

EVALUATING NEURAL MACHINE TRANSLATION ACCURACY FOR SPECIALIZED ARCHITECTURAL TERMINOLOGY

Abdullayeva Zumradxon Yusubjon qizi

PhD researcher

Uzbekistan State World Languages University

zumradibrokhimova@gmail.com

Abstract. This article evaluates the accuracy and limitations of Neural Machine Translation (NMT) and Large Language Models in rendering specialized architectural terminology. The study identifies key semantic risks, including over-generalization and hallucination when processing culturally-bound architectural realia and historical vocabulary. The paper highlights the essential role of Machine Translation Post-Editing and proposes pedagogical strategies for integrating AI tools into translator training.

Keywords: neural machine translation, artificial intelligence, architectural terminology, non-equivalence, cultural realia, post-editing (MTPE), domain-specific translation, translation pedagogy.

Annotatsiya. Ushbu maqolada neyron mashina tarjimasi (NMT) va AI modellarining maxsus me'moriy terminologiyani tarjima qilishdagi aniqligi tahlil qilinadi. Tadqiqotda tarixiy-me'moriy leksika va madaniyatga xos realiyalarni uzatishdagi semantik yo'qotishlar hamda terminologik xatolar ko'rib chiqilgan. Maqolada tarjimonlarni tayyorlashda mashina tarjimasidan so'ng tahrir qilish ko'nikmalarini rivojlantirish bo'yicha amaliy tavsiyalar berilgan.

Kalit so'zlar: neyron mashina tarjimasi, sun'iy intellekt, me'moriy terminologiya, ekvivalentsizlik, madaniyatga xos realiya, post-tahrir (MTPE), sohaviy tarjima, tarjima pedogogikasi.

Аннотация. В статье исследуется точность систем нейронного машинного перевода (NMT) и AI-моделей при передаче специальной архитектурной терминологии. Анализируются основные типы семантических потерь и терминологических ошибок при обработке историко-архитектурной лексики и культурных реалий. На основе полученных результатов предложены методические рекомендации по развитию навыков постотредактирования в процессе подготовки переводчиков.

Ключевые слова: нейронный машинный перевод, искусственный интеллект, архитектурная терминология, безэквивалентность, культурные

реалии, постотредактирование (МТРЕ), специальный перевод, методика обучения.

The rapid evolution of Artificial Intelligence (AI) and Neural Machine Translation (NMT) engines—such as Google Translate, DeepL, and Large Language Model (LLM) architectures—has significantly transformed language services, translation workflows, and Language for Specific Purposes (LSP) pedagogy. AI applications demonstrate remarkable fluency in general language translation; however, their performance in domain-specific technical fields, particularly historical architectural terminology, remains a subject of ongoing academic evaluation.

Architectural discourse presents unique translation challenges. It combines highly technical structural engineering concepts with culturally-bound historical terms, regional building techniques, and ethno-cultural realia. While general technical terms (e.g., *foundation*, *cantilever*, *load-bearing structure*) often possess direct cross-lingual equivalents, historical architectural terms frequently lack direct target-language matches due to distinct regional construction traditions and socio-religious influences.

Evaluating the capacity of AI-driven translation tools to manage non-equivalent architectural terminology is critical not only for maintaining translation accuracy in cultural heritage documentation but also for training future translators in Machine Translation Post-Editing (MTPE).

Neural Machine Translation models operate on deep neural networks that map context through vector embeddings. Large Language Models enhance this approach by analyzing token sequences across vast training datasets. Despite these advances, AI models encounter systemic issues when processing non-equivalent specialized vocabulary. The primary failure mechanisms include:

1. **Over-Generalization (Semantic Compression):** When an AI model encounters a highly specific regional architectural term that is underrepresented in its training corpus, it tends to substitute the term with a broad hypernym (for example, reducing a specific historic vaulted hall to a generic "building" or "room").

2. **False Equivalence (Literal Calque):** AI engines often attempt word-for-word component translation of compound terms, producing target forms that violate domain-specific terminology norms in English.

3. **Hallucination:** Generative AI models may invent plausible-sounding but technically inaccurate architectural definitions when contextual cues are ambiguous.

Due to these potential errors, raw outputs generated by AI tools routinely require structured Human Machine Translation Post-Editing (MTPE) interventions by subject-matter experts. To measure the efficacy of AI translation tools, specialized architectural terms can be classified into three primary operational categories:

1. **Universal Engineering Terms:** This group includes terms describing physical, structural, or geometric properties, such as *arch*, *dome*, *reinforcement*, and *facade*. AI engines demonstrate high accuracy (above 90–95%) in this category due to abundant bilingual parallel data and standardized technical usage. The primary failure mode is limited to minor register mismatches, requiring minimal human post-editing.

2. **Functional Structural Units:** This category encompasses terms describing functional parts of historical complexes, such as *fortifications*, *courtyards*, or *decorative tilework*. AI performance in this area is moderate (ranging between 60% and 75%), frequently suffering from stylistic inconsistencies. Human translators must apply contextual refinement during the post-editing stage to ensure domain-specific appropriateness.

3. **Ethno-Cultural Architectural Realia:** This classification includes unique historical building types, decorative techniques, and regional architectural forms that completely lack target-language equivalents. In this category, unassisted AI translation accuracy drops below 40%. AI engines routinely fail by over-generalizing concepts or hallucinating definitions. For these terms, human experts must override machine translation using specialized strategies, such as transliteration combined with descriptive glosses.

The findings underscore that while AI applications cannot completely automate specialized heritage translation, they serve as powerful productivity assistants when paired with human expert oversight. Consequently, LSP and translation pedagogy must shift from prohibiting AI tool usage to teaching structured **Machine Translation Post-Editing (MTPE)** methodologies.

Effective integration of AI tools in specialized translation training involves four core steps:

1. **Error Detection Training:** Students are trained to identify AI semantic compression, hypernym replacement, and hallucinated technical terms in AI outputs.

2. **Strategy Selection:** Trainees learn when to accept AI draft outputs and when to override them using human translation strategies—such as applying transliteration with explanatory glossing for cultural realia.

3. **Prompt Engineering for Translation:** Students learn to provide contextual constraints, domain glossaries, and style guidelines directly within LLM prompts to improve initial draft quality.

4. **Termbank Integration:** Educating students to build customized digital termbases to constrain AI translation engines within Computer-Assisted Translation (CAT) environments.

Evaluating Neural Machine Translation and LLMs in specialized architectural translation reveals a clear boundary between technical standardization and cultural adaptation. AI tools demonstrate high precision with standardized structural engineering vocabulary but struggle significantly with culturally-bound architectural realia and historical terminology. To preserve cultural heritage and ensure accuracy in interlingual communication, human expertise remains indispensable. Modern translator training programs must incorporate structured AI post-editing skills, equipping future specialists to leverage AI efficiency while maintaining strict domain-specific precision.

References

1. Bowker, L., & Buitrago-Cirio, J. (2019). *Machine Translation and Global Research: Towards Improved Translation Literacy*. Emerald Publishing Limited.
2. Cabré, M. T. (2000). Elements for a theory of terminology: Towards an alternative paradigm. *Terminology*, 6(1), 35–57.
3. Faber, P. (Ed.). (2012). *A Cognitive Linguistics View of Terminology and Translation*. Berlin: De Gruyter Mouton.
4. Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces*, 6(2), 291–309.
5. Koponen, M. (2016). Is machine translation post-editing worth the effort? *AELJU*, 15, 131–148.
6. O'Brien, S. (2022). Machine translation and translation teaching. In *The Routledge Handbook of Translation and Education* (pp. 212–227). Routledge.
7. Temmerman, R. (2000). *Towards New Ways of Terminology Description: The Sociocognitive Approach*. Amsterdam/Philadelphia: John Benjamins.