

Снігур О.М.

кандидат педагогічних наук, доцент,

Національний педагогічний університет імені М.П. Драгоманова

ОБЛАСТІ ЗАСТОСУВАННЯ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ В ЛІНГВІСТИЦІ

Нині в умовах стрімкого науково-технічного прогресу будь-який спеціаліст має вільно орієнтуватися у світовому інформаційному просторі, має вміти знаходити, здобувати, обробляти, зберігати інформацію за допомогою сучасних інформаційно-комунікаційних технологій, комп'ютерних систем і мереж. Не є винятком і лінгвісти.

Лінгвістика, також мовознавство – наука, що вивчає мову в усій складності її прояву; наука про мову взагалі й окремі мови світу як індивідуальних її представників. Це гуманітарна наука, яка є розділом культурології (нарівні з мистецтвознавством і літературознавством) і філології (нарівні з літературознавством), а також галуззю семіотики [3].

У гносеологічному аспекті мовознавство включає: спостереження; реєстрування й описування фактів мовлення; висування гіпотез, які пояснюють ці факти; формулювання гіпотез у вигляді теорій і моделей, які описують мову; їх експериментальну перевірку і спростування; прогнозування мовної поведінки.

Щоб визначити розділ лінгвістики, який найбільш тісно пов'язаний з використанням інформаційних технологій, доцільно звернутися до розподілу лінгвістики на теоретичну та прикладну.

Теоретична лінгвістика є розділом, ядром загального мовознавства. Вона досліджує мовні закони і формулює їх як теорії. Теоретична лінгвістика може бути емпіричною або нормативною. Емпірична лінгвістика описує реальну мову, нормативна ж вказує правильність вимови і написання слів.

Теоретична лінгвістика формує саме уявлення про те, які властивості мови є суттєвими, а які – ні. Створювані в теоретичній

лінгвістиці концептуальні моделі мови не просто описують спостережувані факти, а й претендують на їх пояснення. Іншими словами, класифікації мовних фактів і концептуальні моделі теоретичної лінгвістики претендують на опис того, як дійсно влаштовано мову. Крім розробки лінгвістичних моделей теоретична лінгвістика займається пошуком мовних універсалій, тобто спільних ознак, властивих усім природним мовам світу [5].

Прикладна лінгвістика – розділ мовознавства, пов'язаний із практичним розв'язанням питань, які пов'язані з вивченням мови. До традиційних проблем прикладної лінгвістики належать укладання словників, розробка алфавітів і систем письма, транскрипції усного мовлення та транслітерації іншомовних слів, лінгвістичне обґрунтування викладання рідної та іноземних мов, переклад з однієї мови на іншу, стандартизація й уніфікація науково-технічної термінології, укладання спеціальних лінгвістичних довідників, створення штучних мов, удосконалення орфографії і пунктуації, мовна культура тощо. До нових проблем прикладної лінгвістики належать автоматичний (машинний) переклад, створення інформаційних мов, автоматичне анотування та індексування документів, лінгвістичне забезпечення роботи інформаційних систем, автоматична переробка текстової інформації, лінгвістичне забезпечення автоматичних систем управління (АСУ), автоматичний аналіз (розпізнавання) й автоматичний синтез тексту тощо. Прикладна лінгвістика тісно пов'язана з математикою, кібернетикою та інформатикою [4].

Термін «інформаційні технології» визначається як сукупність методів і технічних засобів збирання, організації, збереження, опрацювання, передачі й подання інформації, що розширює знання людей і розвиває їхні можливості щодо керування технічними і соціальними проблемами [8, с. 166].

Конкретизуючи визначення поняття «інформаційні технології» по відношенню до лінгвістики, можна сказати, що *інформаційні технології в лінгвістиці* – це сукупність методів і технічних засобів

збирання, організації, збереження, опрацювання, передачі й подання інформації про мову і закони її функціонування за допомогою комп'ютера та інших електронних засобів.

Можна відзначити, що поняття «інформаційні технології» в лінгвістиці відноситься в основному до задач прикладної лінгвістики. До них можна віднести: створення систем штучного інтелекту; створення систем автоматичного перекладу; створення систем автоматичного анотування і реферування текстів; створення систем породження текстів; створення систем навчання мови; створення систем розуміння усного мовлення; створення систем генерації мовлення; створення автоматизованих інформаційно-пошукових систем; створення систем атрибуції та дешифровки анонімних і псевдоанонімних текстів; розробка баз даних (словників, карток, каталогів, реєстрів тощо); розробка різного типу автоматичних словників; розробка систем передачі інформації в мережі Інтернет.

Виділяють такі області застосування інформаційно-комунікаційних технологій в лінгвістиці: автоматичний аналіз та синтез мовлення; автоматичне розпізнавання тексту; автоматичне анотування та реферування тексту; автоматичний аналіз та синтез тексту.

Автоматичний синтез тексту (АС), операція, в якій за заданою граматичною і семантичною інформацією будується той, що містить цю інформацію текст на природній мові; операція виконується по деякому алгоритму відповідно до заздалегідь розробленого опису даної мови. АС підрозділяється на три етапи: 1) семантичний – перехід від смислового запису фрази до її синтаксичної структури; 2) синтаксичний – перехід від синтаксичної структури фрази до ланцюжка лексико-граматичних характеристик словоформ, що представляє фразу; 3) лексико-морфологічний – перехід від лексико-граматичної характеристики до реальної словоформи. АС – необхідний етап в різних видах автоматичної обробки текстів, зокрема при машинному перекладі [1].

Зворотна операція називається автоматичним аналізом тексту. Автоматичний аналіз друкованого тексту, також як і аналіз усної мови, починається з його введення в комп'ютер. Для введення інформації в комп'ютер використовуються спеціальні пристрої – клавіатура, миша та ін., але найбільш зручним інструментом для введення великої кількості друкованих текстів є сканер. Щоб комп'ютер зміг перевести графічне (растрове) зображення символів в текстову форму, при якій у кожного символу є свій двійковий код (наприклад, в системі кодувань ASCII), потрібна програма автоматичного розпізнавання символів (англ. OCR = Optical Character Recognition).

Автоматичний аналіз тексту включає ряд вельми складних операцій, які комп'ютер виконує над текстом відповідно до заданого алгоритму. При *автоматичному аналізі* текст послідовно перетворюється в його лексико-морфологічні, синтаксичні та семантичні уявлення, зрозумілі комп'ютеру. Зворотний процес перетворення лексико-морфологічних, синтаксичних і семантичних комп'ютерних уявлень в текст на природній мові називається автоматичним синтезом тексту [2].

Автоматичний аналіз тексту включає ряд етапів: граматичний аналіз; морфологічний аналіз; синтаксичний аналіз; семантичний аналіз.

Граматичний аналіз визначається також як токенізація (від англ. token – окреме слово, фраза або будь-який інший значущий елемент тексту). Формальними сигналами кордонів текстових елементів виступають роздільники різного роду: прогалини, що позначають кордони між словами, великі літери та розділові знаки, які позначають межі між реченнями і складовими частинами речень, абзацні відступи, що позначають межі між пов'язаними за змістом групами пропозицій тощо [6, с. 167-183].

При *морфологічному аналізі* кожне використане в тексті слово зводиться до його початковій форми і визначається набір морфологічних характеристик текстової форми слова: частина мови; рід, число і відмінок для іменників, число і особа для дієслів тощо.

При синтаксичному аналізі необхідно визначити ролі слів у реченні і їх зв'язку між собою. Результатом цього етапу автоматичного аналізу є уявлення синтаксичних зв'язків кожного речення у вигляді моделей, наприклад у вигляді дерева залежностей.

Семантичний аналіз являє собою, мабуть, найбільш складний напрямок автоматичного аналізу тексту. В цьому випадку потрібне встановлення семантичних відносин між словами в тексті, об'єднання різних мовних визначень, що відносяться до одного і того ж поняття [7, с. 134-166].

Процес перекладу тексту – це складний інформаційний процес, який не можна звести до простої заміни слів однієї мови словами іншої. Необхідно проводити аналіз структури, враховувати синтаксичну, а найголовніше – семантичну будову тексту.

Список використаних джерел:

1. Всі слова. [Електронний ресурс]. – Режим доступу: http://vseslova.com.ua/word/%D0%90%D0%B2%D1%82%D0%BE%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%BD%D0%B8%D0%B9_%D1%81%D0%B8%D0%BD%D1%82%D0%B5%D0%B7-835u.
2. Всі слова. [Електронний ресурс]. – Режим доступу: http://vseslova.com.ua/word/%D0%90%D0%B2%D1%82%D0%BE%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%BD%D0%B8%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D0%B7-831u.
3. Вікіпедія. [Електронний ресурс]. – Режим доступу: <https://uk.wikipedia.org/wiki/%D0%9C%D0%BE%D0%B2%D0%BE%D0%B7%D0%BD%D0%B0%D0%B2%D1%81%D1%82%D0%B2%D0%BE>.
4. Вікіпедія. [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/%D0%9F%D1%80%D0%B8%D0%BA%D0%BB%D0%B0%D0%B4%D0%BD%D0%B0_%D0%BB%D1%96%D0%BD%D0%B3%D0%B2%D1%96%D1%81%D1%82%D0%B8%D0%BA%D0%B0.
5. Вікіпедія. [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/%D0%A2%D0%B5%D0%BE%D1%80%D0%B5%D1%82%D0%B8%D1%87%D0%BD%D0%B0_%D0%BB%D1%96%D0%BD%D0%B3%D0%B2%D1%96%D1%81%D1%82%D0%B8%D0%BA%D0%B0.

6. Соколова Е.Г., Кудряшева И.М. Компонент синтаксического анализа в системе ФРАП. Актуальные вопросы практической реализации систем автоматического перевода, 1982, Москва, с. 167-183.

7. Леонтьева Н.Н., Никогосов С.Л. Система ФРАП как информационная система. Актуальные вопросы практической реализации систем автоматического перевода, 1982, Москва, с. 134-166.

8. Освітні технології: навч.-метод. посіб. / О.М. Пехота, А.З. Кіктенко, О.М. Любарська та ін.; За заг. ред. О.М. Пехоті. – К.: А.С.К., 2001. – 256 с. – Бібліогр. в кінці розд.