

---

## СУЧАСНІ ТЕНДЕНЦІЇ РОЗВИТКУ ПРИКЛАДНОЇ ЛІНГВІСТИКИ (Із досвіду розробки міждисциплінарного курсу "Автоматизація обробки природної мови і моделювання мовленнєвої діяльності")

---

УДК 811.161.2:378.14

**Черницький В.Б. (Chernytskyu V.B.)**

Національний університет кораблебудування ім. адмірала С.О. Макарова

*Проаналізовано структура дисципліни АОПМ та описано основні моменти побудови курсу в рамках дисципліни "Прикладна лінгвістика".*

*Ключові слова: мовленнєва діяльність, математична лінгвістика, лінгвостатистика, лінгвістичне забезпечення.*

*The article comprises multiple examples of defining various trends of applied linguistics based on the examples of its modern achievements and author's research experience in the field.*

*Key words: speech actualization, mathematical linguistics, linguistical statistics, lingware.*

Існує, щонайменше, три причини, які повертають нас до розглядання предмету і напрямків розвитку прикладної лінгвістики. По-перше, це розвиток теоретичних досліджень і практичне втілення їхніх результатів, що само по собі відображує еволюцію знань про природу мови і мовлення на традиційних трьох методологічних рівнях дослідження мовних можливостей: фундаментального вивчення мовних явищ, типового когнітивно-алгоритмічного моделювання, розробки методів лінгвістичного забезпечення прикладних систем.

По-друге, за останні роки якісно змінився зміст існуючих комплексних наукових дисциплін, а на базі відносно нових із них почали швидко розвиватись перспективні напрямки, такі, наприклад, як контент-аналіз та лінгвістика політичного дискурсу, сурдолінгвістика, судова лінгвістика тощо.

По-третє: коли використання комп'ютерів стало звичним у повсякденному житті, глобальна інформатизація діяльності суспільства

призвела до якісно нового розуміння тієї базової ролі, яку відіграє лінгвістичне забезпечення баз даних (БД), баз знань (БЗ), як основ різноманітних систем штучного інтелекту (ШІ), - експертних систем, природномовних інтерфейсів, які обслуговують комунікацію між людиною й ЕОМ, - а також мережевих комунікацій тощо. Наприклад, комп'ютеризація лексикографічної діяльності дала можливість широкого застосування машинних фондів мови (лексичних і лексикографічних БД), що забезпечує вирішення низки окремих завдань вивчення національних мов в цілому, їх функціональних різновидів, автоматичного й автоматизованого створення різноманітних словників, машинного перекладу (МП), накопичення, зберігання й підтримки матеріалів лексичних БД для ПС, електронних бібліотек тощо. Вирішення таких завдань реалізується на основі результатів розробки такої базової дисципліни як автоматизація аналізу и синтезу текстів на природній мові тощо.

З іншого боку, сучасна практика всебічного впровадження потужних ПС в промисловість, транспорт, як невід'ємних елементів управління комплексними процесами у сфері обслуговування, освіти призвели до інтенсифікації розробок в галузі розпізнавання образів, щільно пов'язаних з компетенцією лінгвістики в удосконаленні роботи пристроїв машинного зчитування рукописних текстів. Всі ці досягнення потребують систематизації або зміни структури і завдань окремих галузей і напрямлень прикладної лінгвістики.

Крім того, якщо в 60-90-ти роки минулого століття роль мовознавства як базової науки при вивченні мови й мовленнєвих процесів дуже часто нехтувалась за рахунок запозичення методів теорії наукової інформації, кібернетики та інших галузей знання, то сьогодні цілком очевидним є факт необхідності теоретичного й експериментального вивчення властивостей мови з точки зору її когнітивних й комунікативних можливостей. Тому є дуже корисним інвентаризація основних завдань, цілей, методів, обсягу й структури головних галузей й напрямків прикладної лінгвістики, в тому числі й у сфері освіти заради уточнення предмету тієї галузі прикладної лінгвістики, яка є

винесеною в підзаголовок статті, а саме „*Моделювання мовленнєвої діяльності і АОПМ*”.

Визначення предмету і завдань усіх існуючих „лінгвістик” дає можливість виявити їх належність до конкретних дисциплін або областей науки і, таким чином, сформувані більш-менш чітке розуміння предмета аналізу.

В історичному плані розвитку прикладної лінгвістики логічно почати із *математичної (або структурно-математичної) лінгвістики*. Назва цієї дисципліни увійшла в наукову практику в 50-ті роки минулого століття і на сьогодні набула широкого розповсюдження. Однак, існує ціла низка трактувань цієї дисципліни, серед яких є і, вочевидь, помилкові. З одного боку, ця галузь згадується в дуже широкому і розпливчастому сенсі, тобто її застосовують в лінгвістичних дослідженнях, якщо в них, хоча й в незначному виді використовується математика, або так здається дослідникам-початківцям. До цієї області відносять і праці, пов'язані із створенням моделей мови, на базі використання апарату алгебри, математичної логіки, теорії алгоритмів. А також роботи, пов'язані із статистикою, навіть суто лінгвістичні розвідки, в яких використовуються комп'ютери, у тому числі і з *автоматичного (або машинного) перекладу*.

Нам здається, що найбільш вдале визначення *математичної лінгвістики*, як „...певного кола математичних, по суті, робіт, які виникли із спроб чітко описати факти природних мов і які містять результати, які можуть бути корисними для лінгвістів” було надано Гладким і Мельчуком [2]. До цього кола робіт не мають відношення ті напрямки досліджень, у яких мова є лише одним із можливих додатків (зокрема роботи із *квантитативної лінгвістики*). Для *математичної лінгвістики* характерним є використання лише таких математичних методів, які у певному сенсі є специфічними для мови як такої. Таким чином, ми можемо підкреслити, що „*математична лінгвістика* є математична дисципліна, яка „спрямована” у бік природної мови і лінгвістики” [2, 16].

Таким чином, щоб „закрити” питання про належність *математичної лінгвістики* до *прикладної лінгвістики*, і, зокрема, до *АОПМ*, достатньо стисло

визначити об'єкт і методи математичної лінгвістики за допомогою зрозумілих для лінгвістів уявлень. Математична лінгвістика є не кількісною дисципліною, хоча вона й використовує апарат теорії алгоритмів і алгебри. Її мета не підрахування типу „скільки разів це слово, клас тощо, зустрічається у тексті”, тому що визначення числа яких-небудь об'єктів у більшості випадків виконується не експериментально, а чисто дедуктивно, тобто за допомогою суджень, які є вихідними із їх абстрактних властивостей. Насправді, до теорії алгоритмів і алгебри також нерідко залучають кількісні міркування, котрі, проте, залишаються на другому плані.

Може здаватися, що цілком природно такі судження безпосередньо мають відношення до *квантитативної лінгвістики* або до *статистичної лінгвістики* (інша назва – *лінгвостатистика*). Але до основних задач *лінгвостатистики* входять статистичний відбір учбового матеріалу що є часткою *методики викладання іноземної та рідної мови*. В *психолінгвістиці* це статистичні закономірності відображення авторського мовлення в залежності від жанру і стилю, результатів впливу висловлювання на читача тощо. В *нейролінгвістиці* статистичні обстеження стосуються взаємодії й взаємозалежності між будовою мови і можливостями мозку. Це, перш за все, об'єму пам'яті, властивостей запам'ятовування і відтворення мовних або мовленнєвих ланцюжків як психофізіологічних характеристик, які безпосередньо впливають на мовленнєві.

Прикладна лінгвістика, в цілому, використовує статистичні дослідження при вирішенні завдань *автоматизації редакційно-видавничих процесів*, *квантитативної типологізації текстів*, *атрибуції текстів* анонімних або невідомих авторів (встановлення авторства рукописів). Якщо ми приймемо припущення про мову і мовлення як складну систему дискретних одиниць, то їх кількісні властивості можна використати у дослідженнях історії розвитку і класифікації мов як суспільних явищ, в порівняльно-історичному мовознавстві, в галузі лексикографії при укладанні частотних словників тощо. Це є традиційні галузі *лінгвостатистики*, які обслуговують базові завдання загального мовознавства. Але набагато більше значення, як свідчить один із

найавторитетніших експертів прикладної лінгвістики В.І. Перебийніс, - "...особливе значення мають ці [статистичні – В.Ч.] методи для вирішення цілої низки проблем, так чи інакше пов'язаних з автоматичною обробкою лінгвістичних даних за допомогою ЕОМ". [ 4, 12] Таким чином, ми з'ясували, що найбільш серйозні розробки в прикладній лінгвістиці, з точки зору використання статистичних методів, безпосередньо пов'язані з автоматичною обробкою природної мови.

Поява комп'ютера стала величезною допомогою в клопіткій праці дослідника із прикладної лінгвістики. На перших етапах використання комп'ютера в лінгвістиці (треба зауважити, що першими користувачами ЕОМ стали лексикографи) виявилось, що він є дуже зручною друкарською машинкою з можливостями витирання, заміни літер, слів і навіть великих частин тексту, або його зберігання без порушення цілості.

Пізніше комп'ютер додав лексикографу й такі канцелярські інструменти повсякденної праці, як ножиці, пензлик, клей. Далі, за допомогою спеціалізованого ПЗ були підпорядковані функції, що здійснювались технічними редакторами при підготовці рукопису словника в набір, і ЕОМ стала своєрідним мікровидавництвом.

В галузі *АОПМ* планшетні, а потім ручні сканери незрівнянно прискорили процес набору текстів і картотек у вигляді графічних файлів; *FineReader* став незамінним коректором. Водночас лінгвісти опрацювали можливості нового ПЗ – функції автоматичного пошуку, заміни, роботи з індексованими текстами, що призвело до створення множинних входів у словники. У свою чергу, це прискорило розробку лінгвістичних БД і інформаційних підходів створення лінгвістичних ПС на основі програмного середовища СУБД. Саме автоматизовані лексикографічні системи стали основою лінгвістичного забезпечення автоматизованого й машинного редагування, інформаційного пошуку, розпізнавання і корегування текстів.

Автоматизовані лексичні БД і термінологічні тезауруси зараз становлять невід'ємну основу систем машинного перекладу та правки текстів, а також

переробки великих текстових масивів та їх зберігання як окремих завдань створення й підтримки мовних корпусів й електронних бібліотек.

Курс „*АОПМ і моделювання мовленнєвої діяльності*”, який на протязі 5 років викладається в Національному університеті кораблебудування ім. адм. Макарова та в Миколаївському державному університеті ім. Сухомлинського для студентів спеціальності „Прикладна лінгвістика”, базується на класичних роботах, матеріалах Internet ресурсів, а також на моєму особистому досвіді теоретичних і практичних розробок в галузі автоматичної обробки природної мови [1],[5]. На практичних заняттях моделюється вирішення лінгвістичних завдань, націлених на формування вмінь розробки компонентів лінгвістичного процесора, автоматизації словникарства. В ході лабораторних робіт засвоюються знання про відтворення мовних явищ в програмному середовищі існуючих прикладних ПС, електронних словниках, алгоритмах, які працюють зі словниками на корпусних масивах текстів. Типовими завданнями курсового проекту є: - 1) редагування і створення статей Wikipedia та аналогічних проєктів Wiktionary, Wikiquote тощо українською мовою; - 2) створення компонентів лінгвістичного процесора; - 3) розробка словникових статей і програмування за допомогою DSL тематичних підрозділів електронного англо-українського словника із суднобудування тощо.

Робочий приклад змісту дисципліни:

1. *Загальні проблеми автоматичної обробки природно мовних текстів.* Принципи і методи моделювання мовних явищ. Синтагматичні й парадигматичні відношення. Мовні й мовленнєві моделі у вітчизняних і зарубіжних системах ШІ.

2. *Системи автоматичної обробки тексту.* Автоматичний лінгвістичний аналіз тексту на природній мові: морфологічний, синтаксичний та семантичний. Практична реалізація принципів автоматичної обробки тексту.

3. *Лінгвістичні процесори, принципи їх побудови.* Моделі і методи построения языков представления знаний. Декларативне і процедурне представлення знань. Фрейми. Рекурсивні мережі переходів.

4. *Сучасні мовленнєві технології*. Комп'ютерна лексикографія. Історія і сучасний стан електронного словникарства. Основні напрямки машинного перекладу. Автоматичний аналіз і синтез письмового тексту комп'ютерними системами. Програми-коректори. Автоматичне анотування, реферування й індексування.

#### **Література:**

1. *Вербицкий В.Л., Грязнухина Т.А., Дарчук Н.П. и др.*. Автоматизация анализа научного текста. Киев: Наукова думка, 1984.- 247 с.

2. *В.Гладкий, И.А.Мельчук*. Элементы математической лингвистики. – М: Наука, 1969.- 192 с.

3. *Там само*, С.16

4. *В.І.Перебийніс*. Статистичні методи для лінгвістів: Навчальний посібник/Вінниця: Нова книга, 2001, С.12.

5. *В.Б.Черницький*. Комп'ютерна лексикографія: Навчальний посібник.- Миколаїв: НУК, 2004.-84с.