

Василь Маценко

Огляд програмних засобів прикладної лінгвістики

*Чернівецький національний університет імені Юрія Федьковича,
Чернівці, Україна
E-mail: vgm2006@mail.ru*

Мета цієї роботи - описати найпоширеніші засоби комп'ютерної лінгвістики. Найбільш традиційною формою подання лексичної інформації є словники. Електронні словники - це словники в спеціальному машинному форматі, що призначені для використання на ЕОМ. Існують різні типи словників (тлумачні, двомовні, орфографічні, морфологічні, тезауруси, етимологічні, синонімічні, антонімічні і т.д.). Онлайн-словники - це електронні словники, що розміщені в Інтернеті. Їх розміщують у себе і пошукові портали (slovari.yandex.ru, dict.rambler.ru, lingvo.mail.ru, translate.google.com.ua, translate.meta.ua). Інтернет ресурс <http://e2u.org.ua> містить цілу низку англо-українських і україно-англійських словників. Велика колекція німецьких словників розміщена на сайті <http://www.duden.de>, французьких - на сайті <http://www.larousse.fr>. "Словники України" - комп'ютерна програма надання користувачу лексикографічної інформації про слова української мови, вона включає в себе більше 40 тлумачних словників української мови. Он-лайн версія цієї програми знаходиться на сайті <http://lcorp.ulif.org.ua/dictua> (український лінгвістичний портал).

Єдиним способом справитись з необхідністю перекладу в наш час є застосування новітніх технологій. Першим проявом інформаційних технологій у житті перекладача стали електронні словники. Наступним кроком в розвитку перекладацьких інформаційних технологій стали програми машинного перекладу, які значно полегшили переклад тексту. Нині в практиці перекладу використовується два види перекладу: повністю автоматичний переклад, коли програма здійснює повний переклад тексту, та автоматизований переклад, коли здійснюється переклад окремих слів, фраз, тобто частково. Найбільш часто використовуються допоміжні програми автоматизованого перекладу.

Лінгвістичний корпус - це великий, поданий в електронній формі, уніфікований, структурований, розмічений масив мовних даних,

що призначений для розв'язування конкретних лінгвістичних задач. Британський національний корпус - це один з великих еталонних корпусів (сайт <http://corpus.byu.edu>). В ньому міститься 100 млн. слів. Всі тексти цього корпусу сегментовані за реченнями. Корпус текстів української мови побудований як інформаційно-довідкова система. Він розміщений на сайті <http://www.mova.info> і містить 13 млн. словоформ. Списки національних корпусів можна знайти на сайті <http://www.ruscorpora.ru>.

Ще одне застосування комп'ютерної лінгвістики - це інформаційний пошук і зв'язані з ними задачі індексування, реферування, класифікації, рубрикації документів. Найбільш відомою в наші дні системою автоматичного реферування текстів будь-якого обсягу і будь-якої складності є система Лібретто. Дещо спрощений варіант реферату може видавати система Text Analist.

До прикладної лінгвістики належить розділ теорії і методики вивчення іноземних мов (лінгводидактика). Основна особливість сучасного етапу розвитку комп'ютерної лінгводидактики - це використання різних комп'ютерних програм з метою створення інтегрованого середовища навчання. Прикладами лінгвістичних тренажерів для початкового вивчення іноземних мов є такі комп'ютерні фонетичні системи: Reward (комп'ютерна версія Оксфордського підручника для вивчення англійської мови), професор Хіггінс - англійська без акценту, Bridge to English, Learn to Speak English, Talk to Me, репетитор English та ін.

Існують програми аналізу та синтезу тексту: (<http://asknet.ru>), програми синтаксичного аналізу (<http://sz.ru/parser>), програми підтримки розмови зі співбесідником в режимі реального часу, тобто автоматичного синтезу діалогу (мови) (<http://www.web4design.ru>, <http://orphale.icp.grenet.fr>). Є програми, що читають текст різними голосами на різних мовах. Існують комп'ютерні програми автоматичної генерації текстів, побудови спеціальних документів, моделювання сюжету, наприклад, програма TALE моделює сюжет казок. Наведені приклади програм свідчать про багатогранність програмного забезпечення прикладної лінгвістики та необхідність їх подальшого удосконалення.