

ФРАНЧУК Н.П.

*кандидат педагогічних наук, доцент,
доцент кафедри інформаційних технологій і програмування
Український державний університет імені Михайла Драгоманова, Київ;
старший науковий співробітник відділу відкритих освітньо-наукових
інформаційних систем
Інститут цифровізації освіти Національної академії
педагогічних наук України, Київ*

КОМП'ЮТЕРНІ МЕТОДИ ЛІНГВІСТИЧНОГО АНАЛІЗУ ТА ПЕРЕКЛАДУ: ІНТЕГРАЦІЯ ТЕХНОЛОГІЙ У СУЧАСНОМУ ОПРАЦЮВАННІ МОВИ

У сучасному світі значний обсяг даних генерується різними мовами, і завдяки розвитку комп'ютерних технологій з'являються нові можливості для опрацювання мовних даних. Лінгвістичний аналіз та переклад стали стратегічними напрямками для міжнародного спілкування, науки, бізнесу та культури.

Метою написання даної публікації є демонстрація того, як використання комп'ютерних методів лінгвістичного аналізу та перекладу інтегруються в мовне опрацювання, покращуючи якість перекладу та сприяють автоматизації лінгвістичних досліджень (Рис. 1).

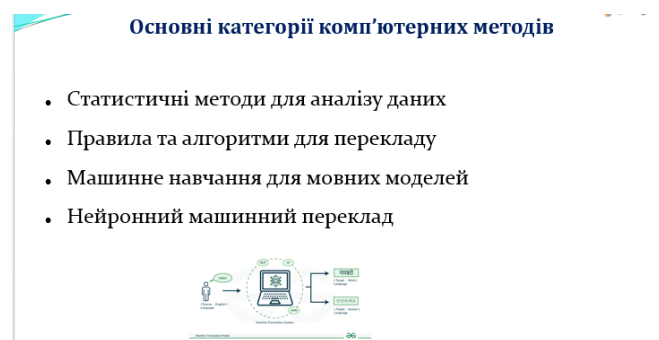


Рис. 1. Основні категорії комп'ютерних методів

Статистичний машинний переклад (Statistical Machine Translation, SMT) базується на паралельних корпусах текстів. Використання глибоких нейронних мереж (наприклад, трансформерів) значно покращило якість автоматичного перекладу, як-от у DeepL чи Google Translate. Моделі типу трансформерів та

нейронні мережі: GPT, T5, BERT як приклади трансформерних моделей, можуть розпізнавати контексти та забезпечувати високоточний переклад [1].

До основних напрямків комп'ютерного лінгвістичного аналізу відносять:

– *морфологічний та синтаксичний аналіз*: це системи, за допомогою яких можна автоматично визначати структуру слів та речень. У попередніх дослідженнях було описано використання певних програмних засобів для здійснення синтаксичного аналізу тексту [2]. До прикладу:

- ✓ Morphological <http://morphological.org/>
- ✓ Stilus <https://www.mystilus.com/en/tools/morphosyntactic-analyzer>
- ✓ Delhp-in <https://delph-in.github.io/delphin-viz/demo/>
- ✓ SDU <https://visl.sdu.dk/visl/en/parsing/automatic/parse.php>

– *семантичний аналіз*: це технології, за допомогою яких можна визначати значення слів у контексті, такі як Word Embeddings (GloVe, Word2Vec) і більш сучасні, наприклад, моделі на базі нейромереж типу BERT.

- ✓ NLTK (Natural Language Toolkit) <https://github.com/nltk/nltk>
- ✓ Hugging Face Transformers – бібліотека з великою кількістю моделей для NLP, включаючи сучасні нейронні мережі типу BERT, GPT та інших трансформерів <https://github.com/RaRe-Technologies/gensim>

– *дискурсивний та прагматичний аналіз* – оцінка мовленнєвих актів та визначення контекстуальних залежностей, що важливо для правильної інтерпретації змісту тексту. Для дискурсивного та прагматичного аналізу можна поєднувати різні моделі та інструменти на базі Google Colab (або іншої платформи з підтримкою Python), що дозволить поетапно проводити лексичний, синтаксичний, семантичний, а потім і дискурсивний аналіз тексту.

Застосування та практична користь комп'ютерних методів полягає в аналізі великих обсягів тексту, інтеграції з інтерфейсами штучного інтелекту та автоматичному перекладі у реальному часі.

Комп'ютерний лінгвістичний аналіз і машинний переклад суттєво впливають на автоматизацію роботи з мовними даними, забезпечуючи високі швидкість та точність опрацювання великих обсягів тексту. Використання цих

технологій дозволяє здійснювати масовий аналіз текстових корпусів та оперативно виконувати переклади, що значно розширює можливості міжмовної комунікації в наукових, технічних та соціальних галузях. Водночас існують певні обмеження таких підходів, серед яких – складнощі з адекватним перекладом культурно-специфічних елементів та труднощі у передачі емоційного й тонального забарвлення тексту, що є характерним для міжособистісного спілкування. Крім того, незважаючи на прогрес у розпізнаванні контексту, такі моделі все ще мають ризик помилок у його тлумаченні, що може призводити до викривлення змісту (Рис. 2).

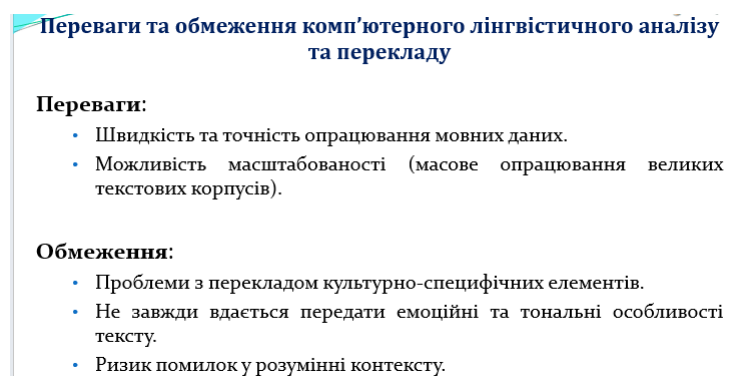


Рис. 2. Переваги та обмеження комп'ютерного лінгвістичного аналізу та перекладу

Інтеграція технологій глибокого навчання, трансформерів та мультиагентних систем у подальшому покращить якість та точність лінгвістичного аналізу та перекладу, оскільки застосування комп'ютерних методів лінгвістичного аналізу та перекладу відкривають нові горизонти у дослідженні мовних процесів та автоматизації перекладу, стаючи важливим інструментом у глобалізованому світі.

Список використаних джерел

1. Devlin, J., et al. (2018). «BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.» *arXiv preprint: <https://arxiv.org/pdf/1810.04805>*.

2. Франчук Н.П. Використання програм для автоматизованого синтаксичного аналізу тексту. *Теорія і практика використання інформаційних*

ХАЛЕЦЬКА З.П.

*кандидат фізико-математичних наук, доцент,
доцент кафедри математики та цифрових технологій
Центральноукраїнський державний університет
імені Володимира Винниченка*

СУЧАСНІ ТЕНДЕНЦІЇ ВПРОВАДЖЕННЯ STEM У ВИЩІЙ ОСВІТІ

На сьогоднішній день STEM (наука, технології, інженерія, математика) є одним з головних трендів освітньої політики багатьох розвинутих країн світу. Розвиток цифрових технологій у Європі відкриває безмежні можливості для трансформації різних сфер життя, особливо освіти. STEM-освіта стає все більш пріоритетною, а інтеграція цифрових інструментів у навчальний процес – невід’ємною частиною сучасної освіти.

В Україні спостерігається підвищена зацікавленість до навчання за STEM-напрямами. У наукових та науково-практичних працях з’ясовується зміст та понятійна система впровадження STEM-освіти. Сьогодні STEM-підходи реалізуються в багатьох українських школах та позашкільних закладах. Не оминув цей процес і заклади вищої освіти. Основною особливістю STEM-освіти є інтегроване навчання застосування науково-технічних знань у реальному житті. Науково-методичні засади створення моделі STEM-освіти полягають у переході від традиційного навчання до інноваційного. Тому, перед кожним закладом вищої освіти, незалежно від обраної стратегії, стоїть завдання пройти цифрову трансформацію.

Так, на кафедрі математики та цифрових технологій Центральноукраїнського державного університету імені Володимира Винниченка реалізується науково-дослідна тема «Цифровізація освітнього середовища та STEM-технології» (0122U201725). В рамках функціонування