



КОРПУСНА ЛІНГВІСТИКА — галузь мовознавства, предметом дослідження якої є принципи та методи формування корпусів текстів, а також розроблення комп'ютерних систем для їхнього опрацювання. Корпус розглядають як модель мовної системи, реалізовану в текстах різних функціонал. стилів, структури, тематики, призначення, території та часу їх створення. Залежно від призначення виділяють фундам. і пошукові (дослідн.) корпуси, для роботи з якими створюють спец. текстові процесори — системи опрацювання інформації про мовні одиниці морфол., синтаксич., логіко-семантич. структур. рівнів тексту. Такі системи зорієнтовані на анотацію (розмічення) тексту, яка передбачає формаліз. виділення у ньому мовних одиниць певних типів. Завдяки наявності автоматиз. засобів аналізу інформації корпуси придатні для виконання за допомогою комп'ютера теор. і практ. завдань мовознавства. За повнотою подання текстів, їхнім структуруванням та наявністю засобів автоматиз. аналізу корпуси відрізняють від електрон. б-к, картотек і баз цитат. Сучасні нац. корпуси мають здебільшого фундам. характер, їхній обсяг сягає 200 і більше млн слововживань. Такими є, напр., Британ. нац. корпус (British National Corpus), Нац. корпус рус. язика, Narodowy Korpus Języka Polskiego або корпус текстів (Frantext) для укладення словника «Trésor de la langue française» («Скарбниця французької мови»). Корпуси текстів різних типів та призначення мають естон., італ., китай., нім., польс., рос., серб., словен., словац., турец., укр. і чес. мови. В Україні перші корпуси україномовних текстів з'явилися у 1990-х рр., зокрема корпус наук.-реферат. та публіцист. текстів, на базі якого співроб. відділу структурно-матем. лінгвістики (нині у складі Інституту української мови

НАНУ, Київ) створили систему автомат. орфогр. контролю тексту «Рута» та систему машинного рос.-укр. і укр.-рос. перекладу «Плай», а також корпус укр. худож., публіцист. і наук. текстів, сформов. у лаб. комп'ютер. лінгвістики Київ. університету. Над укладанням фундам. нац. корпусів україномовних текстів працюють науковці Укр. мовно-інформ. фонду НАНУ (Київ) та Інституту української мови НАНУ. Від 2009 при Міжнар. комітеті славистів працює Комісія з корпус. лінгвістики на чолі з М. Лазинським (Варшава), яка координує діяльність розробників корпусів слов'ян. мов. Від 1996 виходить «International Journal of Corpus Linguistics».

Рекомендована література

1. Фрэнсис У. Н. Проблемы формирования и машинного представления большого корпуса текстов / Пер. с англ. // Новое в зарубеж. лингвистике. Вып. 14. Проблемы и методы лексикографии. Москва, 1983;
2. Автоматизация анализа научного текста. К., 1984;
3. Баранов А. Н. Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема // Русистика сегодня. 1998. № 1-2;
4. Корпусна лінгвістика. К., 2005;
5. Демська-Кульчицька О. М. Основи національного корпусу української мови. К., 2005;
6. Резникова Т. И. Славянская корпусная лингвистика: современное состояние ресурсов // Нац. корпус рус. языка: 2006-08. Новые результаты и перспективы. С.-Петербург, 2009.

Є. А. Карпіловська

Бібліографічний опис:

Корпусна лінгвістика / Є. А. Карпіловська // Енциклопедія Сучасної України [Електронний ресурс] / Редкол.: І. М. Дзюба, А. І. Жуковський, М. Г. Железняк [та ін.] ; НАН України, НТШ. — К. : Інститут енциклопедичних досліджень НАН України, 2014. — Режим доступу: <https://esu.com.ua/article-5251>

2001-2025 © Ця енциклопедична стаття захищена авторським правом згідно з чинним законодавством України (докладніше).