

УДК 811.581.11

DOI <https://doi.org/10.24919/2308-4863/43-2-15>**Олександр КОЗОРИЗ,***orcid.org/0000-0002-4502-583X*

кандидат філологічних наук,

асистент кафедри мов і літератур Далекого Сходу та Південно-Східної Азії  
Інституту філології Київського національного університету імені Тараса Шевченка  
(Київ, Україна) *davinci@3g.ua*

## АВТОМАТИЗАЦІЯ ОЦІНКИ ЯКОСТІ ПЕРЕКЛАДУ ТЕКСТІВ ПАРАЛЕЛЬНИХ КОРПУСІВ

Якість лінгвістичного корпусу може визначати низка чинників: *type / token ratio (TTR)*, середня довжина речення *ASL*, *type*, *token*, автоматичний індекс читабельності *ARI*, показники кількості слів у реченні тощо. У статті автором описано продовження власних досліджень у цій сфері, висвітлено інший окремий аспект – перевірку якості текстів перекладів. Було розроблено унікальну методологію оцінки якості перекладів текстів паралельних корпусів на основі метрики *BLEU* в автоматичному режимі. Написано програму мовою Python для обрахунку величини *BLEU* для текстів корпусів за допомогою сервісу Google в автоматичному режимі.

Питання оцінки якості перекладів докладно розглядалися ще з минулого століття, коефіцієнт *BLEU* відомий вже майже 20 років, але поєднання цих двох напрямів дослідження запропоновано автором уперше. Перевагою оцінки *BLEU* є уніфікація та швидкість оцінювання для будь-якого обсягу матеріалів, а також уникнення суб'єктивного людського чинника.

Із сайту *OpenSubtitles* було завантажено 8 різних за мовними напрямками, обсягом і тематичним спрямуванням корпусів. Успішно оцінено якість перекладів для мов 6 різних мовних сімей і 8 мовних пар, таких як: китайська, японська, в'єтнамська, корейська, арабська, індонезійська, англійська, українська. Розроблено методологію роботи з корпусами східних мов: китайської та японської, що не містять проміжків між словами, а також арабської мови із письмом справа наліво.

Визначено граничну межу достовірності перекладів із коефіцієнтом не нижчим 0,26 одиниць (переклади, нижчі за цей показник, мають сумнівну якість, імовірно, є помилковими). Виявлено також, що короткі фрази, які мають варіативні переклади, можуть бути помилково оцінені у 0 балів. Помічено іншу проблему високих коефіцієнтів – 0,9 і більше, такі переклади виглядають занадто ідеальними, ймовірно, через те, що зроблені системами машинного перекладу. Нині в інтернеті вже існує велика кількість таких механічних перекладів.

Оремо було перевірено вади методики шляхом введення зовсім неправильних перекладів, перекладів іншої довжини, повторів частково правильних словосполучень тощо – у всіх випадках коефіцієнт не перетинав меж у 0,26 одиниць, що може вважатися необхідним мінімумом перекладу достатньої якості.

**Ключові слова:** якість лінгвістичного корпусу, паралельний корпус, системи автоматичного перекладу, китайська мова, японська мова, в'єтнамська мова, корейська мова, арабська мова, індонезійська мова.

**Oleksandr KOZORIZ,***orcid.org/0000-0002-4502-583X*

Ph. D. in Philology,

Teaching Assistant at the Department of Languages and Literatures of the Far East and Southeast Asia  
Taras Shevchenko National University of Kyiv  
(Kyiv, Ukraine) *davinci@3g.ua*

## AUTOMATIZATION OF TRANSLATION QUALITY ASSESSMENT OF TEXTS OF PARALLEL CORPORA

A number of factors can determine the quality of the linguistic corpus: *type / token ratio (TTR)*, average sentence length *ASL*, *type*, *token*, automatic readability index (*ARI*), indicators of the number of words in a sentence, etc. In the article the author continued his own research in this area, highlighted another separate aspect - checking the quality of translated texts. The unique methodology for assessing the quality of translations of texts in parallel corpora based on *BLEU* metrics in automatic mode has been developed. The program has been written in Python to calculate the *BLEU* value for corpus texts using the Google service in automatic mode.

The issues of translation quality assessment have been considered in detail since the last century, the *BLEU* coefficient has been known for almost 20 years, but the author has proposed the combination of these two areas of research for the first time today. The advantage of *BLEU* evaluation is the unification and speed of evaluation for any volume of materials, as well as the avoidance of subjective human factor.

*From the OpenSubtitles site were downloaded eight different in language areas, volume and thematic direction of the corpora. The quality of translations for the languages of six different language families and 8 language pairs has been successfully assessed: Chinese, Japanese, Vietnamese, Korean, Arabic, Indonesian, English and Ukrainian. The methodology for working with the corpora of Oriental languages has been developed: Chinese and Japanese, which do not contain spaces between words, as well as Arabic with writing from right to left.*

*The limit of reliability of translations with a coefficient not lower than 0.26 units has been determined (translations below this indicator have doubtful quality, probably erroneous). It was also found that short phrases that have variable translations could be mistaken for 0 points. Another problem with high coefficients (0.9 and more) has been noticed, such translations look too ideal, probably because they are made by machine translation systems. There are already a large number of such mechanical translations online.*

*Some defects of the methodology were checked separately by entering completely incorrect translations, translations of different lengths, repetitions of partially correct phrases, etc. In all cases, the coefficient did not exceed 0.26 units, which can be considered a necessary minimum of sufficient quality translation.*

**Key words:** translation quality assessment, parallel corpus, automatic translation systems, Chinese, Japanese, Vietnamese, Korean, Arabic, Indonesian.

**Виклад основного матеріалу.** Нейронний машинний переклад (NMT), розроблений компанією Google (GNMT) та представлений у листопаді 2016 року, відкрив нову еру в галузі машинного перекладу; використовує штучну нейронну мережу для підвищення швидкості та якості перекладу, застосовує метод машинного перекладу, що базується на прикладах. Така система «навчається» перекладати на мільйонах прикладів. Якість перекладів такої системи переважно залежить від якості попередньо створених паралельних корпусів. Під паралельними корпусами розуміємо саме вирівняні (aligned) паралельні тексти, де існує зв'язок між реченнями як одиницями перекладу.

У мережі з'явилася величезна кількість паралельних корпусів на основі різноманітних матеріалів (книжки, новини, блоги, сайти тощо), але не всі із цих корпусів мають високу якість: поганий переклад, невіривняний переклад, іншомовні вкраплення тощо. Саме тому створення якісних різноматематичних паралельних корпусів набуває вагомого значення. Одним із аспектів, що визначають якість корпусу, виступає точність і достовірність його перекладів. Тому виникла ідея автоматизувати оцінку якості перекладів текстів корпусів, оскільки це один із критеріїв якості всього корпусу.

Задля досягнення мети дослідження виконаємо такі завдання: визначимо сутність якості перекладу; запропонуємо методику та створимо програму автоматичної оцінки якості перекладу; перевіримо ефективність роботи програми на різних мовних напрямках.

Для характеристики результатів перекладацького процесу В. Н. Комісаров використовує такі терміни, як адекватний та еквівалентний переклад, а також точний, буквальный і вільний переклад (Комиссаров, 1990). Усі вони й досі застосовуються, щоб сформулювати вимоги до якості перекладу, виконаного людиною.

Оцінка перекладу найчастіше проводиться за парадигмою адекватності та плавності, де під адекватністю (точністю) розуміють ступінь, до якого переклад фіксує значення оригіналу, а під плавністю (зрозумілість) – ступінь, до якого вихідний текст відповідає критеріям цільової мови.

Стівен Доєрті у своїй статті «Питання оцінки якості людського та автоматичного перекладу» розглядає проблеми, що виникають при оцінюванні перекладу людиною, зауважує, що надійність та ефективність оцінки залежить від якості написаних інструкцій і вказівок. Також згадує про високий ступінь суб'єктивності, коли існує більше одного оцінювача. Для вимірювання надійності між оцінювачами пропонується використовувати коефіцієнт Каппа Коєна (Doherty, 2017).

У контексті глобалізації та інформатизації суспільства у зв'язку зі зростанням кількості перекладеної інформації виникла потреба створення автоматичної метрики оцінки машинного перекладу, яка оцінює якість перекладу шляхом порівняння вихідних даних системи *машинного перекладу* з набором «еталонних» перекладів, здійснених людиною.

**BLEU** – один із перших показників, який високо корелює з оцінками якості, здійснених людиною, на рівні корпусу. Нині є одним із найпопулярніших у цій галузі. Основна ідея метрики полягає в тому, що «чим ближче *машинний переклад* (МП) до професійного людського перекладу, тим він кращий». Метрика обчислює бали для окремих сегментів, зазвичай речень, а потім усереднює ці бали по всьому корпусу для остаточної оцінки. Вихідним показником завжди є число між 0 і 1. Це значення вказує, наскільки перекладений текст за допомогою МП відповідає еталонному перекладу. BLEU використовує модифіковану форму точності, щоб запобігти поліпшенню оцінки за рахунок невиправданого повтору окремих слів.

Кішор Папінені у своїй статті «BLEU: метод автоматичного оцінювання машинного

перекладу» детально описує усі особливості методу, експериментальним шляхом доводить, що співвідношення BLEU високо корелює з людським судженням, щодо перекладу на англійську мову з чотирьох цілком різних мов (арабської, китайської, французької, іспанської), що представляють 3 різні мовні сім'ї. Перетинання цієї прірви для китайсько-англійського перекладу є значним викликом сьогодення (Parineni, 2002).

Крім BLEU також застосовуються інші метрики: NIST, Word error rate (WER) – коефіцієнт помилок у словах, Position-independent word error rate (PER), METEOR, LEPOR тощо. Загальний огляд метрик автоматичного оцінювання якості машинного перекладу знаходимо у статті В. Яковини та В. Масюкевич (Яковина, 2013).

Розглянемо процес роботи метрики BLEU на практиці. Один із варіантів можна побачити, застосувавши Інтерактивний оцінювач метрики BLEU (Interactive BLEU score evaluator), який рахує оцінку BLEU для перекладених файлів, один або кілька, порівнюючи їх із еталонними перекладами. Має зручний інтерфейс, де можна продивлятися переклади у зіставленні з оригіналом по реченню, із підсвіткою відмінностей, або ж завантажити Excel документ із результатами перевірки. Інший варіант. Напишемо невеличку програму мовою Python, див. рис. 1.

У сучасному світі завдяки сервісу Google translate будь-яка людина може певною мірою стати перекладачем будь-якої мови. Так само будь-який філолог за допомогою сервісу Google-пошук певною мірою

може стати програмістом. Елегантний і простий синтаксис мови Python робить її ідеальною для написання скриптів і швидкої розробки прикладних програм у багатьох галузях науки, у тому числі лінгвістики.

Запропонований скрипт використовує модуль Natural Language Toolkit (NLTK), спеціально розроблений для аналізу природних мов. Він порівнює два файли з «еталонним» і «машинним» перекладами й автоматично підраховує індекс BLEU для визначення якості перекладу кожного рядка корпусу. Як машинний переклад можна використовувати сервіс Google translate. Запропонована програма не працює для коротких речень довжиною менше чотирьох слів. Для коротких речень можна запропонувати такий рядок коду: `BLEUScore = nltk.translate.bleu_score.sentence_bleu([reference], hypothesis, weights = (0.5, 0.5))`.

Також можна порахувати середній BLEU для цілого корпусу і визначити його якість. Напишемо ще одну програму, яка працює з єдиним файлом-корпусом, де паралельний текст розділено знаком «>». Тобто рядок корпусу виглядає таким чином: «但是 友谊的确深厚。 > І це таки було дружбою». Програма автоматично завантажує оригінал (ліва частина), перекладає через Google translate і порівнює його з перекладом у корпусі (права частина), що ми перевіряємо; також рахує загальний BLEU корпусу як середній арифметичний усіх рядків. У програмі використано модуль mtranslate (Google Translate API), що сьогодні є одним із небагатьох безкоштовних варіантів роботи із сервісом Google translate, див. рис. 2.

```
import nltk
x = 1
with open('text1.txt', 'rt', encoding='utf-8') as f1:
    with open('text2.txt', 'rt', encoding='utf-8') as f2:
        with open('text3.txt', 'wt', encoding='utf-8') as f3:
            while True:
                hypothesis = f1.readline()
                reference = f2.readline()
```

```
BLEUScore =
nltk.translate.bleu_score.sentence_bleu([reference],
hypothesis)
line = str(x) + ' ' + str("%.4f" % BLEUScore) +
'\t' + hypothesis[0:-1] + '\t' + reference
if not hypothesis:
    break
print(line, file=f3, sep=",", end="")
x += 1
```

Рис. 1. Оцінювач метрики BLEU мовою Python для двох файлів

```
import nltk
from mtranslate import translate
x = 1
corpus_total = 0
with open('test.txt', 'rt', encoding='utf-8') as f1:
    with open('text.txt', 'wt', encoding='utf-8') as f3:
        while True:
            s = f1.readline()
            n = s.find(" > ")
            hypothesis = s [n+3:-1]
            reference = translate(s [0:n], "uk")
```

```
BLEUScore =
nltk.translate.bleu_score.sentence_bleu([reference],
hypothesis)
line = str(x) + '\t' + s [0:n] + '\t' +
str("%.4f" % BLEUScore) + '\t' + hypothesis + '\t' +
reference
corpus_total += BLEUScore
if not hypothesis:
    print("%.4f" %
(corpus_total/(x-1)), file=f3)
    break
print(line, file=f3, sep=",", end='\n')
x += 1
```

Рис. 2. Оцінювач метрики BLEU мовою Python для монофайлу

Ми застосуємо метод оцінювання BLEU для перевірки якості корпусу. Пропонуємо переглянути табл. 1, яка демонструє результати роботи програми та результати оцінки BLEU для англійської мови. Столпчики «оригінал» та «еталонний переклад» взято нами з корпусу, колонку «переклад» виконано Google translate, BLEU – пораховано нашою програмою, переклади рядків 21–25 створені штучно для перевірки алгоритму оцінки BLEU, не враховані у середній показник. Середній BLEU для рядків від 1 до 20 – 0,6 одиниць. Загальну якість перекладів корпусу можна оцінити як високу. Результати роботи нашої програми з оцінювання BLEU дещо різняться з

результатами Інтерактивного оцінювача, але загалом корелюють.

Як бачимо, рядок 1 – повний збіг, оцінюється як одиниця. Чим ближчий лексичний і структурний збіг, тим вища оцінка, результати варіюються від 0,37 до 1. Менші результати дають рядки 19–20, тут у нас синонімічні інваріанти перекладу, які також є правильними, втім, мають досить низькі показники. Рядки № 21–22: тут ми проекспериментували з кількістю схожих слів, коефіцієнт BLEU варіюється від 0,48 до 0,78. У рядках № 23–24 ми задали завідомо неправильні переклади, коефіцієнт BLEU варіюється від 0,09 до 0,22; останній показник рядка № 23

Таблиця 1

**Ілюстративна таблиця підрахунку коефіцієнта BLEU для англійської мови**

№	Оригінал	BLEU	Еталонний переклад	Машинний переклад Google
1	非洲是一块非常大的大陆。	1,0000	Africa is a very large continent.	Africa is a very large continent.
2	你是哪里人呢？	0,9457	Where are you from?	Where are you from?
3	市政大厅前有一群人。	0,7947	There is a crowd of people in front of the town hall.	There is a group of people in front of the city hall.
4	她坐在窗旁。	0,5086	She sat at the window.	She is sitting by the window.
5	她从马上摔了下来。	0,8235	She fell off her horse.	She fell off the horse.
6	这条街不准停车。	0,5000	You can't park in this street.	No parking is allowed on this street.
7	咖啡和茶使我感到兴奋。	0,6943	Coffee and tea stimulate me.	Coffee and tea excite me.
8	山那边有一个村庄。	0,6124	There is a village beyond the hill.	There is a village over the mountain.
9	我希望有更多的时间。	0,5330	I wish I had more time.	I hope to have more time.
10	目前，他正在度假。	0,2820	At present, he is on holiday.	He is currently on vacation.
11	你想看这个电视节目吗？	0,7382	Do you want to watch this TV programme?	Do you want to watch this TV show?
12	除非他亲口对我说，否则我决不相信。	0,6155	I will not believe it unless and until he tells me so.	Unless he tells me personally, I will never believe it.
13	法国客人就住在那边的旅馆。	0,6096	The French are staying in the hotel just over there.	French guests live in the hotel over there.
14	我尊重他这位作家，也尊重他这个人。	0,5693	I respect him as a writer and as a man.	I respect him as a writer, and I also respect him as a man.
15	孩子的新毛衣上勾了一个洞。	0,5470	The child has picked a hole in his new jumper.	There was a hole in the child's new sweater.
16	二十世纪的经验和教训	0,9009	The experience and lessons of the twentieth century	Experience and lessons of the twentieth century
17	但他们也会带来欢乐。	0,4047	But they will also give delight.	But they also bring joy.
18	你也不错。	0,3757	You're looking good yourself.	You're good too.
19	我累死了。	0,2819	I am worn out.	I am exhausted.
20	我就要份牛排。	0,2826	I will take the beefsteak.	I want a steak.
21	我就要份牛排。	0,4843	I will take the beefsteak.	<i>I want a beefsteak.</i>
22	我就要份牛排。	0,7801	I will take the beefsteak.	<i>I will take the steak.</i>
23	我就要份牛排。	0,2221	I will take the beefsteak.	<i>She fell off the horse.</i>
24	我就要份牛排。	0,0903	I will take the beefsteak.	<i>The French are staying in the hotel just over there.</i>
25	我就要份牛排。	0,2638	I will take the beefsteak.	I take I take I take.



відносно високий, він виникає через схожу кількість слів еталону і перекладу, але він все ще менший за правильні переклади.

У рядку № 25 ми спробували заплутати програму, ввели повторювану двослівну послідовність із приблизно правильною загальною кількістю слів у реченні, як результат, отримали досить високий коефіцієнт BLEU – 0,26. Це дійсно хіба алгоритму підрахунку, але у реальному житті таких перекладів не буває.

Так само отримуємо оцінювання для китайської мови, див. табл. 2. Єдина відмінність полягає у тому, що треба додати проміжки між ієрогліфами до початку роботи програми, а після

закінчення – їх видалити. Можемо зазначити, що загальна оцінка для китайської мови дещо вища, ніж для англійської, ймовірно, через те, що початково коефіцієнт було розроблено для мов, що мають літери. Особливо добре це видно на прикладі № 9 і № 10, де переклад суттєво відрізняється від еталона за структурою, а оцінка лишається досить високою. У прикладі № 11 ми так само навмисне дали завідомо неправильний варіант із тотожною кількістю знаків і отримали коефіцієнт 0,16. Середнє 0,71 без врахування останнього рядка.

Аби бути об'єктивними щодо якості роботи скрипту, ми провели оцінювання коефіцієнта

Таблиця 2

Ілюстративна таблиця підрахунку коефіцієнта BLEU для китайської мови

№	Оригінал	BLEU	Еталонний переклад	Машинний переклад Google
1	Do you want to watch this TV programme?	1,0000	你想看这个电视节目吗?	你想看这个电视节目吗?
2	Africa is a very large continent.	0,8895	非洲是一块非常大的大陆。	非洲是一个非常大的大陆。
3	I wish I had more time.	0,8709	我希望有更多的时间。	我希望我有更多的时间。
4	She fell off her horse.	0,8470	她从马上摔了下来。	她从马上摔下来了。
5	There is a village beyond the hill.	0,7509	山那边有一个村庄。	山上有一个村庄。
6	There is a crowd of people in front of the town hall.	0,6179	市政大厅前有一群人。	市政厅前有很多人。
7	The French are staying in the hotel just over there.	0,6023	法国客人就住在那边的旅馆。	法国人住在那边的酒店里。
8	You can't park in this street.	0,5385	这条街不准停车。	你不能在这条街上停车。
9	I will not believe it unless and until he tells me so.	0,5291	除非他亲口对我说, 否则我决不相信。	除非并且直到他告诉我, 否则我不会相信。
10	Coffee and tea stimulate me.	0,4700	咖啡和茶使我感到兴奋。	咖啡和茶刺激了我。
11	The French are staying in the hotel just over there.	0,1608	法国客人就住在那边的旅馆。	法国法国法国法国法国法国。

Таблиця 3

Ілюстративна таблиця підрахунку коефіцієнта BLEU для в'єтнамської мови

№	Оригінал	BLEU	Еталонний переклад	Машинний переклад Google
1	It never happened.	0,2529	Hãy xem như chuyện đó chưa từng xảy ra.	Nó không bao giờ xảy ra.
2	You know who he is, right?	0,5365	Ông biết hắn là ai, đúng không?	Bạn biết anh ta là ai, phải không?
3	And people pay for that.	0,6920	Và người ta trả tiền cho việc đó.	Và mọi người trả tiền cho điều đó.
4	That's it. That's it, you're leaving.	0,1479	Thế thôi, về nhà đi.	Đó là nó. Vậy đó, bạn đang đi.
5	I'll tell you what you wanna know.	0,2069	Tôi sẽ nói cái anh muốn	Tôi sẽ cho bạn biết những gì bạn muốn biết.
6	Yeah, we're gonna take it slow.	0,3427	Chúng ta sẽ lái thật chậm.	Phải, chúng tôi sẽ làm nó chậm lại.
7	That is life. That's what it is.	0,3472	Đó là cuộc sống, thế đấy.	Đó là cuộc sống. Đó là những gì nó được.
8	Tonight, the future begins.	0,8710	Đêm nay, tương lai sẽ bắt đầu.	Đêm nay, tương lai bắt đầu.
9	What would you do with him?	0,5956	Cô định làm gì với anh ấy chứ?	Bạn sẽ làm gì với anh ấy?
10	Come on, you have to go.	0,2573	Đi nào. Mọi người phải đi ngay.	Nào, bạn phải đi.

BLEU для інших популярних для перекладу мов із різних мовних сімей. Для в'єтнамської мови (див. табл. 3.) із сайту Opus (OPUS) було завантажено паралельний корпус субтитрів і вибрано довільно 10 рядків. Як видно з оцінювання (див. табл. 3), переклади № 1, 4, 5 і № 10 мають досить низький BLEU – менше 0,26, а отже, мають сумнівну якість для корпусу. Загалом переклад субтитрів може містити певні суб'єктивні неточності, спричинені їхньою специфікою: розмовним стилем, акустичною довжиною репліки тощо.

Так само було проведено оцінювання коефіцієнта BLEU для української мови. Із сайту Opus (OPUS) було завантажено паралельний корпус субтитрів і вибрано довільно 10 рядків. Усі переклади більшою мірою правильно передають зміст китайського оригіналу (див. табл. 4), однак рядок № 9 містить альтернативний лексико-синтаксичний варіант перекладу, і машина, на жаль, оцінила BLEU у нуль балів. Також переклади № 7 і № 8 мають низькі показники, але тут скоріше

хиба «занадто художнього», вільного перекладу української частини, Google переклав ближче до оригіналу, тому оцінку BLEU можна вважати правильною.

Оцінювання коефіцієнта BLEU для японської мови дещо ускладнено відсутністю проміжків між словами, тому довелося скористатися сторонньою програмою SegmentAnt (SegmentAnt), яка ділить китайські та японські речення на слова; також написана на Python і за потреби може бути інтегрована в один програмний код. Із сайту Opus (OPUS) було завантажено паралельний корпус субтитрів і вибрано довільно 10 рядків. Після додавання проміжків переклад японською мовою виглядає таким чином (див. табл. 5). Усі переклади більшою мірою правильно передають зміст українського оригіналу, лише рядок № 5 має нульовий BLEU, хоча переклад правильний, але еталон занадто метафоричний; так само низький BLEU має рядок № 9, оскільки еталонний переклад занадто спрощений.

Таблиця 4

**Ілюстративна таблиця підрахунку коефіцієнта BLEU для української мови**

№	Оригінал	BLEU	Еталонний переклад	Машинний переклад Google
1	世界上所不知道的是，两个独裁者的贪欲远远不止波兰。	0,4038	Ось чого світ не знав, що обидва диктатори узгодили більше ніж Польщу.	Світ не знає, що жадібність двох диктаторів виходить далеко за межі Польщі.
2	在克里姆林宫签署一份秘密协议显示 在战争开始前的一个星期，	0,5753	У таємному протоколі, підписаному у Кремлі за тиждень до початку війни	Підписання секретної угоди в Кремлі показало, що за тиждень до початку війни
3	希特勒和斯大林就达成了 分割欧洲的协议。	0,4951	Гітлер та Сталін узгодили поділ Європи.	Гітлер і Сталін досягли згоди розділити Європу.
4	该协议的签署称为 希特勒和斯大林侵略 一系列欧洲国家的基础。	0,4245	Цим таємним протоколом Гітлер дав зелене світло Сталіну на окупацію певної кількості європейських країн.	Підписання угоди було названо основою агресії Гітлера та Сталіна проти низки європейських країн.
5	名单上首先是芬兰。	0,4710	Першою за переліком була Фінляндія.	Першою у списку є Фінляндія.
6	俄罗斯轰炸机再次袭击赫尔辛基。	0,7143	Знову російські бомбардувальники шугають над Гельсінкі.	Російські бомбардувальники знову атакували Гельсінкі.
7	此刻，城市正在被摧毁。	0,1973	Знову розкішне місто затряслося від війни та зруйнування.	У цей момент місто руйнується.
8	我们看到一栋栋的建筑物 被炸弹彻底摧毁！	0,2041	Без затримки і б'ючи швидше, як ви бачите, будівлі падають від влучених снарядів.	Ми бачили будівлю за будівлею, повністю зруйновану бомбами!
9	这是一场灾难。	0,0000	То було лихо.	Це катастрофа.
10	我的父亲失去了4个兄弟。	0,3884	Мій батько втратив чотирьох із п'яти його братів у тій війні.	Мій батько втратив 4 братів.

Далі було проведено оцінювання для корейської мови. Як бачимо з результатів (див. табл. 6), хоча між словами є проміжки та не треба додатково доопрацьовувати матеріал, загальна оцінка коефіцієнта BLEU досить низька. Тут основних причин кілька: велика кількість граматичних форм корейської мови – суфіксів і закінчень, що

впливає на якість перекладу від Google; через віддаленість мов присутня суттєва граматична варіативність у перекладі. Ці фактори зменшують ефективність автоматичного оцінювання для цього мовленнєвого напрямку.

Для оцінювання арабської мови із сайту Opus (OPUS) було завантажено англо-арабський пара-

Таблиця 5

Ілюстративна таблиця підрахунку коефіцієнта BLEU для японської мови

№	Оригінал	BLEU	Еталонний переклад	Машинний переклад Google
1	«Хто контролює минуле – той контролює майбутнє»	0,7062	過去を支配する者は未来まで支配する	「過去を支配する者が未来を支配する」
2	Будівельники нашого світу, борються захищають, спливають кров'ю помирають.	0,3158	我々の世界の建設者である争い戦い血を流し命を捨てて	私たちの世界の建設者たちは、保護のために戦い、血を吐き、死にかけています。
3	На вулицях наших міст і на полі бою борються проти спотворення наших мрій та надій.	0,2988	わが都市の街路で遥かなる戦場で我々の希望と夢を砕く敵と戦って	私たちの街の通りや戦場で、彼らは私たちの夢や希望の歪みと戦っています。
4	Хто вони?	0,4367	彼らは何者だ?	彼らは誰なの?
5	Їсти хочеш?	0,0000	腹は減ってるか?	あなたは食べたいですか
6	Сідайте, будь ласка.	0,2320	どうぞ、座って	座ってください。
7	Я зараз хочу побути сама.	0,4946	今は一人になりたいの	私は今一人になりたいです。
8	Ти крадеш моє авто.	0,3642	俺の車を盗んでるぜ	あなたは私の車を盗んでいます。
9	Відвезу їх до поліцейського відділку.	0,1382	警察署に連行する	彼らを警察署に連れて行きます。
10	Ти не хочеш цього робити, запевняю.	0,3044	あなたはこんな事したくない約束する	あなたはそれをしたくない、私はあなたを保証します。

Таблиця 6

Ілюстративна таблиця підрахунку коефіцієнта BLEU для корейської мови

№	Оригінал	BLEU	Еталонний переклад	Машинний переклад Google
1	Це фільм не про піратство.	0,3354	이 영화는 해적행위에 대한 이야기가 아닙니다.	이것은 불법 복제에 관한 영화가 아닙니다.
2	Ці ліжка просто жахливі.	0,1880	이 침대만 있으면 불안해요	이 침대는 끔찍합니다.
3	Думаєш він довго протримається в нафтовій індустрії?	0,1095	석유시추하는데 오래 있을거 같아요?	석유 산업에서 오래 지속될 것이라고 생각하십니까?
4	Мамо, я хочу перебраться додому.	0,3594	엄마, 집에 가고 싶어	엄마, 집으로 이사하고 싶어요.
5	Я не хочу вмирати!	0,5378	죽고 싶지 않아!	나는 죽고 싶지 않다!
6	Напиши, якщо будуть питання.	0,2289	궁금한 거 있으면 메일 보내요	질문이 있으면 작성하십시오.
7	це чудове запитання.	0,2896	좋은 질문이에요	이것은 좋은 질문입니다.
8	Він дав нам другий шанс.	0,3876	그분께서 두 번째 기회를 주셨어요	그는 우리에게 두 번째 기회를 주었습니다.
9	Така честь з вами познайомитися.	0,0000	만나뵈게 돼서 너무 영광이에요	만나서 영광입니다.
10	Вона спеціаліст з теоретичної фізики елементарних часток.	0,1863	소립자 물리학 전공이죠	그녀는 소립자 이론 물리학의 전문가입니다.

лельний корпус англійської телепередачі TED Talks. Середня оцінка BLEU становить 0,6534 одиниці (див. табл. 7). Такі відносно високі показники виникли через однозначність смислу текстів оригіналу, а також низьку варіативність перекладів. Текст арабського «еталону» було перекладено з англійської, оскільки зазначено що англійське. Загалом якісь переклади корпусу можна оцінити як достойну, оскільки коефіцієнт усіх речень розташувався у межах від 0,35–0,86 одиниць.

Для оцінювання індонезійської мови із сайту Opus (OPUS) було завантажено достатньо великий паралельний корпус (більше 9 млн пар речень), вибіркоче дослідження якого показало високу середню оцінку BLEU – 0,81 одиниць (див. табл. 8), але тут теж виникає сумнів щодо якості корпусу, оскільки, можливо, корпус було перекладено саме за допомогою сервісу Google.

Оскільки у дослідницьких умовах увесь корпус поділено на окремі речення, які сегментовані крапками або іншими розділовими знаками, контекст є важливим аспектом, котрий не враховується в оцінці BLEU. Також абсолютно не враховується варіативність перекладів, що може призводити до заниження оцінювання у деяких випадках. Необхідний «еталонний» переклад, заздалегідь підготовлений людиною або систе-

мою автоматичного перекладу, також може містити певний суб'єктивний чинник.

Натомість автоматичне оцінювання має низьку перевагу: уніфіковане та швидке оцінювання для будь-якого обсягу матеріалів; уникнення суб'єктивних чинників оцінювання – переклад буде завжди мати однаковий бал за однакових вхідних даних; оцінка BLEU високо корелює з людською оцінкою, уникаючи суб'єктивності останньої.

Загалом в автоматичному режимі успішно було оцінено мови 6 різних мовних сімей і 8 мовних пар (див. табл. 9). У всіх випадках коефіцієнт показав свою правомірність. Якість оцінки BLEU залежить від довжини речення, кількості варіантів альтернативних перекладів, якості перекладу сервісу Google, адекватності еталонного перекладу, віддаленості мов мовної пари та інших чинників. Тому цифри, зазначені у таблиці щодо середнього, мінімального та максимального BLEU різних мов, мають лише довідковий характер і не можуть порівнюватися між собою. Переклади з коефіцієнтом нижчим за 0,26 мають сумнівну якість, хоча у поодиноких випадках на коротких реченнях через варіативність і метафоричність перекладу правильний переклад може бути помилково оцінений як «нуль».

Таблиця 7

Ілюстративна таблиця підрахунку коефіцієнта BLEU для арабської мови

№	Оригінал	BLEU	Еталонний переклад	Машинний переклад Google
1	So, let me show you a picture of one form of diversity.	0.8462	تروص مكيريأ ينوع ديل اتلاب عونتلأ نم دحاو لكششل	دحاو لكششل تروص مكيريأ ينوع دحاو عونتلأ نم
2	That suit would cost me about a million and a half dollars.	0.5531	براقى ام قزبلا فلكتسو رالود فصنو نوي لم	يل اوح ينفلكتست قلدبلا كلت رالود فصنو نوي لم
3	Of course a lot of people will ignore this advice.	0.7248	نم ريثكلا لالحا عييبطبو عحيصنلا هذه للاحتي سانلا	نم ريثكلا للاحتي عييبطاب عحيصنلا هذه سانلا
4	Within three weeks it will be everywhere in the world.	0.8494	نوكيس عيباسا قثالث لالحا مل اعلا يف نالكم لك يف	نوكيس عيباسا قثالث نوضع يف مل اعلا يف نالكم لك يف
5	You know the old adage «laughter is the best medicine?»	0.6011	ميدقلا لثمل انوفرعت له «؟جالع لصفأ وه كحضلا»	ميدقلا روثملا لوقلا فرعت تنأ «؟اواد لصفأ وه كحضلا»
6	So our camera system can't actually see it directly.	0.5827	نم اريم الكلا ماظن نكمتي ال اذل رشابم لكشش متيور	اريم الكلا ماظن نكمتي ال لذل قرشابم امتيور انب صاخلا
7	Now, it doesn't require a natural disaster for this to work.	0.4981	قثراك بلطنت ال ،نال لمتعلا اذه لعييبط	قثراك رمالا بلطنتي ال ،نال اذه حجني يتح لعييبط
8	What led to their global dominance for so many years?	0.8255	امتنيه لىلأ يذأ يذلا ام ؟قديع تاو نسل لىل اعلا	مل اعلا لىل عمتنيه لىلأ يذأ يذلا ام ؟قديع تاو نسل
9	So let me go forward to the new science of happiness.	0.8684	لىلأ لقتننأ لىل اوح مسا اذل ديدجلا قداعسلا ملع	ملع لىلأ مدقتأ نأ لىل اوح مسا اذل ديدجلا قداعسلا
10	It was an absolute precondition for our being able to speak.	0.3543	قبسم يس اساطرش لذل و مال الكلا لىل اعلا نتردقل	اقلطم اقبسم اطرش نالك دقل مال الكلا نم ان نكمتل
11	People started to ask me, «Where is this guy now?»	0.4844	لىل اعلا لىل اوسلا حرطب سانلا ادب ؟نال صخشلا اذه وه نيا	اذه نيا ، ين نول أسى سانلا ادب «؟نال لجرلا



Таблиця 8

## Ілюстративна таблиця підрахунку коефіцієнта BLEU для індонезійської мови

№	Оригінал	BLEU	Еталонний переклад	Машинний переклад Google
1	I'm sorry I ever doubted you.	1.0000	Maaf saya pernah meragukan Anda.	Maaf saya pernah meragukan Anda.
2	Ah, who could blame you, I'm very unlikable.	0.9109	Ah, siapa yang bisa menyalahkanmu, Saya sangat tidak disukai.	Ah, siapa yang bisa menyalahkanmu, aku sangat tidak disukai.
3	Velma, I need you at the stage.	0.6382	Velma, aku butuh kamu di panggung.	Velma, aku membutuhkanmu di atas panggung.
4	Yeah, you guys might not be the loosest cats around, but you never gave up on the party and that's the true spirit of the Bash.	0.7130	Ya, kalian mungkin tidak Kucing longgar di sekitar, Tapi Anda tidak pernah menyerah pada pesta itu Dan itulah semangat sejati Bash.	Ya, kalian mungkin bukan kucing yang paling longgar, tetapi Anda tidak pernah menyerah pada pesta dan itulah semangat Bash yang sebenarnya.
5	It's the nursery.	1.0000	Ini kamar bayi.	Ini kamar bayi.
6	We're retiring from the party animal life.	0.6933	Kami pensiun Dari pesta kehidupan binatang	Kami pensiun dari kehidupan hewan pesta.
7	Fun was fun, but deep down our hearts are as square as they get.	0.6718	Menyenangkan memang menyenangkan, tapi jauh di lubuk hatiku Hati adalah sebagai persegi seperti yang mereka dapatkan.	Kesenangan itu menyenangkan, tetapi jauh di lubuk hati kami, hati kami sama seperti yang mereka dapatkan.
8	But I can think of three worthy replacements who never gave up on this case.	0.8090	Tapi aku bisa berpikir Dari tiga pengganti yang layak Yang tidak pernah menyerah dalam kasus ini	Tapi saya bisa memikirkan tiga pengganti yang layak yang tidak pernah menyerah dalam kasus ini.
9	Oh, kids, how could you?	0.8277	Oh, anak-anak, bagaimana bisa kamu?	Oh, anak-anak, bagaimana bisa?
10	What are your orders, my captains?	0.8798	Apa perintah anda, kapten saya	Apa perintah Anda, kapten saya?

Таблиця 9

## Зведена таблиця характеристик коефіцієнта BLEU

Мовна сім'я	мова	Середній BLEU	Мінімальний BLEU	Максимальний BLEU
індоєвропейська	1. англійська	0,6	0,28	1
Сино-тибетська	2. китайська	0,71	0,47	1
Австроазійські мови	3. в'єтнамська	0,425	0,14	0,87
індоєвропейська	4. українська	0,39	0	0,71
Ізольована	5. японська	0,33	0	0,70
Ізольована	6. корейська	0,26	0	0,54
Семіто-хамітська	7. арабська	0,65	0,35	0,86
Малайсько-полінезійська	8. індонезійська	0,84	0,67	1

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Комиссаров В. Н. Теория перевода (лингвистические аспекты) : учебник. Москва : Высш. шк., 1990. 253 с.
2. Яковина В., Масюкевич В. Огляд та аналіз метрик оцінювання якості машинного перекладу. *Вісник Національного університету «Львівська політехніка». Комп'ютерні науки та інформаційні технології*. 2013. № 771. С. 101–107.
3. Google Translate API : веб-сайт. URL: <https://github.com/mouuff/mtranslate> (accessed 1 June 2021).
4. Interactive BLEU score evaluator : веб-сайт. URL: <https://www.letsmt.eu/Bleu.aspx> (accessed 1 June 2021).
5. Kishore Papineni, Salim Roukos, 2002, BLEU: a Method for Automatic Evaluation of Machine Translation : веб-сайт. URL: [https://www.researchgate.net/publication/2588204\\_BLEU\\_a\\_Method\\_for\\_Automatic\\_Evaluation\\_of\\_Machine\\_Translation](https://www.researchgate.net/publication/2588204_BLEU_a_Method_for_Automatic_Evaluation_of_Machine_Translation) (accessed 1 June 2021).
6. NLTK : веб-сайт. URL: <http://www.nltk.org/> (accessed 1 June 2021).
7. SegmentAnt : веб-сайт. URL: <https://www.laurenceanthony.net/software/segmentant/> (accessed 1 June 2021).
8. Stephen Doherty, 2017, Issues in human and automatic translation quality assessment : веб-сайт. URL: [https://www.researchgate.net/publication/314261771\\_Issues\\_in\\_human\\_and\\_automatic\\_translation\\_quality\\_assessment](https://www.researchgate.net/publication/314261771_Issues_in_human_and_automatic_translation_quality_assessment) (accessed 1 June 2021).
9. OPUS : веб-сайт. URL: <https://opus.nlpl.eu/OpenSubtitles.php> (accessed 1 June 2021).

#### REFERENCES

1. Google Translate API. URL: <https://github.com/mouuff/mtranslate> (accessed 1 June 2021).
2. Interactive BLEU score evaluator. URL: <https://www.letsmt.eu/Bleu.aspx> (accessed 1 June 2021).
3. Kishore Papineni, Salim Roukos, 2002, BLEU: a Method for Automatic Evaluation of Machine Translation. URL: [https://www.researchgate.net/publication/2588204\\_BLEU\\_a\\_Method\\_for\\_Automatic\\_Evaluation\\_of\\_Machine\\_Translation](https://www.researchgate.net/publication/2588204_BLEU_a_Method_for_Automatic_Evaluation_of_Machine_Translation) (accessed 1 June 2021).
4. Komissarov V. N. Teoriya perevoda (lingvisticheskie aspekty) [Translation Theory (Linguistic Aspects)] : uchebnyk. Moskva : Vyssh. shk., 1990. 253 p. [in Russian].
5. NLTK. URL: <http://www.nltk.org/> (accessed 1 June 2021).
6. SegmentAnt. URL: <https://www.laurenceanthonny.net/software/segmentant/> (accessed 1 June 2021).
7. Stephen Doherty, 2017, Issues in human and automatic translation quality assessment. URL: [https://www.researchgate.net/publication/314261771\\_Issues\\_in\\_human\\_and\\_automatic\\_translation\\_quality\\_assessment](https://www.researchgate.net/publication/314261771_Issues_in_human_and_automatic_translation_quality_assessment) (accessed 1 June 2021).
8. OPUS. URL: <https://opus.nlpl.eu/OpenSubtitles.php> (accessed 1 June 2021).
9. Yakovyna V., Masiukevych V. Ohliad ta analiz metryk otsiniuvannia yakosti mashynnoho perekladu. [Review and analysis of machine translation quality metrics]. Visnyk Natsionalnoho universytetu "Lvivska politehnika". Kompiuterni nauky ta informatsiini tekhnologii. 2013. № 771. pp. 101–107 [in Ukrainian].