

Когут Владислав Сергійович, здобувач вищої освіти факультету міжнародних відносин
Хмельницький національний університет, Україна

Науковий керівник: Тарасова Ольга Володимирівна, канд. пед. наук, доцент,
доцент кафедри практики іноземної мови та методики викладання
Хмельницький національний університет, Україна

КОМП'ЮТЕРНА ЛІНГВІСТИКА ЯК ГАЛУЗЬ МОВОЗНАВСТВА

Комп'ютери посідають важливе місце в житті сучасного суспільства. Комп'ютерна лінгвістика – це порівняно молода наука, яка займається проблемами використання природної мови в системах автоматичної обробки інформації. За основу дослідження було використано праці відомих вітчизняних лінгвістів Марчук Ю. (Y. Marchuk), Шемякін Ю. (Shemyakin Y.), Боярський К. (Boyarsky K.) та ін.

Розвиток теорії і практики мовознавства вимагає використання точних і об'єктивних методів для аналізу мови й тексту. Разом з тим, використання математичних прийомів при систематизації, вимірюванні та інтерпретації лінгвістичного матеріалу у поєднанні з якісним аналізом результатів дозволяє мовознавцям ретельно досліджувати особливості будови мови й утворення тексту. [Белоногов 2004, 192 с.]

Контакти мовознавства з іншими науками (акустикою, фізіологією вищої нервової діяльності, кібернетикою та обчислювальною технікою тощо) постійно розширюються і можуть існувати виключно за умов використання математичної мови, яка має високий ступінь узагальнення та універсальності для різних галузей знань. В існуючих системах машинного перекладу, автоматичного анотування, людино-машинного діалогу будь-яке повідомлення природною мовою перекодовується в математичну мову комп'ютера.

Під терміном «*комп'ютерна лінгвістика*» розуміємо широкий спектр використання комп'ютерних інструментів – програм, технологій організації та обробки даних – для моделювання функціонування мови в певних умовах, ситуаціях, а також сфера застосування комп'ютерних моделей мови не лише в лінгвістиці, а й у суміжних із нею дисциплінах. Сфера комп'ютерної лінгвістики охоплює практично все, що пов'язано з використанням комп'ютерів у мовознавстві. Термін «комп'ютерна лінгвістика» орієнтує на використання комп'ютерів для вирішення різноманітних наукових і практичних завдань, пов'язаних із мовою. Комп'ютерна лінгвістика посідає проміжне місце між прикладною лінгвістикою та інформатикою. [Марчук 2007]

Застосування математичних методів у мовознавстві має на меті замінити звичайно дифузну, інтуїтивно сформульовану і таку, що не має повного розв'язку, лінгвістичну задачу однією або декількома простішими, логічно сформульованими математичними задачами. Таке розділення складної лінгвістичної проблеми на простіші задачі, які алгоритмізуються, називають математичною експлікацією лінгвістичного об'єкту або явища. [Шемякін 2001, 81 с.]

Автоматична обробка природної мови – напрям комп'ютерної лінгвістики, який передбачає створення, перетворення й аналіз текстів із застосуванням природної або штучної мов, результатом чого може бути формування машинних фондів національних мов, автоматичних словників, термінологічних банків, комп'ютерних картотек, баз даних, комп'ютерних граматик тощо. Створення текстів відбувається у процесі автоматичного синтезу на підставі семантичного, синтаксичного й лексемно-морфологічного представлення вихідної інформації. Перетворення текстів здійснюється при автоматизованому редагуванні – внесенні виправлень і доповнень, форматуванні – членуванні тексту, нумерації сторінок; при лексикографічній обробці – уведенні щодо

кожного слова словникової дефініції й іншої потрібної інформації або підготовці на основі текстів автоматизованих лексикографічних систем; реферуванні – скороченні вихідного тексту відповідно до програми. Аналіз текстів передбачає членування тексту на одиниці, доступні комп'ютерній лексикографічній обробці, виокремлення ключових слів, персоналій, термінів. Аналіз і синтез текстів здійснюється на базі лінгвістичних процесорів – програмно-лінгвістичних комплексів багаторівневого типу, орієнтованих на граматичний, семантичний або когнітивний аналіз текстової інформації та діалог із користувачем. Процесори останнього покоління спрямовані на самонавчання шляхом гіпотез відносно явищ і категорій мови. Аналіз є основою машинного перекладу з однієї мови на іншу, а також перетворення текстів природною мовою на мови програмування, комп'ютерні кодові системи.

Комп'ютерна лінгвістика тісно пов'язана з дослідженнями у галузі штучного інтелекту. Штучний інтелект – розділ комп'ютерології та інформатики (computer science), розробляє «розумні» комп'ютерні системи, тобто системи, які виявляють певні характеристики в людській поведінці (розуміння мови, навчання, міркування, вирішення завдань тощо, а також етичні аспекти). Вивчення розуміння людської мови комп'ютером неможливе без заглиблення в теоретичні засади комп'ютерної лінгвістики. [Белоногов 2004, 169 с.]

Перші експерименти з машинному перекладу, які підтвердили можливість його реалізації, було проведено 1954 р. в Джорджтаунському університеті (Вашингтон, США). Згодом в промислово розвинених країнах світу розпочато дослідження і розробки, спрямовані на створення систем машинного перекладу. З досягнень комп'ютерної лінгвістики слід виділити, передусім, систему під назвою «RETRANS». Перша промислова версія цієї системою була розроблена 1993 року. Згодом було створено ще кілька систем. До 1998 р. вони були призначені до роботи у середовищі ОС MS DOS, а в 1998-2000 рр. були адаптовані для операційних систем Windows9x/NT і різних варіантів UNIX. У 2001 р. було розроблено нову версію системи для фразеологічного машинного перекладу «RETRANS».

Крім основних політематичних машинних словників до складу системи RETRANS входять дванадцять додаткових тематичних словників, що включають лексику споріднених тематик. Ці словники містять словникові статті тільки тих слів і словосполучень із політематичних словників, які в рамках кожної тематичної групи мають пріоритетні перекладні еквіваленти, відмінні від пріоритетних перекладних еквівалентів основних словників. У додаткових тематичних словниках лексичні одиниці (слова і словосполучення) можуть мати не більше двох перекладних еквівалентів.

Засоби автоматизованого поповнення та налаштування машинних словників, створені на початкових етапах розробки системи RETRANS, відіграли велику роль в її подальшому розвитку. Але досвід практичної експлуатації цієї системи виявив необхідність внесення певних змін до структури словників. Таким чином було додано словники користувачів. Словник користувача дає можливість використовувати при перекладі раніше накопичені словникові масиви. На відміну від основного політематичного словника і додаткових тематичних словників, вміст яких може змінюватися тільки розробниками системи машинного перекладу, словник користувача може створюватись та змінюватись користувачем. Користувач системи RETRANS може створювати практично необмежену кількість словників з різних тем, але в процесі перекладу конкретного тексту використовується тільки один словник. За необхідності, можна об'єднати кілька словників в один. В інтерактивному режимі роботи користувач має можливість переглядати всі варіанти перекладу слів і словосполучень, які містяться в словниках, змінювати їх пріоритет, додавати нові перекладні еквіваленти, а також виконувати інші операції над проміжними результатами перекладу. [Боярский 2013, 72 с.]

Висновки. Підсумовуючи зазначене вище, можемо зробити висновок, що комп'ютерна лінгвістика – це галузь знань, пов'язана з вирішенням завдань автоматичної обробки інформації, представленої на природній мові. У системах автоматичної обробки текстової інформації важливу роль відіграє комп'ютерна лінгвістика. Комп'ютерна лінгвістика як наука має значний потенціал, оскільки безперервно розвивається. Водночас, як наука комп'ютерна лінгвістика досить молода, тому потребує подальших досліджень.

Список використаних джерел:

1. Белоногов Г. Г. (2004) *Комп'ютерна лінгвістика і перспективні інформаційні технології*. Москва. 192 с.
2. Большакова Е.И. (2011) *Автоматическая обработка текстов на естественном языке и компьютерная лингвистика*. МИЭМ. 272 с.
3. Боярский К. К. (2013) *Введение в компьютерную лингвистику*. Учебное пособие. СПб: НИУ ИТМО. 72 с.
4. Марчук Ю.Н. (2007) *Компьютерная лингвистика*. Восток-Запад 81 с.
5. Шемякин Ю.И. (2001) *Начала Компьютерной лингвистики*. Москва. 81 с.