УДК 61-50

Щербина Ю.М., Висоцька В.А., Шестакевич Т.В.

МЕТОДИ ТА РОЗДІЛИ МАТЕМАТИЧНОЇ ЛІНГВІСТИКИ В СТРУКТУРІ КУРСУ «КОМП'ЮТЕРНА ЛІНГВІСТИКА»

В статті подано методи математичної лінгвістики, застосовні для вирішення теоретичних та практичних завдань прикладної лінгвістики та мовознавства. Описано методи мовознавства та математики для вивчення двох частин математичної лінгвістики – квантитативної та комбінаторної лінгвістики. Запропоновано структуру курсу «Математична лінгвістика».

Ключові слова: *математична лінгвістика, квантитативна лінгвістика, комбінаторна лінгвістика.*

В статье приведенны методы математической лингвистики, применимые для решения теоретических и практических заданий прикладной лингвистики и языкознания. Описаны методы языкознания и математики для изучения двух частей математической лингвистики – квантитативної и комбинаторной лингвистики. Предложена структура курса «Математическая лингвистика».

Ключевые слова: *математическая лингвистика, квантитативная лингвистика, комбинаторная лингвистика.*

In this paper theoretical and applied tasks, solvable with the help of methods of mathematical linguistics, are resulted. Linguistic and mathematical methods are described for the study of two sections of mathematical linguistics – computational and combinatorical linguistics. The structure of the course «Mathematical linguistics» is offered.

Key words: *mathematical linguistics, квантитативна linguistics, combinatorical linguistics*

ВСТУП

Повний та різносторонній опис мови та мовлення вимагає врахування і якісних, і кількісних характеристик лінгвістичних об'єктів. Використання кількісних методів у мовознавстві має об'єктивні підстави: очевидним є вплив кількісних характеристик на якісні (кількість голосних морфем, наприклад, впливає на звукову якість слів), можемо обчислити кількість абзаців, речень, слів, букв, звуків та інших кількісних характеристик тексту, частоти різних елементів мови у мовленні підлягають ймовірнісно-статистичним законам.

Прикладна лінгвістика виникла наприкінці 1920-х років, і з появою комп'ютерів у цій галузі активізуються дослідження в частині практичного застосування накопичених лінгвістичних знань, завдяки чому виникають і з середини XX століття стрімко розвиваються математична лінгвістика, комп'ютерна лінгвістика, машинний переклад тощо. Однією з перших в Україні фахівців з прикладної лінгвістики почав готувати Національний університет «Львівська політехніка» [6].

МАТЕМАТИЧНА ЛІНГВІСТИКА ЯК СКЛАДОВА ЧАСТИНА ПРИКЛАДНОЇ ЛІНГВІСТИКИ

Прикладна лінгвістика – це галузь мовознавства, об'єктом якої є розробка методів та засобів застосування у соціальній практиці людини результатів фундаментальних лінгвістичних досліджень, спрямованих на пізнання онтологічних властивостей мови [4]. Складовими прикладної лінгвістики є комп'ютерна (машинна, обчислювальна, інженерна) лінгвістика, структурна лінгвістика, математична (алгебраїчна, обчислювальна) лінгвістика, лінгводидактика (викладання рідної або іноземної мови), термінознавство, перекладознавство, контрастивна (співставна) лінгвістика, статистична лінгвістика і т. д.

Теоретичним підґрунтям прикладної лінгвістики вважають **математичну лінгвістику** – галузь науки на межі *мовознавства* та *математики*, яка вивчає можливості застосування математичних методів для опису та дослідження природних і деяких штучних мов, для пояснення лінгвістичних подій [4]. Підкреслюючи спільність поняттєвого апарату, математичну лінгвістику іноді розділяють на галузі мовознавства та математики [1, 2], а також відзначають, що в частині ***використання*** розроблених математичних моделей для опису будови природних мов математична лінгвістика належить до такої галузі досліджень, як *штучний інтелект*.

Відзначимо специфічні завдання мовознавства та прикладної лінгвістики зокрема [5], які з допомогою взаємодії відповідних математичних методів вирішує математична лінгвістика:

- Побудовані з допомогою обернених тригонометричних функцій моделі можна використовувати для опису структурних зсувів (наприклад, появи або зникнення лінгвістичних одиниць) в області лексикології, морфології, фонології, синтаксису та стилістики.
- У роботах з діахронічної лінгвістики часто наводяться таблиці, що містять кількісні дані про використання досліджуваного явища на різних етапах його історії. Такі таблиці показують динаміку лінгвістичного процесу, більш поглиблено аналізувати який можна з допомогою елементарних функцій, графічно, а також з використанням аналітичних моделей досліджуваного процесу.
- Класична глоттохронологія має на меті наближено датувати процес розходження діалектів та споріднених мов, а також кількісно оцінити ступінь їх спорідненості – для чого використовує статистичні методи.
- Методи математичної статистики також допоможуть у створенні словників (у тому числі частотних та статистичних), автоматичних словників, тезаурусів (лексикографія), у вирішенні завдань стилеметрії.
- Постійне розширення сфери діяльності сприяє неухильному збільшенню термінологічного словника людини, і для лексикографічної практики, а особливо під час побудови систем машинного перекладу та реферування, важливо мати прогноз кількісного росту термінології в різних областях знань. Таке прогнозування дало б можливість свідомого планування випуску та ефективного поповнення електронних політехнічних та галузевий словників та довідників.
- Для визначення швидкості зміни в мові та мовленні, а також їх кількісних оцінок, використовується поняття похідної.
- Тригонометричні функції та поняття границі дозволяють моделювати циклічний і стрибкоподібний характер лінгвістичних процесів.
- Можна також описати процеси приросту та накопичення нових мовних елементів – використовуючи теорію рядів та інтегрування.
- Приблизні оцінки вимірів обмежень, що накладаються на використання лінгвістичних одиниць системою та нормою мови, можна проводити з допомогою комбінаторики.
- Лінгвістичні дослідження в частині підготовки систем машинного перекладу та інформаційного пошуку вимагають, зокрема, знання об'єму вибірки, необхідного для забезпечення з заданою імовірністю появи певної лінгвістичної події хоча б раз.

Знаходження достатнього об'єму вибірки необхідне для визначення достовірності отриманих висновків.

Математична лінгвістика також бере участь у вирішенні таких універсальних проблем прикладної лінгвістики [1, 4, 6]:

- Розробка формальних моделей природних та штучних мов.
- Вирішення питань практичної транскрипції та транслітерації.
- Дешифрування невідомих писемностей.
- Усний та письмовий переклад, розробка систем автоматичного перекладу.
- Авторська та часова атрибуція твору.
- Вивчення процесів та створення правил утворення нових назв товарів, виробів тощо.
- Створення систем стенографії, систем письма для сліпих.
- Судова та кримінальна лінгвістика.
- Лінгводидактика:
  - навчання рідній та іноземній мові,
  - розробка відповідних лінгвістичних методик (навчання дітей та дорослих, навчання емігрантів тощо).
- Розробка раціональної та стабільної орфографії.
- Автоматичне розпізнавання та синтез мови.
- Розробка автоматизованих систем опрацювання текстової інформації:
  - автоматичне зняття полісемії,
  - автоматичне анотування,
  - індексування та реферування,
  - синтаксичний аналіз,
  - системи автоматичної компресії текстів,
  - автоматизований інтелектуальний аналіз тексту з метою виділення неявного змісту («читання між рядків»).
- Лінгвістичне забезпечення автоматизованих систем управління.
- Комп'ютеризація навчання та автоматизація викладання з допомогою комп'ютера:
  - автоматизовані системи навчання,
  - системи дистанційного навчання,
  - автоматизоване робоче місце перекладача та викладача.
- Вирішення задач прикладного термінознавства:
  - систематизація та автоматизація терміносистем певних галузей, урахування динамічності їх розвитку,
  - укладання термінологічних словників,
  - стандартизація й уніфікація науково-технічної термінології тощо.
- Створення автоматизованих редакційно-видавничих систем.

ПРЕДМЕТ МАТЕМАТИЧНОЇ ЛІНГВІСТКИ. СТРУКТУРА КУРСУ

Для вирішення зазначених завдань і проблем математична лінгвістика спирається та використовує специфічні методи розділів мовознавства (фонетика та фонологія, морфеміка та словотвір, морфологія, лексика та лексикографія, синтаксис, стилістика і т.д.) та математики (теорія множин, теорія алгоритмів, математична логіка, математична статистика, математичний аналіз, комбінаторика, теорія імовірності, теорія інформації і т. д.) [1, 4].

У математичній лінгвістиці умовно виділяють дві частини: квантитативну та обчислювальну лінгвістику [5]. Структуру курсу «Математична лінгвістика» можна представити у вигляді схеми на рис. 1.

Перша частина – **квантитативна лінгвістика** (обчислювальна лінгвістика, лінгвометрія) – вивчає та описує лінгвістичні явища за допомогою методів «кількісної» математики

(комбінаторика, теорія імовірності, математична статистика, математичний аналіз, теорія інформації тощо).

Для оволодіння основними методами квантитативної лінгвістики пропонується вивчення таких тем [1, 5]:

- ***Комбінаторні методи у лінгвістиці*** *(основні формули комбінаторики та їх застосування).*
- ***Елементарні та складні лінгвістичні події*** *(операції над лінгвістичними подіями, імовірність лінгвістичних подій, апріорні та апостеріорні оцінки імовірності лінгвістичної події).*
- ***Незалежні лінгвістичні випробування в тексті*** *(проста, поліноміальна та пуассонівська схеми).*
- ***Теорія інформації*** *(інформаційні виміри кодування інформації, розрахунок швидкості передачі інформації та пропускної здатності інформаційних каналів зв'язку, визначення надлишковості повідомлень, оптимальне кодування).*
- ***Основи статистичної лінгвістики*** *(основні поняття, параметричні та непараметричні методи досліджень, лінгвометрія).*

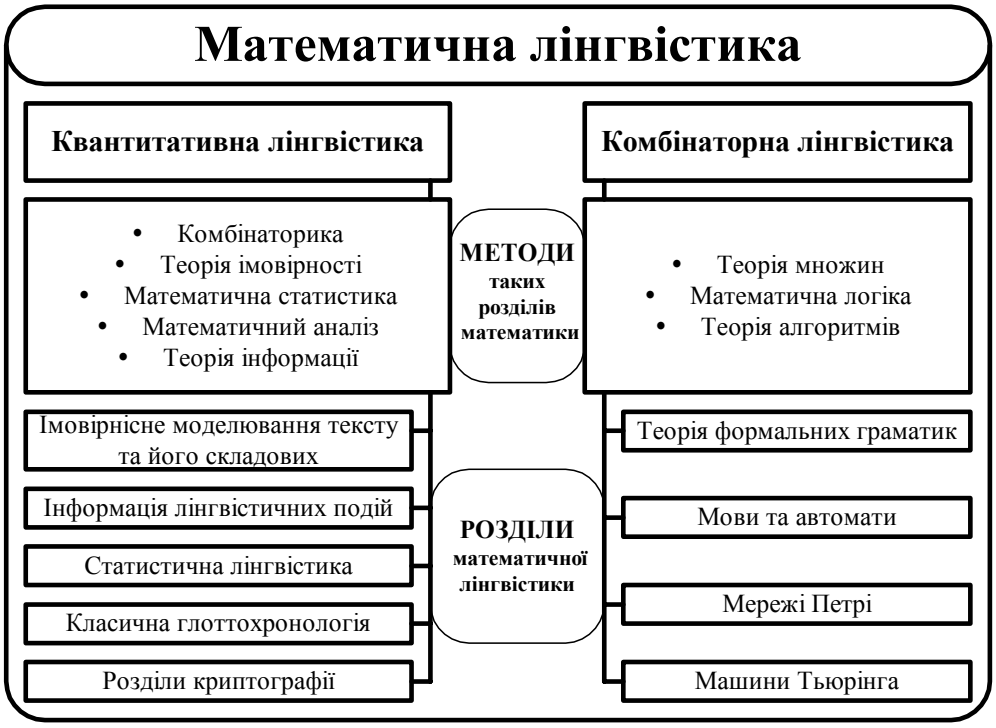

Рис. 1. Схематична структура курсу «Математична лінгвістика»

Оволодіння зазначеними методами дозволяє вивчати такі основні розділи квантитативної лінгвістики:

- Імовірнісне моделювання тексту та його складових.
- Інформація лінгвістичних подій.
- Статистична лінгвістика.
- Класична глоттохронологія.
- Розділи криптографії.

Друга частина математичної лінгвістики – це **комбінаторна лінгвістика**, яка вивчає та описує лінгвістичні явища за допомогою методів «некількісної» математики (теорія множин, математична логіка, теорія алгоритмів).

Для оволодіння основними методами комбінаторної лінгвістики пропонується вивчення таких тем [3]:

- ***Основні поняття теорії формальних граматик*** *(розпізнавальні та генеративні граматики).*
- ***Теорія автоматів*** *(автомати скінченні з виходом та без, недетермінований автомат, аналіз скінченних автоматів, взаємозв'язок між мовами, автоматами та граматиками).*
- ***Теорія алгоритмів*** *(алгоритми комп'ютерної математики, алгоритми ідентифікації текстових рядків і скінченні автомати).*
- ***Теорія обчислень і математична логіка*** *(універсальні обчислювальні моделі – машини Тьюрінга).*

Після вивчення зазначених методів можна перейти до вивчення основних розділів комбінаторної лінгвістики:

- Теорія формальних граматик.
- Мови та автомати.
- Мережі Петрі.
- Машини Тьюрінга.

Застосування методів математичної лінгвістики ілюструється прикладами, містить завдання кількох рівнів складності; кожен розділ присвячений вирішенню певної задачі мовознавства (прикладної лінгвістики) і включає комп'ютерний практикум. З метою автоматизації навчального процесу ведеться розробка системи незалежного тестового контролю знань, яка передбачає автоматичне генерування завдань до вивченої теми та розділу курсу, автоматичну перевірку та оцінювання відповідей. Завдання формуються на основі опрацювання прізвища та імені студента, номера його студентського квитка та (або) залікової, дати складання тесту і т. д.

Запропонована схема викладення матеріалу ґрунтується на практичному досвіді ([3, 7], пройшли апробацію методичні вказівки до виконання лабораторних робіт з курсу «Математична лінгвістика» для студентів базового напрямку «Філологія» № 1-5, видані у 2006 р. та № 1-8, видані у 2007 р., укладені кандидатом фізико-математичних наук, доцентом кафедри дискретного аналізу та інтелектуальних систем Львівського національного університету ім. І. Франка Щербиною Ю.М. та асистентами кафедри інформаційних систем та мереж Національного університету «Львівська політехніка» Висоцькою В.А. та Шестакевич Т.В.) і дозволяє охопити та проілюструвати ті методи й розділи математичної лінгвістики, що необхідні для здобуття фаху в галузі прикладної лінгвістики. Запропонований додатковий матеріал для поглибленого вивчення розділів допоможе скерувати зацікавленого студента у власних наукових дослідженнях та практичних розробках з використанням сучасних інформаційних технологій.

ВИСНОВОК

Існування тісного взаємного зв'язку між кількісними та якісними характеристиками мови для їх опису та вивчення передбачає спільне застосування математичного апарату квантитативної та комбінаторної лінгвістики. Основні методи та розділи цих складових математичної лінгвістики структуровано у відповідному курсі. Застосування опанованого матеріалу дозволяє вирішувати широке коло завдань прикладної лінгвістики, що підкреслює вагоме місце математичної лінгвістики у вирішенні проблем мовознавства в цілому.

ЛІТЕРАТУРА

1. Бук С.Н. Основи статистичної лінгвістики: Навчально-методичний посібник / Відп. ред. проф. Ф.С. Бацевич. – Л. : Видавничий центр ЛНУ ім. І. Франка, 2008. – 124 с.
2. Математична лінгвістика // Великий тлумачний словник сучасної української мови / Уклад. та голов. ред. В.Т. Бусел. – К.; Ірпінь: ВТД «Перун», 2004. – С. 513.
3. Нікольський Ю.В. Дискретна математика: підруч. для студ. ВНЗ / Ю.В. Нікольський, В.В. Пасічник,

Ю.М. Щербина. – К. : Видавнича група BHV, 2007. – 368 с: іл.
4. Перебийніс В.С. Математична лінгвістика // Українська мова: Енциклопедія / Редкол.: В.М. Русанівський та інші. – К. : Українська енциклопедія, 2000, – С. 302-287.
5. Пиотровский Р.Г. Математическая лингвистика: Учебное пособие / Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А. – М.: Высшая школа, 1977. – 384 с.
6. Шульгіна В.І. Прикладна лінгвістика як навчальна та наукова спеціальність [Електронний ресурс]. – Режим доступу до статті: http://www.bdpu.org/scientific_published/akt_probl_sl_filol-12/1.doc.
7. Щербина Ю.М. Предмет математичної лінгвістики // Вісник НУ «Львівська політехніка», серія «Інформаційні системи та мережі». – 2002. – № 464. – С. 340-349.

Рецензенти: д.т.н., проф. Данілов В.Я.,
д.т.н., проф. Фісун М.Т.

© Щербина Ю.М., Висоцька В.А.,
Шестакевич Т.В., 2009

Стаття надійшла до редакції 02.09.09