]; and Catford described cases of «cultural untranslatability», where a source concept has no functional counterpart in the target culture [2]. Cultural nuance — idiom, connotation, politeness, affect, allusion — is precisely where meaning and form diverge, and therefore where automation is most severely tested.

A gap remains in how this is assessed. Much MT evaluation relies on automatic metrics such as BLEU applied to news or technical corpora, which under-measure cultural adequacy and reward surface overlap [8]. Studies of MT on literary and culturally dense text report a persistent quality gap relative to human translators [9]. What is still needed is a systematic, category-based account of where and why AI translation loses cultural nuance.

The aim of this article is therefore to compare human and machine translation across the main categories of culturally loaded meaning, to determine the structural limits of AI in this domain, and to draw practical conclusions — including for lower-resource languages such as Uzbek, which are under-represented in training data. The novelty of the study lies in combining translation-studies categories with a focused comparison oriented toward cultural nuance rather than aggregate fluency.

2. Materials and methods

The study is qualitative and interpretive, using an illustrative comparative design. Rather than aggregate scoring, it examines representative culturally loaded items so that the mechanisms of nuance loss can be observed directly.

A purposive sample of source items was selected across six categories: (1) idioms and proverbs; (2) culture-specific realia (food, dress, concepts); (3) honorifics, register and forms of address; (4) kinship terms and context-dependent polysemy; (5) pragmatics, implicature and connotation; and (6) items from a lower-resource language (Uzbek) translated into English. For each item, typical general-purpose MT/LLM output was compared with a professional human rendering.

Evaluation followed a cultural-adequacy framework synthesised from Newmark's cultural categories and House's model of translation quality assessment (the overt/covert and functional-equivalence distinctions) [3; 5], complemented by Baker's taxonomy of non-equivalence above word level [4] and Catford's notion of cultural untranslatability [2]. Each comparison was coded for the type of nuance at risk. The design has clear limitations — the sample is small and illustrative, the technology evolves rapidly, and outputs vary across systems and versions — which are addressed in the discussion.

3. Results

3.1. Idioms and proverbs. MT most visibly fails on fixed figurative expressions, which it tends to render compositionally. A phrase whose meaning is non-literal is translated word by word, producing output that is grammatical but semantically wrong. Human translators, by contrast, recognise the unit and substitute a functional equivalent or gloss, preserving the intended sense.

3.2. Culture-specific realia. Names of foods, garments, rituals and institutions that have no exact target equivalent are frequently replaced by an approximate or misleading term. Such substitution erases distinctiveness and can even mislead the reader. Human translators deploy explicit strategies — borrowing with a gloss, cultural substitution, or a short explanatory addition — that MT applies only inconsistently.

3.3. Honorifics, register and address. Markers of politeness, deference and affection are routinely flattened. Affective suffixes, honorific titles and the T–V distinction (formal vs informal «you») often collapse into a single neutral form, stripping the text of its interpersonal meaning. Because register signals social relationships, its loss distorts tone and, with it, the reader's perception of the speakers.

3.4. Kinship and polysemy. Context-dependent terms — especially kinship words that map one-to-many across cultures — are mis-selected when the system defaults to the most frequent sense rather than the contextually correct one. Human translators resolve such ambiguity from situational and cultural knowledge that the system lacks.

3.5. Pragmatics, implicature and connotation. The deepest limitation lies in pragmatics: irony, indirect speech acts, politeness strategies and culturally specific connotations depend on what is meant beyond what is said. MT optimises for the literal proposition and therefore misses implicature, humour and emotional colour — features that carry much of a text's cultural identity.

3.6. The lower-resource dimension (Uzbek). For languages with limited parallel data, these problems intensify. Sparser training material and frequent reliance on a high-resource pivot language (e.g. Russian) compound errors and import artefacts of the pivot. As a result, the cultural distance between Uzbek source texts and their English MT output is typically greater than for well-resourced pairs. Table 1 gives illustrative Uzbek–English examples of the recurrent error types, and Figure 1 summarises the two contrasting pathways.

Table 1. Illustrative Uzbek–English examples of cultural-nuance loss in MT

Source item (Uzbek)	Typical MT rendering	Human rendering	Cultural nuance at risk
«Qo'li ochiq»	«his hand is open»	«generous, open-handed»	Idiom read literally
«Osh»	«soup» / «pilaf»	«osh (plov), the festive rice dish»	Culture-specific realia flattened
«Dilnoza-jon»	«Dilnoza» (suffix dropped)	«dear Dilnoza» (affection kept)	Affective / honorific marker lost
«Kelin»	«bride» (always)	«daughter-in-law» (in context)	Polysemy & social role mis-selected
«Bir yostiqqa bosh qo'ygan»	«put heads on one pillow»	«a married couple / sharing a life»	Metaphorical idiom lost

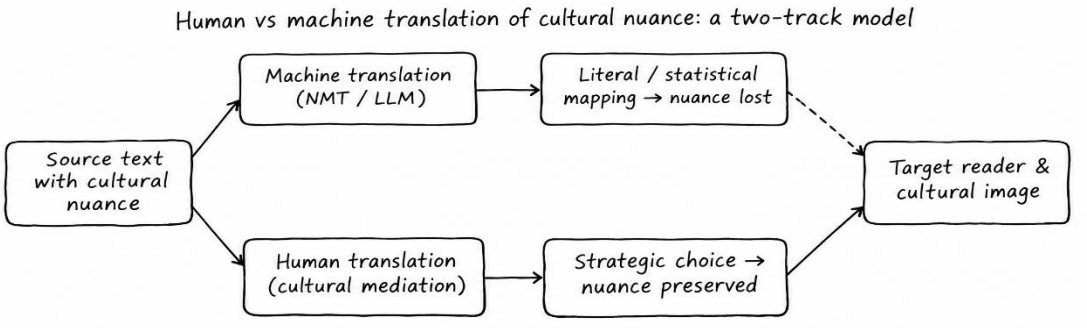


Figure 1. Human vs machine translation of cultural nuance: a two-track model

4. Discussion

The results point to a structural, not merely temporary, explanation. MT — including LLM-based systems — operates by modelling statistical regularities in text; it does not possess the lived, embodied and social knowledge from which cultural meaning arises. It therefore reproduces patterns well but reasons about cultural intention poorly. Training data skewed toward high-resource, denotative content reinforces this bias, and evaluation metrics focused on surface overlap conceal it.

LLMs narrow some of the gap: they handle longer context and can produce idiomatic equivalents when prompted. Yet they also introduce new risks — fluent but fabricated cultural detail («hallucination»), inconsistent register, and an absence of accountability for the choices made.

Fluency, crucially, is not the same as cultural adequacy, and a confident wrong rendering can be more misleading than an obviously literal one.

This also has methodological consequences. Because cultural adequacy is largely invisible to overlap-based metrics, it can be captured only through human-centred evaluation — structured error typologies, expert judgement and, ideally, reception testing with target readers. Relying on automatic scores alone risks declaring parity precisely on the dimension where the gap is widest. Any claim that MT has «matched» human translation must therefore specify the text type and the evaluation method before it can be assessed.

Human translators retain their advantage precisely where theory predicts: as cultural mediators who make deliberate, context-sensitive choices between domestication and foreignization and who bear ethical responsibility for the result [6]. The practical implication is not a contest but complementarity: a human-in-the-loop model in which MT provides a rapid draft and the human performs culturally informed post-editing. The translator's role shifts from first-draft production toward evaluation, cultural judgement and quality assurance.

For Uzbek and other lower-resource languages, the conclusions are concrete. Investment is needed in high-quality parallel data, in direct (non-pivot) language pairs, in the training of translators who can post-edit effectively, and in hybrid workflows for culturally sensitive genres such as literature, diplomacy and heritage. The limitations of the present study — its small, illustrative sample and the fast pace of technological change — mean that its examples are indicative rather than definitive; future work should test these categories on larger, systematically scored corpora and on the latest systems.

5. Conclusion

Artificial intelligence has made translation faster and more accessible and performs strongly on denotative, culturally neutral content. Its limits, however, are clearest where translation matters most as a cultural act: in idiom, realia, honorifics, polysemy and pragmatics, where meaning depends on shared cultural knowledge rather than surface form. Across these categories, human translators remain indispensable as cultural mediators. The most productive path is therefore not full automation but complementarity — AI as a powerful tool within a human-governed process — a conclusion that is especially pressing for lower-resource languages, where cultural nuance is both most valuable and most at risk.

References

1. Nida E. A. Toward a Science of Translating. — Leiden: E. J. Brill, 1964.
2. Catford J. C. A Linguistic Theory of Translation. — Oxford: Oxford University Press, 1965.
3. Newmark P. A Textbook of Translation. — New York: Prentice Hall, 1988.
4. Baker M. In Other Words: A Coursebook on Translation. — 3rd ed. — London; New York: Routledge, 2018.
5. House J. Translation Quality Assessment: Past and Present. — London; New York: Routledge, 2015.
6. Venuti L. The Translator's Invisibility: A History of Translation. — London; New York: Routledge, 1995.
7. Hutchins W. J., Somers H. L. An Introduction to Machine Translation. — London: Academic Press, 1992.
8. Koehn P. Neural Machine Translation. — Cambridge: Cambridge University Press, 2020.

9. Toral A., Way A. What Level of Quality Can Neural Machine Translation Attain on Literary Text? // Translation Quality Assessment / ed. J. Moorkens et al. — Cham: Springer, 2018. — P. 263–287.