

# Дніпровський гуманітарний університет



## СУЧАСНА ПРИКЛАДНА ЛІНГВ ІСТИКА

Частина II

## ЗМІСТ

<b>ПЕРЕДМОВА</b> .....	5
<b>Алексеев Артур.</b> Взаємодія знань корпусної лінгвістики та інформаційних технологій у навчальному процесі .....	6
<b>Андрішко Олег.</b> Матеріали до словника міжмовних омонімів: українська мова й ізоляти / некласифіковані мови Африки .....	15
<b>Варакута Марина.</b> Мистецтво говоріння і слухання: ключі до ефективного спілкування .....	25
<b>Василенко Наталія.</b> Літературний портрет Катеринослава. Іван Манжура (1851-1893) – український поет-казкар, вчений-фольклорист. Концептуальні роздуми та музейні інтерпретації .....	34
<b>Глущенко Ганна.</b> Практичний досвід перекладу літератури за допомогою нейромережі: переваги та виклики .....	43
<b>Дудник Олена.</b> Політика українізації напр. ХХ – на поч. ХХІ ст. ....	53
<b>Квітинська Ліана.</b> Короткий екскурс у історію етносеміотики, питання та перспективи розвитку етносеміотики як науки .....	61
<b>Корсунська Вікторія.</b> Kommunikative Kompetenz als Element der interkulturellen Kommunikation und der beruflichen Interaktion eines Individuums in der gegenwärtigen Stufe der gesellschaftlichen Entwicklung .....	73
<b>Крашеніннікова Тетяна.</b> Роль мов у сприйнятті народів як націй .....	87
<b>Лимар Андрій.</b> Державна політика формування гуманітарної складової професійної ідентичності особистості студента .....	93
<b>Natsiuk Maryana. Osidak Viktoriia.</b> Applying multilingualism in foreign language teaching .....	103
<b>Негодченко Ольга.</b> Відображення етнокультури у творах Миколи Пимоненка .....	113
<b>Нежива Ольга.</b> Предметно-мовне інтегроване навчання іноземних мов у закладах вищої освіти .....	124
<b>Поповський Анатолій.</b> Дбаймо про культуру мови – державотворчого коду нації .....	132
<b>Rakutienė Silvija, Butrimė Edita.</b> Development of personal competences at a specialized university – a changing paradigm of traditional education .....	140
<b>Руколяньська Наталія.</b> Терміни на позначення абстрактних понять у субмові кримінального процесуального права .....	151
<b>Сергієнко Наталія.</b> Мовні аспекти юридичної термінології у сучасному законодавстві .....	159
<b>Сковронська Ірина.</b> Сучасні виклики та чинники збереження національної ідентичності українців .....	167
<b>Sukhoviyi Vitalii.</b> In the beginning was the word? Language and national consciousness. Ukrainian case .....	175
<b>Хмель-Дунай Галина.</b> До питання збереження автентичності в етнокультурному просторі .....	208

<b>Холод Ірина.</b> Особливості використання методології CLIL у викладанні англійської мови професійного спрямування .....	219
<b>Шапошнікова Анастасія.</b> Концепт «Паляниця» в етнографічних і фольклорних матеріалах, надрукованих наприкінці XIX – на початку XX ст. ....	227
<b>Шостак Оксана.</b> Лінгвістичні інструменти для створення психологічної адаптації .....	236
<b>Юрченко Анна.</b> Музейне заняття «Як козаки у розвідку ходили»: захід на стику історії, археології та патріотичного виховання .....	244
<b>ВІДОМОСТІ ПРО АВТОРІВ</b> .....	253

- (1883–2008 pp.) / упоряд. та автори передм.: І. Прокопенко, Леся Степовичка. Дніпропетровськ: ПП «Ліра ЛТД», 2009. Ч. І. С. 14–179.
8. Василенко Н. Є. Українська дитяча книжка в Катеринославі (1880–1910-і pp.). Історія і культура Придніпров'я: невідомі та маловідомі сторінки: наук. щорічник / гол. ред. Г. К Швидько. Київ: Нац. гірничий університет, видавець Олег Філюк, 2015. Вип. 11. С. 43–52.
9. Вирський Д. С. Манжура Іван Іванович. Енциклопедія історії України: у 10 т. / редкол.: В. А. Смолій (голова) та ін.; Інститут історії України НАН України. Київ: Наук. думка, 2009. Т. 6: Ла–Мі. С. 491.
10. Заремба В. І. Зажурена калина. Заремба В.І. Три шляхи до вівтаря: Історичні повісті. Дніпропетровськ: Січ, 1992. 455 с.; С. 4–174.
11. Іваннікова Л. В. Видатний учень вченого О. О. Потебні (І. І. Манжура). Іваннікова Л. В. Фольклористика Півдня України: сторінки історії. Запоріжжя, 2008. С. 51–80.
12. Каширіна Л. Іван Манжура та українська народна пісня. Народні пісні в записках Івана Манжури. Київ: Музична Україна, 1974. С. 5–13.
13. Поповський А. М. Південноукраїнські джерела в історії формування української літературної мови: монографія. Дніпро: Ліра, 2018. С. 250, 258.
14. Світленко С. І. Іван Манжура. Світленко С. І. Суспільний рух на Катеринославщині у 50–80-х роках ХІХ століття: Монографія. Дніпропетровськ: Вид-во ДНУ, 2006. С. 87–95.

## **ПРАКТИЧНИЙ ДОСВІД ПЕРЕКЛАДУ ЛІТЕРАТУРИ ЗА ДОПОМОГОЮ НЕЙРОМЕРЕЖІ: ПЕРЕВАГИ ТА ВИКЛИКИ**

**Ганна ГЛУЩЕНКО**

*старший викладач кафедри соціально-гуманітарних дисциплін  
та мовної підготовки КЗВО «Хортицька національна навчально-  
реабілітаційна академія»*

Великі мовні моделі, як-от ChatGPT, можуть надавати послідовні, релевантні, природні та адекватні рішення для багатьох завдань обробки природної мови (Natural language processing, NLP). Ця стаття використовує машинний переклад на рівні цілісного літературного твору, щоб оцінити здатність нейромережі перекладати з урахуванням контексту (як у межах тексту, так і поза ним), стильових особливостей і стилістичних засобів. Для оцінки якості перекладу використовувалися як метричні показники (sacreBLEU, TER, document-level sacreBLEU), так і суб'єктивне оцінювання людьми (носіями мови та професійними перекладачами), які звертали увагу на природність, послідовність, цілісність, вибір слів тощо. За результатами оцінювання нейромережа продемонструвала високі продуктивність і потенціал у перекладі на рівні літературного тексту, здатність досліджувати лінгвістичні дані. Робота окреслює проблеми та можливості ChatGPT у машинному перекладі, які стануть у пригоді як перекладачам-практикам, так і розробникам мовних моделей.

**Ключові слова:** великі мовні моделі (LLM), нейромережа, NLP, ChatGPT

Large language models, such as ChatGPT, can provide consistent, relevant, fluent and adequate solutions for many natural language processing (NLP) tasks. This article utilizes machine translation at the level of a literary work to evaluate the neural network's ability to translate with account for context (both within and beyond the text), stylistic features and means. Both evaluation metrics (sacreBLEU, TER, document-level sacreBLEU) and subjective evaluation by humans (native speakers and professional translators) were used to assess translation quality with human focus on consistency, coherence, word choice, and more. On evaluation, the large language model demonstrated high performance and showed potential in translating at the level of a literary work, demonstrated its ability to explore linguistic data. The work outlines the challenges and opportunities of ChatGPT in machine translation, which will be beneficial for both practicing translators and language model developers.

**Keywords:** large language models (LLM), NLP, ChatGPT

Питання автоматизації перекладу давно розглядаються (і вирішуються) лінгвістами усього світу як у теоретичній, так і в практичній площині. За понад сорок років ми пройшли довгий шлях від провальних спроб створити програму, що перекладатиме, як людина, до принципово іншого підходу з використанням CAT-інструментів, принципів машинного навчання, залучення великих даних і, зрештою, – великих мовних моделей (Large Language Models, LLMs), які й лягли в основу нейромереж.

Обсяг даних, які потребують перекладу, невпинно зростає, і деякі завдання неможливо виконати силами людей-перекладачів. Так, керівник підрозділу Європейської Комісії з досліджень та інновацій і розробник європейської стратегії у сфері даних Кіммо Россі пригадує, що під час вступу його батьківщини, Фінляндії, до Євросоюзу необхідно було перекласти всі нормативно-правові документи ЄС фінською мовою, щоб усі громадяни нової країни-члена отримали повний доступ і цілковите розуміння цих документів і почувалися рівноправними громадянами Євросоюзу. На той момент обсяг роботи становив понад 60 000 сторінок, і переклад виконувався людьми, але за десяток років обсяг виріс до 6 мільйонів сторінок. За розрахунками Россі, перекладачам Єврокомісії знадобилося б близько 100 років, щоб перекласти таку кількість нормативних документів усіма офіційними мовами ЄС – і це не враховуючи 2–3 млн сторінок документів, що додаються щороку. Лише для перекладу цих нових документів штат людей-перекладачів довелося б збільшити до 40–50 тисяч, щоб контент

залишався актуальним [6]. Натомість цю роботу виконує система MT@EC, в якій реалізовано парадигму статистичного машинного перекладу (Statistical Machine Translation, SMT) і машинного навчання на основі якісних перекладів, зроблених людиною. До системи вже додавали нову мову (норвезьку), тож після вступу України до ЄС систему можна буде навчити й українському перекладу.

І це лише один із багатьох аспектів. Наприклад, єдиний ринок європейських країн передбачає, що люди повинні мати змогу замовляти товари різних виробників, а, відповідно, розуміти інформацію на їхніх вебсайтах. Це стосується й глобальної економіки, тож великі маркетплейси, як-от Amazon чи AliExpress, давно розробляють і використовують власні системи, які базуються на тих ж принципах.

Коли наявні алгоритми (deep neural learning) поєдналися з критичною масою даних і збільшеною обчислювальною потужністю, відбувся той якісний стрибок, який уможливив ChatGPT і дав неспеціалістам шанс скористатися можливостями штучного інтелекту, а спеціалістам – дослідити ці можливості.

За останні кілька років машинний переклад досяг значного прогресу завдяки використанню попередньо навчених моделей, зокрема BERT, GPT-2 і T5. Ці моделі продемонстрували вражаючу продуктивність машинного перекладу [1; 2; 3], однак здебільшого роботи стосувалися перекладу на рівні речень, що призводить до генерації цільових перекладів, які логічно не пов'язані або не враховують контекст. Останнім часом спостерігається зростання інтересу до перекладу на рівні документів, що передбачає переклад цілісного тексту та моделювання певних рис дискурсу [4; 5; 9]. Найпопулярніша велика мовна модель ChatGPT демонструє здатність підтримувати зв'язність і послідовність розмови, враховуючи попередні повороти бесіди. Крім цього, її навчали на великій кількості діалогових

даних, що дозволило їй вивчити шаблони й умовності людського спілкування та покращило здатність до розуміння та генерування на рівні документів.

Мета цієї роботи – з’ясувати, як великі мовні моделі, зокрема ChatGPT, працюють на рівні літературного твору, забезпечуючи узгодженість, послідовність і цілісність у перекладі, а також оцінити здатність нейромережі перекладати з урахуванням контексту (як у межах тексту, так і поза ним), стильових особливостей і стилістичних засобів.

Дослідження проводилося на матеріалі різних за стилем і розміром творів із використанням різних мовних пар, як показано в таблиці 1.

Стиль	Назва	Мовна пара	Розмір (слів)
Художній	«God's gift», оповідання сучасної шотландської письменниці Алі Сміт	EN>UA	2940
Публіцистичний	«How to bag a geek», стаття з журналу The Economist	EN>DE EN>UA	1012

*Таблиця 1. Дані матеріалів перекладу.*

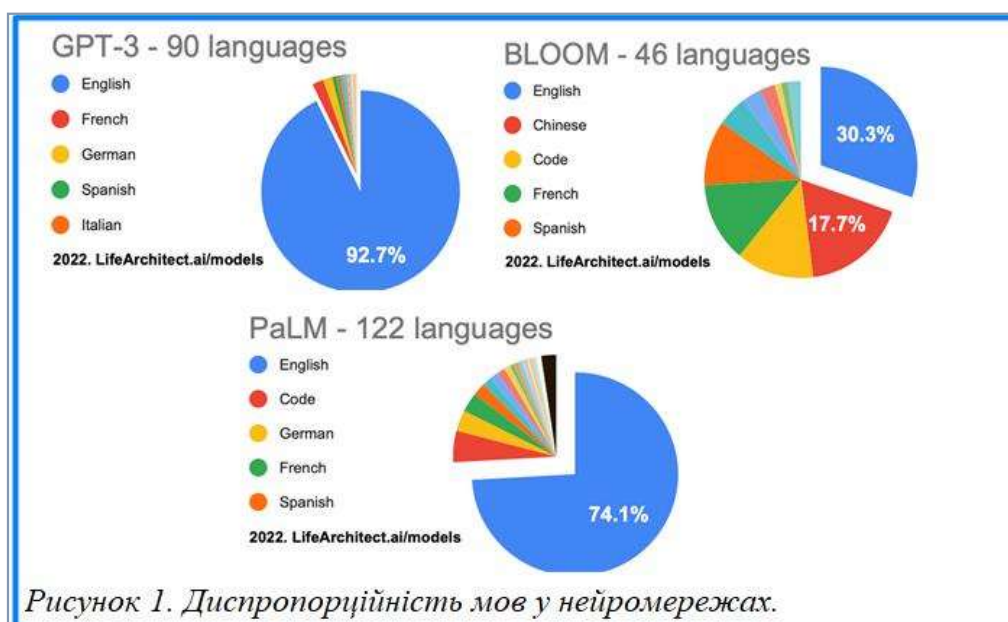
Для оцінювання якості перекладу використовувалися класичні метричні показники: sacreBLEU (Post, 2018) і TER (Snover 2006) на рівні речення, а також d-BLEU (document-level sacreBLEU) на рівні тексту [10]. Крім цього, проводилося суб’єктивне оцінювання людьми, які звертали увагу на природність, послідовність, цілісність, вибір слів тощо. Кожен переклад оцінили 4 людини (носії мови та професійні перекладачі) за шкалою 1–5 балів.

Найкращі показники отримано для публіцистичного тексту мовної пари EN>DE за умови використання контекстних підказок (таблиця 2).

BLEU			TER			d-BLEU			Людина		
худ	публ UA	публ DE	худ	публ UA	публ DE	худ	публ UA	публ DE	худ	публ UA	публ DE
13,3	22,1	26,7	76,3	59	62,2	16,5	27,4	29,1	2,5	2,7	3,5

*Таблиця 2. Показники якості перекладу.*

Частково ці результати були очікуваними. По-перше, існує значна диспропорція використання мов для створення й навчання всіх LLM або нейромереж (рис. 1) [7].



Ми бачимо, що, незважаючи на різноманіття мов, моделі здебільшого навчають саме на англійському контенті (окрім китайських систем, де доля китайської переважає). І якщо німецька входить до першої десятки мов, на українську припадає менше одного відсотка. Найкращим перекладом за таких умов був би переклад на англійську, а не з англійською на іншу, тут би





потенціал ШІ розкрився повною мірою. Крім цього, важливий і зміст контенту, на якому навчають LLM. Серед 10 основних наборів даних GPT-3 художня література не має переваги (рис. 2) [7].

Незважаючи на те, що тексти перекладалися реченнями, ChatGPT продемонстрував здатність моделювати довготривалу взаємозалежність. Що довшим ставав текст, то більше кількість вхідних даних підвищувала якість перекладу. Під час перекладу ChatGPT не завжди дотримувався меж і структури речення джерела, уникаючи інтерференції мови оригіналу в перекладі й імітуючи природній підхід, прийнятий людьми для перекладу текстів, який не вимагає перекладу від речення до речення. ChatGPT намагався забезпечити структурну цілісність і передати стиль автора.

Серед беззаперечних переваг перекладу за допомогою ChatGPT можна назвати такі:

- швидкість: штучний інтелект здатний обробляти великі обсяги тексту значно швидше, ніж людина; це прискорює переклад, що особливо важливо під час перекладу значних обсягів, як-от книжок;
- швидкий доступ до ресурсів: завдяки постійному доступу до баз даних (зокрема словників, термінологічних баз, довідкових матеріалів тощо) завжди є можливість знайти найкращий відповідник для перекладу не лише окремих слів чи термінів, а й сталих виразів, сленгових одиниць, безеквівалентної лексики тощо;
- нормативна правильність: хоча LLM не ідеальні, вони зменшують кількість граматичних, орфографічних і пунктуаційних помилок, опісок тощо;
- оптимізація відповідно до контексту: штучний інтелект аналізує та враховує контекст на рівні речення, що допомагає вибрати правильний варіант і уникнути двозначності в перекладі;

- дотримання стилю: навчені на великій кількості даних моделі можуть вдало відтворювати відповідний тон висловлювання, стиль автора, атмосферу оригіналу;

- багатоваріантність: у відповідь на запит ШІ може пропонувати альтернативні варіанти перекладу, допомагаючи підібрати найкращий відповідник у разі неоднозначності перекладу, стилістичних засобів тощо.

Варто взяти до уваги, що в питаннях стилю, контексту та варіативності довелося використовувати підказки, щоб спрямовувати ChatGPT до потрібного результату. Незважаючи на переваги, які надає використання штучного інтелекту, цілком покладатися на нього в перекладі художніх і публіцистичних текстів не варто, особливо в разі перекладу на таку «другорядну» в ієрархії LLM мову, як українська. ШІ не забезпечує уніфікацію скорочень і термінів протягом усього тексту, не завжди обирає точний відповідник чи розпізнає okazionalizmi, гру слів, метафори. Загалом відсоток речень, які не потребували корегування людиною (для перекладу на українську), становив лише 22 %.

Розглянемо приклад.

Оригінал: *Chunky signing-on bonuses and «precatations» – paid vacations before taking up a new position – are being dangled in front of tech folk to tempt them to jump ship.*

Переклад ChatGPT: *Перед технічними фахівцями висовуються великі бонуси за підписання контракту та «передплата» – оплачені відпустки перед прийняттям нової посади, щоб змусити їх перейти на інші проєкти.*

Перекласти *tech folk* як «технічні фахівці» було правильно по суті, але не по духу. Після корекції запиту ChatGPT миттєво запропонував кілька цікавих варіантів:

Технічні гуру

ІТ-віртуози

Програмістська братія

Комп'ютерні експерти

Гік-спільнота

Кодерське братство

Тех-маги

Інженери інновацій

Електронні генії

Інформаційні волхви

Корекція запиту спрацювала і з перекладом виразу *jump ship*, тобто скорегувати запит до нейромеріжі було швидше й простіше, ніж підбирати ці варіанти самостійно. А от у випадку *are being dangled* виявилось простіше перекласти самостійно, ніж пояснити метафоричність виразу (усілякі «приємності» виступають у ролі морквини перед носом віслюка, змушуючи його радо бігти в потрібному напрямку). Ось як виглядає переклад того ж речення людиною:

*Технарів заманюють і переманюють чималими вхідними бонусами й «передпустками» (оплаченими наперед відпустками перед вступом на нову посаду).*

Як ми бачимо, переклад людини виглядає природніше, враховує контекст, стиль і образність.

Крім цього, інколи терміни перекладаються ШІ буквально, без урахування специфічного контексту; зрідка трапляються граматичні й орфографічні помилки. Переклад не завжди точний із семантичної точки зору, не адаптований до цільової аудиторії, йому бракує творчого підходу. Нейромережа спирається на дані, на яких навчалася, а вони теж можуть бути помилковими. Так, у нашому тексті правильним варіантом перекладу *Silicon Valley* була б «Кремнієва долина», але нейромережа використала розповсюджений помилковий переклад «Силіконова долина», бо вона не може бути розумнішою, ніж її «вчитель» – масив даних. Але найголовніше – лише людина-перекладач може, в першу чергу, побачити ці прогалини,

скорегувати запит, спрямувати ШІ до довідкових матеріалів, пояснити задачу тощо.

Отже, під час роботи з нейромережею необхідно враховувати загальні обмеження великих мовних моделей, сформульовані Андреем Василевсом [8]:

- Розуміння контексту: хоча LLM здатні розуміти контекст у межах тексту, їм складно обробляти ширший контекст і бракує знань про реальний світ, яким їх не навчали.
- Залежність від даних: LLM не можуть генерувати точну інформацію, яка виходить за межі даних, на яких їх навчали, та інколи закріплюють наявну в цих даних упередженість.
- Креативність: LLM імітують творчість, спираючись на шаблони, але не творять у людському розумінні слова. Вони не продукують нові ідеї чи концепції так, як люди.
- Емоційний інтелект: у LLM відсутній емоційний інтелект, на відміну від людей вони не можуть по справжньому розуміти чи виражати емоції.
- Етичне судження: LLM не враховують етику й мораль і не розуміють моральні наслідки, які може мати згенерований ними текст.
- Надійність: відповіді LLM можуть бути неправильними чи взагалі не мати сенсу, тому необхідно враховувати це під час обробки важливої інформації.

Аналізуючи описаний досвід перекладу літератури за допомогою ChatGPT, можна дійти висновку, що цей процес потребує постійної взаємодії між людиною-перекладачем і штучним інтелектом, щоб досягти високої якості перекладу, оскільки кожна сторона привносить в процес свої унікальні можливості. Така взаємодія дає змогу поєднати точність і адекватність, адаптувати стиль і контекст, забезпечити якісну перевірку та редагування тексту. Поки що в цьому тандемі (більшою мірою для перекладу на українську) основним рушієм виступає людина. З розвитком алгоритмів і

технологій машинного перекладу ситуація може змінюватися, хоча й не схоже, що штучний інтелект зможе повністю замінити людину в творчому процесі.

Очікується, що штучний інтелект і надалі розвиватиметься в бік кращого розуміння контексту, стилю та культурних особливостей, що дасть змогу використовувати ШІ для попереднього перекладу й аналізу текстів (виявлення ключових термінів, стильових особливостей тощо). Наразі ChatGPT має кілька конкурентів у сфері перекладу (v. MarianMT, DeepL тощо), тож у подальшому буде цікаво порівняти процес і результат перекладу за допомогою інших нейромереж із отриманим досвідом роботи із ChatGPT.

### Список використаної літератури

1. Jinhua Zhu et al. *Incorporating BERT into Neural Machine Translation*. Conference proceedings at ICLR. 2020. URL: <https://arxiv.org/pdf/2002.06823.pdf>
2. Junliang Guo et al. *Incorporating BERT into parallel sequence decoding with adapters*. Advances in Neural Information Processing Systems. 2020. URL: <https://arxiv.org/pdf/2010.06138.pdf>
3. Linting Xue et al. *mT5: A massively multilingual pre-trained text-to-text transformer*. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021. URL: <https://arxiv.org/pdf/2010.11934.pdf>
4. Longyue Wang et al. *Document-Level Machine Translation with Large Language Models*. 2023. URL: <https://arxiv.org/pdf/2304.02210.pdf>
5. Longyue Wang et al. *Findings of the WMT 2023 shared task on discourse-level literary translation*. Proceedings of the Eighth Conference on Machine Translation (WMT). URL: <https://arxiv.org/pdf/2311.03127.pdf>
6. Rossi K. *Translation in the future – The future of translation*. Ukrainian Translation Industry Camp. 2016. URL: <https://utic.eu/en/video/video-uticamp-2016/>
7. Thompson A. D. *Inside language models (from GPT-4 to PaLM)*. URL: <https://lifearchitect.ai/models/#emerging>
8. Vasiljevs A. *Back to the Future. #2023TEF: Achieving excellence in translation*. 2023. URL: <https://2023tef.b2match.io/components/32450?session=c2Vzc2lrbjoxNTQ4NTY%3D>
9. Voita E., Sennrich R., Titov I. *Context-aware monolingual repair for neural machine translation*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019. URL: <https://arxiv.org/pdf/1909.01383.pdf>