

DOI <https://doi.org/10.30525/978-9934-26-383-5-38>

**MACHINE TRANSLATION OF LITERARY TEXT:  
ON THE ISSUE OF CHOOSING TRANSLATION QUALITY  
EVALUATION METRICS**

**МАШИННИЙ ПЕРЕКЛАД ХУДОЖНЬОГО ТЕКСТУ:  
ДО ПИТАННЯ ПРО ВИБІР МЕТРИК ОЦІНЮВАННЯ  
ЯКОСТІ ПЕРЕКЛАДУ**

**Shulzhenko Yu. M.**

*Senior Lecturer  
at the Translation Department  
Kremenchuk Mykhailo Ostrohradskyi  
National University  
Kremenchuk, Poltava region, Ukraine*

**Шульженко Ю. М.**

*старший викладач кафедри  
перекладу  
Кременчуцький національний  
університет  
імені Михайла Остроградського  
м. Кременчук, Полтавська область,  
Україна*

**Artemenko Yu. O.**

*Candidate of Philological Sciences,  
Associate Professor at the Translation  
Department  
Kremenchuk Mykhailo Ostrohradskyi  
National University  
Kremenchuk, Poltava region, Ukraine*

**Артеменко Ю. О.**

*кандидат філологічних наук,  
доцент кафедри перекладу  
Кременчуцький національний  
університет  
імені Михайла Остроградського  
м. Кременчук, Полтавська область,  
Україна*

**Kushnir M. Yu.**

*Master Student at the Translation  
Department  
Kremenchuk Mykhailo Ostrohradskyi  
National University  
Kremenchuk, Poltava region, Ukraine*

**Кушнір М. Ю.**

*магістрант кафедри перекладу  
Кременчуцький національний  
університет  
імені Михайла Остроградського  
м. Кременчук, Полтавська область,  
Україна*

Цифрові технології доволі активно впроваджуються у роботу перекладача, проте, перекладах художньої літератури автоматизований чи машинний переклад застосовується рідше, хоча для оптимізації пошуку або проведення передперекладацької роботи з використанням крос-мовного контенту ці засоби можуть стати в нагоді. Однак зв'язний текст, що породжений машинним перекладачем, усе ще вимагає серйозного доопрацювання і, отже, питання доцільності включення цифрових технологій в роботу сучасного перекладача художніх текстів

та вибору методів і засобів оцінки якості таких перекладів залишаються актуальними.

Отже, постає питання про вибір метрик оцінювання. Наразі існує два основних напрямки оцінки якості машинного перекладу: ручна (людська) оцінка та автоматична оцінка.

Оцінка якості перекладу може бути проведена як автоматично, так і за допомогою людського оцінювання. Кожен підхід має свої переваги та недоліки. Автоматичне оцінювання (метрики BLEU, BLEURT тощо) є швидким та ефективним інструментом, базується на його результати є об'єктивними, оскільки базуються на конкретних алгоритмах та формулах. Проте у випадку з художнім текстом автоматичні метрики можуть не враховувати усі тонкощі та семантичні відтінки перекладу, і це неможливо зробити, використовуючи оцінку з допомогою N-грам (послідовності з n-слів чи знаків) чи порівняння рядків. [4; 2] Уцілому, автоматична метрика оцінює, чи читабельний переклад, чи ні, тому, вважаємо, для оцінки художнього перекладу цього недостатньо. Переклад повинен передавати усі аспекти оригіналу: стиль, настрій, відчуття, інтенції автора, культурні особливості та культурний фон, часові та соціокультурні реалії вдало відтворювати образи, метафори та символи, зберігаючи їх силу, відповідати контексту тощо.

В основі ручного оцінювання лежить оцінка якості перекладу людиною за певними критеріями, у тому числі, й за переліченими вище. Таке оцінювання є витратним, проте дає змогу оцінити семантику, креативність, мету та контекст перекладу. Проте оцінка може бути й суб'єктивною, оскільки рівень розуміння тексту реципієнтом завжди залежить від індивідуальних, отже, суб'єктивних чинників, отже, оцінка перекладу не може бути абсолютно точною [6].

Людська оцінка може відбуватися як подвійне сліпе оцінювання, коли два або більше незалежних експертів оцінюють якість перекладу, не знаючи, яка саме система виконала переклад, або пряме оцінювання, коли експерти надають оцінку перекладу за певною шкалою, виставляючи певну кількість балів за кожен критерій. Найчастіше такі критерії розробляються на основі лінгвостилістичного аналізу тексту. М. Нагао запропонував шкалу оцінки машинного перекладу на основі лінгвостилістичного аналізу, від п'яти до одного бала, що враховує загальний сенс, граматику, слововживання та стиль. [3]

С. Каллісон-Берч зі співавторами вважають, що під час оцінювання тексту можна використовувати рейтингову систему, тобто експерти мають оцінити якість перекладу за низкою критеріїв: метрики плавності (зрозумілості) та адекватності (точності) і скласти рейтинг результатів різних систем автоматичного перекладу, розташувавши їх

за порядком від 1 до 5, де 1 – найгірше і 5 – найкраще [1]. Такі метрики є традиційними шкалами оцінки якості машинного перекладу за параметрами точності і повноти відтворення смислу та правильності і природності цільової мови.

Ще одним способом оцінювання якості перекладу є аналіз помилок, запропонований Д. Віларом зі співавторами [5]. Під аналізом помилок мається на увазі ідентифікація та класифікація помилок у перекладацькому тексті, отриманому під час автоматичного перекладу. Кількісна оцінка обчислюється за емпіричною формулою (1) (чим вищий показник  $E$ , тим вища якість машинного перекладу):

$$E = \frac{N(w)}{N(sem\&lex)*10 + N(synt)*5 + N(morph)*3 + N(morph)*3 + N(cov)*2 + N(style)} \quad (1)$$

де  $E$  – загальна оцінка коректності машинного перекладу;  $N(w)$  – кількість слів у перекладеному тексті;  $N(sem\&lex)$  – кількість семантичних і лексичних помилок;  $N(synt)$  – кількість синтаксичних помилок;  $N(morph)$  – кількість морфологічних помилок;  $N(cov)$  – кількість неперекладених слів;  $N(style)$  – кількість стилістичних помилок.

А. Торал [2] пропонує оцінювати якість машинного перекладу художнього тексту за метрикою креативності, тобто, поєднання інновативності та прийнятності, аналізуючи переклад одиниць з креативним потенціалом: метафор, порівнянь, ідіоматичних виразів, гри слів, історичних та культурних референсів, лексичної різноманітності та неологізмів тощо (загалом 15 критеріїв). Також має бути врахованим рівень відтворення тієї чи іншої креативної одиниці, опущення, помилки, творчі заміни, абстрагування, модифікації та конкретизації (уточнення). Прийнятність та новизна об'єднуються в один бал. Результат множиться на 100, щоб виразити його у відсотках (2):

$$Creativity\ score = \left( \frac{\#CSs}{\#USPs} - \frac{\#error\ points - \#Kudos}{\#number\ of\ words\ in\ ST} \right) * 100 \quad (2)$$

де  $Cs$  (*creative shifts*) – це креативні трансформації,  $UCPs$  (*units of creative potential*) – одиниці з креативним потенціалом, *error points* – кількість помилок, *number of words in ST* – кількість слів у тексті,  $\#Kudos$  – додаткові бали за креативність. Така формула використовувалася у рамках експерименту для експертної оцінки якості машинного перекладу художнього тексту у порівнянні з перекладом, здійсненим людиною. [2] Результати експерименту показали, що на даному етапі машина у більшості випадків не здатна запропонувати

елегантні та креативні рішення перекладацьких проблем, на відміну від людини-перекладача.

Отже, як бачимо, ручне оцінювання можна поєднувати з математичними методами оцінки, виражаючи такі поняття як адекватність чи креативність у числовому відношенні, що дає змогу встановити об'єктивні критерії, дати більш інформативний відгук, зробити результат оцінки перекладу більш консистентним та надійним. Такі експериментальні дослідження можуть слугувати потужним інструментом для покращення та навчання існуючих систем машинного перекладу, дають можливість перекладачам навчатися, аналізуючи конкретні приклади.

### Література:

1. Callison-Burch, C. Evaluation of machine translation. *Proc. Second Workshop on Statistical Machine Translation*. Prague, 2007. P. 136–158.
2. Guerberof-Arenas A. & Toral A. Creativity in translation. Machine translation as a constraint for literary texts. <https://doi.org/10.48550/arXiv.2204.05655>
3. Nagao M. A New Method of N-gram Statistics for Large Number of n and Automatic Extraction of Words and Phrases from Large Text Data of Japanese. *Proceedings of the 15th International Conference on Computational Linguistics (COLING 1994)*. Kyoto, Japan, 1994. P. 611–615.
4. Thai K., Karpinska M., Krishna K., Ray W., Inghilleri M., Wieting John., Iyyer M. Exploring Document-Level Literary Machine Translation with Parallel Paragraphs from World Literature. <https://doi.org/10.48550/arXiv.2210.14250>
5. Vilar, D. Error Analysis of Statistical Machine Translation Output. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, 2006. P. 697–702.
6. White, J.S., T. O'Connell and F. O'Mara. 1994. The ARPA MT Evaluation Methodologies: Evolution, Lessons, and Future Approach. *Proc. of the First Conference of the Association for Machine Translation*. P. 193–205.