

INTEGRATION OF AUTOMATED TECHNOLOGIES AND ARTIFICIAL INTELLIGENCE IN THE TRAINING OF FUTURE APPLIED LINGUISTS

*Nataliia Franchuk*¹

ІНТЕГРАЦІЯ АВТОМАТИЗОВАНИХ ТЕХНОЛОГІЙ ТА ШТУЧНОГО ІНТЕЛЕКТУ В НАВЧАННІ МАЙБУТНІХ ПРИКЛАДНИХ ЛІНГВІСТІВ

Наталія Франчук

Abstract. The article considers automated natural language processing technologies, covering the full range of tasks — from automatic transcription to machine translation. Modern methods, algorithms, and models are analyzed for processing text and audio data. Special attention is paid to integrating speech recognition technologies, syntactic and morphological analysis, and text generation. Achievements in neural networks, particularly transformers, are the basis of modern machine translation systems. The prospects for using these technologies in various areas are highlighted. Key challenges are outlined, particularly the problems of multilingualism, accuracy of text data processing, and ethical aspects. Conclusions are drawn regarding further developing automated natural language processing technologies to ensure effective interlingual communication and automation of routine linguistic tasks. The work aims to investigate and substantiate the introduction of automated technologies and artificial intelligence into the process of training future applied linguists, as well as to place their influence on the development of professional competencies in this field. The research methodology was theoretical and empirical methods, namely, analysis and synthesis of scientific sources on the use of automated technologies in the educational process, a systematic approach to the processing of natural language, observation, surveying, and testing of students who use modern technologies in learning. The scientific novelty of the work lies in the systematization of artificial intelligence technologies that are most effective for teaching applied linguistics; recommendations have been developed for integrating these technologies into curricula, considering the specifics of the professional activities of linguists. The influence of automated tools on improving students' linguistic and analytical skills has been revealed. Using automated technologies and artificial intelligence significantly increases the effectiveness of training future applied linguists.

Keywords: natural language processing, artificial intelligence, future applied linguists, speech recognition, syntactic analysis, morphological analysis, text generation, machine translation systems

¹ Dragomanov Ukrainian State University, Kyiv, Ukraine. Institute for Digitalisation of Education of NAES of Ukraine, Kyiv, Ukraine. n.p.franchuk@udu.edu.ua, <http://orcid.org/0000-0002-0213-143X>

Анотація. У статті розглянуто автоматизовані технології опрацювання природної мови, що охоплюють повний спектр завдань — від автоматичної транскрипції до машинного перекладу. Проаналізовано сучасні методи, алгоритми та моделі, які використовуються для опрацювання текстових та аудіоданих. Особлива увага приділяється інтеграції технологій розпізнавання мовлення, синтаксичного й морфологічного аналізу, а також генерації тексту. Розглянуто досягнення в галузі нейронних мереж, зокрема трансформерів, які є основою сучасних систем машинного перекладу. Висвітлено перспективи використання цих технологій у різних сферах. Окреслено ключові виклики, зокрема проблеми багатомовності, точності опрацювання текстових даних та етичні аспекти. Зроблено висновки щодо важливості подальшого розвитку автоматизованих технологій опрацювання природної мови для забезпечення ефективної міжмовної комунікації та автоматизації рутинних лінгвістичних завдань. Мета роботи: дослідити та обґрунтувати впровадження автоматизованих технологій і штучного інтелекту в процес навчання майбутніх прикладних лінгвістів, а також дослідити їх вплив на розвиток професійних компетентностей у цій галузі. Методологією дослідження слугували теоретичні та емпіричні методи, а саме: аналіз і синтез наукових джерел щодо використання автоматизованих технологій в освітньому процесі, системного підходу до опрацювання природної мови, спостереження, опитування та тестування студентів, які використовують сучасні технології в навчанні. Наукова новизна роботи полягає у систематизації технологій штучного інтелекту, які найбільш ефективні для навчання прикладної лінгвістики; розроблено рекомендації щодо інтеграції цих технологій у навчальні програми з урахуванням специфіки професійної діяльності лінгвістів. Виявлено вплив автоматизованих інструментів на покращення лінгвістичних та аналітичних навичок студентів. Використання автоматизованих технологій і штучного інтелекту значно підвищує ефективність підготовки майбутніх прикладних лінгвістів.

Ключові слова: опрацювання природної мови, штучний інтелект, майбутні прикладні лінгвісти, розпізнавання мовлення, синтаксичний аналіз, морфологічний аналіз, генерація тексту, системи машинного перекладу

Вступ

Актуальність дослідження інтеграції автоматизованих технологій та штучного інтелекту в навчанні майбутніх прикладних лінгвістів обумовлена кількома факторами. Сучасний ринок праці потребує фахівців, які володіють не тільки лінгвістичними знаннями, а й навичками з використанням автоматизованих інструментів, таких як системи машинного перекладу, аналізу текстів та штучного інтелекту. Інтеграція таких технологій у навчання дозволяє підготувати студентів до викликів цифрової епохи. Штучний інтелект активно використовується в автоматичному перекладі, створенні мовних моделей, розпізнаванні мови та тексту. Майбутні прикладні лінгвісти не повинні лише розуміти принципи роботи цієї системи, але й уміти використовувати їх у своїй професійній діяльності. Традиційні методи навчання вже не забезпечують достатньої гнучкості й актуальності для сучасних реалій. Інтеграція автоматизованих технологій у навчальні програми дозволяє зробити навчання інноваційним, практично орієнтованим і конкурентоспроможним.

Використання сучасних технологій у навчанні інформатичних дисциплін майбутніх прикладних лінгвістів, полегшує закріплення складних концепцій і демонструє прикладне застосування теоретичних знань. Ця галузь перебуває на перетині лінгвістики, інформаційних технологій та когнітивних наук. Актуальність дослідження зумовлена потребою в розробці методів навчання, що відносяться до міждисциплінарного характеру професійної діяльності прикладних лінгвістів. У контексті глобалізації й багатомовного спілкування попит на швидкі, точні та ефективні рішення в лінгвістичній сфері зростає. Це вимагає від фахівців розуміння сучасних інструментів для автоматизації перекладу, опрацювання даних і аналізу тексту.

Аналіз останніх публікацій з досліджуваної тематики

Автоматизовані технології опрацювання природної мови (natural language processing, NLP) є однією з ключових сфер штучного інтелекту, яка стрімко розвивається. Застосування цих технологій спрямовані на автоматизацію процесів аналізу, розуміння, генерації та перекладу текстових й мовленнєвих даних. Усе це робить NLP незамінним інструментом для розв'язання сучасних проблем у різних сферах: від освіти до комерційних застосувань [1], [2].

О. Каваль та М. Умрик в роботі [3] описують інтеграцію штучного інтелекту для автоматизації процесу створення освітніх тестів.

На паралельних корпусах текстів базується статистичний машинний переклад (Statistical Machine Translation) [4]. Використання глибоких нейронних мереж (до прикладу, трансформерів) ґрунтовно покращило якість автоматичного перекладу, особливо це помітно в DeepL (<https://www.deepl.com/uk/translator>) та Google Translate (<https://translate.google.com.ua>). Моделі типу трансформерів та нейронні мережі: GPT, T5, BERT як приклади трансформерних моделей, можуть розпізнавати контексти та забезпечувати високоточний переклад [5].

У навчально-методичному посібнику [6] подано основні концепти, цілі та методи комп'ютерної лінгвістики, висвітлено її історичні витоки, сучасний стан розвитку та міждисциплінарні зв'язки. Розглянуто провідні підходи до автоматизованого синтаксичного та семантичного аналізу природної мови, а також передбачено можливості їх практичного застосування. Особливу увагу О.В. Ваховська приділяє ключовим аспектам взаємозв'язку комп'ютерної лінгвістики з технологіями штучного інтелекту.

Мета дослідження: дослідити та обґрунтувати впровадження автоматизованих технологій і штучного інтелекту в процес навчання майбутніх прикладних лінгвістів, а також дослідити їх вплив на розвиток професійних компетентностей у цій галузі.

Подання основного матеріалу

Сучасні технології опрацювання природної мови базуються на досягненнях у галузі машинного навчання, зокрема нейронних мереж, які дозволяють опрацьовувати великі обсяги текстових та аудіоданих з високою точністю. Розглянемо основні методи, алгоритми та моделі, що є актуальними в цій галузі.

До появи нейронних мереж опрацювання тексту значною мірою базувалося на статистичних моделях та методах на основі правил. Для прогнозування наступного слова в тексті на основі попередніх використовувалися статистичні N-грамні моделі. Хоча цей підхід є простим, він має обмеження у врахуванні довгострокових залежностей. Також використовувався алгоритм тематичного моделювання для класифікації текстів за темами.

Розглянемо автоматизовані технології опрацювання природної мови, що охоплюють повний спектр завдань (див. табл. 1) [7].

Таблиця 1. Основні аспекти NLP

<i>Аспект</i>	<i>Характеристики</i>
Автоматична транскрипція мовлення	Автоматичне розпізнавання мовлення (Automatic Speech Recognition, ASR) дозволяє перетворювати аудіофайли в текст. Сучасні системи ASR, наприклад Whisper від OpenAI, використовують нейронні мережі для розпізнавання мови навіть у складних акустичних умовах. Їхнє використання сприяє створенню субтитрів, голосових помічників та аналізу дзвінків.
Морфологічний та синтаксичний аналіз	На цьому етапі текст розбивається на окремі слова або морфеми для аналізу їх граматичних властивостей. Синтаксичний аналіз передбачає побудову дерев залежностей, що дозволяє зрозуміти структуру речень. Ці завдання розв'язуються за допомогою інструментів, таких як SpaCy (бібліотека програмного забезпечення з відкритим вихідним кодом для опрацювання природної мови, написана на мовах програмування Python і Cython).
Семантичний аналіз і моделі трансформерів	Семантичний аналіз дозволяє машинам «розуміти» значення тексту. Тут використовуються трансформери, наприклад BERT або Generative Pre-trained Transformer (GPT), які вивчають контекст кожного слова в реченні. Завдяки цьому можна виконувати завдання контекстного пошуку, емоційного аналізу та автоматичного узагальнення тексту.
Машинний переклад	Системи машинного перекладу, такі як Google Translate чи DeepL, застосовують нейронні моделі для створення високоякісних перекладів. Технологія трансформерів значно підвищила якість перекладу, враховуючи як локальний, так і глобальний контекст речення.

Методи на основі правил залучають граматичні й семантичні конструкції, які задаються вручну. Прикладом є аналіз синтаксичних дерев та частиномовна розмітка. Хоча цей підхід забезпечує контрольованість, він погано масштабується для складних мов.

Тепер все частіше використовуються рекурентні нейронні мережі (RNN) та довгу короткочасну пам'ять Long Short-Term Memory (LSTM) для опрацювання послідовностей, таких як речення, але є обмеження у врахуванні довгих контекстів. Також використовується архітектура трансформерів (Transformers), що частково замінює RNN завдяки здатності паралельно опрацювати дані.

Двонаправлений шифратор репрезентацій з трансформерів — Bidirectional Encoder Representations from Transformers (BERT) використовується для завдань класифікації тексту, пошуку даних та семантичного аналізу.

Генеративний попередньо навчений трансформер — Generative Pre-trained Transformer (GPT) генерує текст із врахуванням контексту, та є основою чат-ботів та «творчих застосунків».

Трансформер передавання тексту в текст — Text-to-Text Transfer Transformer (T5) — це універсальна модель для багатьох задач NLP, включаючи переклад, класифікацію та генерацію тексту [8].

Сьогодні автоматизовані технології опрацювання природної мови застосовуються у всіх сферах, зокрема для: автоматичного оцінювання текстів, використання платформ для інтерактивного навчання [9], створення та використання чат-ботів, аналізу електронних записів, розпізнавання симптомів за описами, автоматизації опрацювання запитів, опрацювання великих обсягів текстових даних, автоматичного аналізу дослідницьких публікацій та ін. У попередніх публікаціях було досліджені деякі з цих питань [10].

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 [10]. Наприклад:
 - Delph-in <https://delph-in.github.io/delphin-viz/demo/>
 - Morphological <http://morphological.org/>
 - SDU <https://visl.sdu.dk/visl/en/parsing/automatic/parse.php>
 - Stilus <https://www.mystilus.com/en/tools/morphosyntactic-analyzer>
- *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 пра-

вільної 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 [5].

Висновки та перспективи подальших досліджень

Автоматизовані технології опрацювання природної мови стали основою сучасних цифрових інновацій. Їх подальший розвиток дозволить значно підвищити ефективність роботи з даними, покращити міжмовну комунікацію та забезпечити доступ до знань для ширшої аудиторії. Інтеграція NLP у повсякденні процеси відкриває нові можливості для автоматизації, підвищення якості обслуговування та розв'язування соціальних викликів.

Однак є певні аспекти, які варто врахувати та детальніше відобразити у подальших дослідженнях: розроблення моделей для мов із невеликою кількістю доступних текстових даних залишається значною проблемою; важливість уникнення дискримінаційних чи упереджених результатів, зумовлених обмеженнями навчальних даних; удосконалення моделей для роботи на пристроях з обмеженими ресурсами.

Література

- [1] Kuzminov I.F., Bakhtin P.D., Timofeev A.A. *et al.* 2021. Modern Natural Language Processing Technologies for Strategic Analytics. *Sci. Tech. Inf. Proc.* 48. P. 467–475. <https://doi.org/10.3103/S0147688221060071>
- [2] The NLP Playbook: From Basics to Advanced Techniques and Algorithms—Let's Data Science. <https://letsdatascience.com/natural-language-processing/> (Дата звернення: 17.10. 2024).
- [3] Коваль О., Умрик М. 2024. Використання штучного інтелекту для автоматизації процесу створення освітніх тестів. *Міждисциплінарні дослідження складних систем.* № 24 (2024). С. 78-96. <https://doi.org/10.31392/iscs.2024.24.078>

- [4] Devlin J., *et al.* 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint*. <https://arxiv.org/pdf/1810.04805> (Дата звернення: 1.11.2024).
- [5] Франчук Н. П. 2024. Комп'ютерні методи лінгвістичного аналізу та перекладу: інтеграція технологій у сучасному опрацюванні мови. *Матеріали VI Всеукраїнської науково-практичної конференції «Сучасні інформаційні технології в освіті і науці» м. Умань. 14–15 листопада 2024 р.* <https://lib.iitta.gov.ua/id/eprint/743482> (Дата звернення: 10.12.2024).
- [6] Ваховська О. В. 2023. Основи комп'ютерної лінгвістики: навчально-методичний посібник. Київ : Видавничий центр КНЛУ, 112 с. http://rep.knlu.edu.ua/xmlui/bitstream/handle/787878787/3115/Ваховська_Основи%20комп%27ютерної%20лінгвістики_посібник%20%281%29.pdf?sequence=1&isAllowed=y (Дата звернення: 12.11.2024).
- [7] Франчук Н. П. 2024. Автоматизовані технології опрацювання природної мови: від транскрипції до машинного перекладу. *Навчання і викладання у багатомовному світі у цифровому форматі в Україні сьогодні = Digitales Lernen und Lehren in der deutschsprachigen Welt in der heutigen Ukraine* : матеріали Міжнародної науково-практичної онлайн-конференції, 29–30 листопада 2024 р. С. 196–198. <https://doi.org/10.31392/UDU-MKfpgoe-2024.40>
- [8] Франчук Н. П. 2024. Використання штучного інтелекту в прикладній лінгвістиці. *Матеріали Міжнародної науково-практичної конференції «Цифрова трансформація освіти: теоретико-методичні засади» з нагоди 70-річчя від дня народження доктора педагогічних наук, професора, заслуженого працівника освіти України, директора Навчально-наукового інституту перепідготовки та підвищення кваліфікації Сергієнка Володимира Петровича.* м. Київ. 28 жовтня 2024 р. С. 408–409. <https://doi.org/10.31392/UDU-MKNino-Sergiyenko70-2024>
- [9] Франчук Н. П. 2017. Створення комп'ютерно-орієнтованого методичного забезпечення навчально-виховного процесу. *Науковий часопис НПУ імені М. П. Драгоманова. Серія № 2. Комп'ютерно-орієнтовані системи навчання.* 19(26). С. 181–187.
- [10] Франчук Н. П. 2023. Використання програм для автоматизованого синтаксичного аналізу тексту. *Теорія і практика використання інформаційних технологій в умовах цифрової трансформації освіти.* 29 червня 2023. м. Київ. С. 69–71. <https://lib.iitta.gov.ua/736682> (Дата звернення: 12.11.2024).

References

- [1] Kuzminov I. F., Bakhtin P. D., Timofeev A. A. *et al.* 2021. Modern Natural Language Processing Technologies for Strategic Analytics. *Sci. Tech. Inf. Proc.* 48. P. 467-475. <https://doi.org/10.3103/S0147688221060071>
- [2] The NLP Playbook: From Basics to Advanced Techniques and Algorithms—Let's Data Science. <https://letsdatascience.com/natural-language-processing/> (Accessed: 17.10.2024).
- [3] Koval O., Umryk M. 2024. Vykorystannia shtuchnoho intelektu dlia avtomatyzatsii protsesu stvorennia osvitynikh testiv. [The use of artificial intelli-

- gence for the automation of the process of creating educational tests.] *Mizhdystyplinarni doslidzhennia skladnykh system = Interdisciplinary Studies of Complex Systems*. № 24 (2024). S. 78–96. <https://doi.org/10.31392/scs.2024.24.078>
- [4] Devlin J. *et al.* 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint*: <https://arxiv.org/pdf/1810.04805> (Accessed: 1.11.2024).
- [5] Franchuk N. P. 2024. Kompiuterni metody lnhvistychnoho analizu ta perekladu: intehratsiia tekhnolohii u suchasnomu opratsiuvanni movy. [Computer methods of linguistic analysis and translation: integration of technologies in modern language processing]. *Materialy VI Vseukrainskoi naukovo-praktychnoi konferentsii «Suchasni informatsiini tekhnolohii v osviti i nauksi»* m. Uman. 14–15 lystopada 2024 r. <https://lib.iitta.gov.ua/id/eprint/743482> (Accessed: 10.12.2024).
- [6] Vakhovska O. 2023. Osnovy kompiuternoii lnhvistyky [Basics of computational linguistics] : navchalno-metodychnyi posibnyk. Kyiv : Vydavnychi tsentr KNLU, 112 s. http://rep.knlu.edu.ua/xmlui/bitstream/handle/787878787/3115/Важовська_Основи%20комп%27ютерної%20лінгвістики_посібник%20%281%29.pdf?sequence=1&isAllowed=y (Accessed: 12.11.2024).
- [7] Franchuk N. P. 2024. Avtomatyzovani tekhnolohii opratsiuvannia pryrodnoi movy: vid transkryptsii do mashynnoho perekladu. [Automated technologies for natural language processing: from transcription to machine translation.] *Navchannia i vykladannia u bahatomovnomu sviti u tsyfrovomu formati v Ukraini sohodni = Digitales Lernen und Lehren in der deutschsprachigen Welt in der heutigen Ukraine : materialy Mizhnarodnoi naukovo-praktychnoi onlain-konferentsii*, 29–30 lystopada 2024 r. S. 196–198. <https://doi.org/10.31392/UDU-MKfpgoe-2024.40>
- [8] Franchuk N. P. 2024. Vykorystannia shtuchnoho intelektu v prykladnii lnhvistytsi. [The use of artificial intelligence in applied linguistics]. *Materialy Mizhnarodnoi naukovo-praktychnoi konferentsii «Tsyfrova transformatsiia osvity: teoretyko-metodychni zasady» z nahody 70-richchia vid dnia narodzhennia doktora pedahohichnykh nauk, profesora, zasluženoho pratsivnyka osvity Ukrainy, dyrektora Navchalno-naukovoho instytutu perepidhotovky ta pidvyshchennia kvalifikatsii Serhiienka Volodymyra Petrovycha*. m. Kyiv. 28 zhovtnia 2024 r. S. 408–409. <https://doi.org/10.31392/UDU-MKNino-Sergiyenko70-2024>
- [9] Franchuk N. P. 2017. Stvorennia kompiuterno-oriientovanoho metodychnoho zabezpechennia navchalno-vykhovnoho protsesu. [Creating a computer-oriented methodical support of the training-educational process]. *Naukovyi chasopys NPU imeni M. P. Drahomanova. Seriiia № 2. Kompiuterno-oriientovani systemy navchannia*. 19(26). S. 181–187.
- [10] Franchuk N. P. 2023. Vykorystannia prohram dlia avtomatyzovanoho syntaktychnoho analizu tekstu. *Teoriia i praktyka vykorystannia informatsiinykh tekhnolohii v umovakh tsyfrovoi transformatsii osvity*. [Using programs for automated syntactic text analysis. *Theory and practice of using information technologies in the context of digital transformation of education*]. June 29 chervnia, 2023. Kyiv. S. 69–71. <https://lib.iitta.gov.ua/736682> (Accessed: 12.11.2024).