INTERNATIONAL MULTI DISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT
eISSN 2394-6334 Impact factor: 7,854 Volume 12, issue 09 (2025)

CULTURAL UNTRANSLATABILITY AND AI'S INTERPRETATIVE LIMITS: LEXICOGRAPHIC IMPLICATIONS FOR LANGUAGE POLICY AND IDENTITY

Yusupova Mushtariy Baxtiyor kizi
Karshi State University
Doctorate (PhD) student
E-mail:mushtariyyusupova1999@gmail.com

Annotation: This paper explores the phenomenon of cultural untranslatability and its implications for AI-driven translation systems. While neural machine translation has significantly advanced linguistic fluency, it continues to face challenges in handling culturally bound expressions. The study also addresses the lexicographic implications of these limitations, particularly the erosion of linguistic identity, the marginalization of minority languages, and the reinforcement of global epistemic imbalances. By advocating for digital sovereignty and decolonial approaches, the paper argues for inclusive AI lexicography that preserves cultural diversity and ensures equitable language representation.

Keywords: cultural untranslatability, AI translation, lexicography, Uzbek language, digital sovereignty, epistemic imbalance, decolonial linguistics

Annotatsiya: Ushbu maqola madaniy untranslatability hodisasi va uning sun’iy intellekt asosidagi tarjima tizimlariga ta’sirini tahlil qiladi. Neyron mashina tarjimasi til ravonligini sezilarli darajada oshirgan bo‘lsa-da, u madaniy jihatdan o‘ziga xos iboralarni tarjima qilishda hanuz qiyinchiliklarga duch kelmoqda. Bu esa algoritmik talqin va inson tafakkuri o‘rtasidagi farqni ko‘rsatadi. Tadqiqot lexikografik oqibatlarga ham e’tibor qaratadi: til identitetining yemirilishi, kichik tillarning chetga surilishi va global epistemologik nomutanosiblikning kuchayishi. Muallif raqamli suverenitet va dekolonial yondashuvlarni ilgari surib, AI lexikografiyasining madaniy xilma-xillikni saqlash va tillarning adolatli vakilligini ta’minlashdagi rolini asoslaydi.

Kalit so‘zlar: madaniy untranslatability, AI tarjima, lexikografiya, o‘zbek tili, raqamli suverenitet, epistemologik nomutanosiblik, dekolonial lingvistika

Аннотация: Данная статья рассматривает феномен культурной непереводимости и его влияние на системы перевода на основе искусственного интеллекта. Несмотря на значительные успехи нейронного машинного перевода в области языковой беглости, он по-прежнему сталкивается с трудностями при передаче культурно-специфических выражений. В статье также анализируются лексикографические последствия этих ограничений: эрозия языковой идентичности, маргинализация малых языков и усиление глобального эпистемологического дисбаланса. Автор подчеркивает важность цифрового суверенитета и деколониальных подходов, утверждая необходимость инклюзивной AI-лексикографии для сохранения культурного разнообразия и справедливого представления языков.

Ключевые слова: культурная непереводимость, AI-перевод, лексикография, узбекский язык, цифровой суверенитет, эпистемологический дисбаланс, деколониальная лингвистика

Introduction

Artificial Intelligence (AI) has become a transformative force in the domain of translation and intercultural communication. While systems such as Google Translate, DeepL, and ChatGPT demonstrate significant progress in fluency and speed, they face persistent limitations when confronted with culturally bound expressions. The phenomenon of cultural untranslatability—where words or phrases resist direct equivalence across languages—remains a fundamental challenge (Nida, 1964; Sapir & Whorf, 1956).

This issue is especially pronounced in languages with deeply embedded cultural, spiritual, or historical contexts, such as Uzbek. Expressions like “qut-baraka,” “o‘zbekona tarbiya,” and “ota-onaning duosi” cannot be adequately translated into English without losing their layered meanings. These examples highlight the epistemological gap between algorithmic interpretation and human cognition.

Furthermore, AI translation tools face difficulties in balancing literal accuracy with idiomatic expression. While Transformer-based models such as GPT-4 and GPT-5 have improved context sensitivity, they still treat language as data rather than lived experience. This distinction becomes critical when translating texts with cultural or religious references, as word choices that seem semantically close may diverge widely in pragmatic force. For example, translating Central Asian proverbs often requires cultural annotations rather than direct substitution. A proverb like “Mehnat qilgan to‘yadi” (literally: one who works will eat well) is not simply about physical nourishment but implies moral legitimacy tied to communal values. AI models rarely account for this layered resonance, which demonstrates the gap between surface-level linguistic training and deeper cultural hermeneutics (Venuti, 1995).

AI Translation Technologies and Their Development

Over the past two decades, translation technologies have shifted from rule-based and statistical systems to neural machine translation (NMT). Neural architectures such as sequence-to-sequence models with attention mechanisms (Bahdanau, Cho, & Bengio, 2015) and the Transformer model (Vaswani et al., 2017) significantly improved translation accuracy.

Despite these advancements, AI models primarily rely on large-scale parallel corpora, which are often biased toward high-resource languages such as English, Mandarin, or Spanish (Bender et al., 2021). This imbalance creates structural disadvantages for minority languages. For Uzbek and other Central Asian languages, limited representation in datasets amplifies the risk of reductionism, where culturally rich expressions are simplified into generic terms.

As a result, AI demonstrates fluency without cultural depth, reflecting a distinction between “linguistic competence” and “pragmatic competence” (Hymes, 1972).

In addition to semiotic theories, anthropological perspectives reinforce the concept of untranslatability. Clifford Geertz (1973) argued that culture is best understood through 'thick description,' an interpretive act that situates symbols in context. Translation without such cultural embedding risks becoming superficial. For example, terms used in Uzbek rituals such as “beshik

to‘yi” (cradle ceremony) or “gap-gashtak” (community gathering) do not have functional equivalents in English. Direct substitution would erase the socio-cultural importance of these practices. As Asad (1986) noted in anthropology, translation always involves power relations; the act of simplifying or omitting cultural concepts is itself a political act. This highlights the stakes of cultural untranslatability in the digital era, where AI-driven tools may unwittingly perpetuate cultural erasure.

Cultural Untranslatability: Theoretical Foundations

The concept of cultural untranslatability has been widely discussed in translation studies and semiotics. Lotman (1990) introduced the notion of the ‘semiosphere,’ emphasizing that cultural texts form unique universes of meaning that resist direct equivalence. Similarly, Sapir and Whorf (1956) argued that language shapes worldview, meaning that some expressions remain untranslatable outside their cultural context.

For instance, “qut-baraka” in Uzbek culture signifies not only blessing but also divine grace, communal well-being, and prosperity. Rendering it merely as “blessing” strips away its multidimensional value. The same is true of “o‘zbekona tarbiya,” which refers to a distinct moral and social upbringing that cannot be reduced to “education” or “training.” These cases illustrate why cultural untranslatability challenges AI systems that lack contextual and embodied knowledge (Venuti, 1995).

Another interpretative limitation of AI is its inability to detect irony, humor, or metaphor rooted in cultural frameworks. For instance, Uzbek humor often relies on wordplay and double meanings that require insider knowledge to appreciate. When AI encounters such expressions, it tends to produce literal translations devoid of humor, flattening the communicative intent. Moreover, AI cannot grasp power dynamics implicit in language use. A phrase uttered respectfully in Uzbek, such as addressing elders with honorifics, may be rendered neutrally in English, losing its social nuance. This demonstrates that AI models, trained primarily on written corpora, lack the multimodal and relational sensitivity that humans naturally deploy in interaction (Hymes, 1972).

AI’s Interpretative Limits

Despite remarkable progress in natural language processing, AI lacks the embodied cognition and cultural situatedness of human translators. Eugene Nida (1964) distinguished between formal equivalence (literal translation) and dynamic equivalence (cultural and functional translation). AI systems often default to the former, producing technically correct but pragmatically insufficient results.

Bender et al. (2021) describe this limitation as “stochastic parroting,” where models replicate patterns from data without genuine understanding. For example, “ota-onaning duosi” is often reduced to “parents’ blessing,” yet in Uzbek culture it implies intergenerational continuity, divine favor, and moral legitimacy. Such nuances require interpretive sensitivity that algorithms cannot easily replicate.

Without pragmatic awareness, AI risks flattening cultural depth into universalized categories, reinforcing epistemic asymmetries between dominant and marginalized languages (Fricker, 2007).

Lexicographic Implications for Language Policy and Identity

Beyond technical translation challenges, cultural untranslatability has implications for lexicography, language policy, and identity. Language is not merely a communication tool but also a vessel of memory, ideology, and collective identity (Fishman, 1999). If AI-powered dictionaries and platforms consistently misinterpret culturally bound terms, several risks arise:

1. **Erosion of linguistic identity:** Native speakers may internalize simplified translations, weakening authentic usage (Spolsky, 2004).

2. **Marginalization of minority languages:** Underrepresentation in digital corpora reinforces linguistic hierarchies (Phillipson, 1992).

3. **Global epistemic imbalance:** Dominant languages shape how AI interprets meaning, marginalizing non-Western worldviews (Mignolo, 2011).

This calls for a decolonial perspective in AI lexicography, where underrepresented cultures are represented on their own terms rather than assimilated into dominant linguistic structures.

Practical initiatives are emerging to resist such erasure. For example, community-driven digital corpora projects in Africa and South Asia demonstrate how localized knowledge can be embedded into AI systems (Bird, 2020). Similar initiatives could be developed for Uzbek and other Central Asian languages, ensuring that AI reflects their epistemic realities.

Lexicographic reduction also risks shaping the way future generations engage with their heritage languages. If students or younger speakers rely on AI tools that consistently mistranslate terms, they may internalize distorted meanings over time. This phenomenon reflects what Phillipson (1992) described as 'linguistic imperialism,' where dominant languages overshadow minority identities. Furthermore, language policies in postcolonial societies often emphasize the revival of indigenous concepts as a means of cultural resistance (Ngugi wa Thiong'o, 1986). AI systems that erase or trivialize these concepts undermine such policy goals. Thus, the stakes of lexicographic representation extend beyond academia to questions of cultural survival and sovereignty.

Moreover, scholars advocate for participatory design in AI development, where local communities are not passive data providers but active co-creators of digital tools (Escobar, 2018). This model aligns with decolonial frameworks, shifting authority away from Western-centric paradigms and toward pluralistic epistemologies.
In the Uzbek context, such approaches would mean integrating oral histories, folklore, and ritual practices into digital corpora, thereby allowing AI not only to translate but also to represent cultural worlds.

Digital Sovereignty and Decolonial Approaches

To counteract these imbalances, scholars argue for digital sovereignty and decolonial approaches in language technology (Ngugi wa Thiong'o, 1986; Mignolo, 2011). Digital sovereignty entails empowering communities to build and control their own linguistic corpora, dictionaries, and translation models.

For Uzbek and other underrepresented languages, this means fostering AI tools that embed cultural values rather than erasing them. Such initiatives align with broader movements for

cultural sustainability, where digital technologies serve as instruments for preserving, not diluting, identity.

Decolonial lexicography emphasizes inclusivity, cultural self-representation, and resisting epistemic injustice in global digital infrastructures (Fricker, 2007).

Global Epistemological Imbalance

The dominance of English, Mandarin, and Spanish in AI training datasets perpetuates a global epistemological imbalance. Languages with fewer speakers and weaker digital footprints risk being marginalized in global discourse. This phenomenon can be linked to Fricker's (2007) concept of 'epistemic injustice,' where certain groups' knowledge systems are systematically undervalued.

In translation technologies, epistemic injustice manifests in how culturally specific terms are reduced, omitted, or misinterpreted. Such reductionism risks homogenizing pluralistic worldviews into Western-centric categories. Addressing this requires intentional policies and inclusive design in AI development (Bender et al., 2021).

Conclusion

Cultural untranslatability exposes the interpretive limits of AI translation systems and raises profound questions for lexicography, language policy, and identity. While neural machine translation has improved technical fluency, it lacks the pragmatic awareness and cultural depth required to handle worldview-bound expressions.

If left unaddressed, AI risks contributing to linguistic homogenization, epistemic imbalance, and erosion of cultural identity. Conversely, by adopting decolonial lexicographic practices, promoting digital sovereignty, and ensuring inclusive representation of minority languages, AI can serve as a tool for cultural preservation rather than assimilation.

Thus, the challenge of cultural untranslatability is not merely technical but also ethical and political, demanding a rethinking of how language technologies are designed and deployed.

References:

- Asad, T. (1986). The Concept of Cultural Translation in British Social Anthropology. In J. Clifford & G. Marcus (Eds.), **Writing Culture: The Poetics and Politics of Ethnography**. University of California Press.
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. arXiv preprint arXiv:1409.0473.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.
- Bird, S. (2020). Decolonising Speech and Language Technology. **Proceedings of the 28th International Conference on Computational Linguistics**, 3504–3519.
- Fricker, M. (2007). Epistemic Injustice: Power and the Ethics of Knowing. Oxford University Press.
- Geertz, C. (1973). The Interpretation of Cultures. Basic Books.

- Hymes, D. (1972). On Communicative Competence. In J. B. Pride & J. Holmes (Eds.), **Sociolinguistics** (pp. 269–293). Penguin Books.
- Lotman, Y. (1990). Universe of the Mind: A Semiotic Theory of Culture. Indiana University Press.
- Mignolo, W. D. (2011). The Darker Side of Western Modernity: Global Futures, Decolonial Options. Duke University Press.
- Phillipson, R. (1992). Linguistic Imperialism. Oxford University Press.
- Sapir, E., & Whorf, B. L. (1956). Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. MIT Press.
- Spolsky, B. (2004). Language Policy. Cambridge University Press.