

ТЕТЯНА МАСЛОВА
Національний технічний
університет України
«КПІ імені Ігоря Сікорського»

ІНСТРУМЕНТИ КОРПУСНОЇ ЛІНГВІСТИКИ У НАВЧАННІ І ДОСЛІДЖЕННІ ІНОЗЕМНИХ МОВ

Корпусна лінгвістика – це розділ прикладного мовознавства, що займається виявленням закономірностей функціонування мови шляхом аналізу лінгвістичних корпусів (баз даних письмових та усних аутентичних текстів) за допомогою сучасних комп'ютерних технологій.

Ресурси корпусної лінгвістики знаходять застосування у різних галузях, наприклад, для машинної обробки мовного матеріалу в комп'ютерній лінгвістиці, для створення репрезентативних колекцій творів конкретного автора або жанру в корпусній стилістиці, а також для перевірки лінгвістичних теорій та у методиці викладання іноземних мов [2; 4].

Перший великий корпус текстів на машинному носії був створений у 60-х роках ХХ ст. у Брауновському університеті (США) з

метою дослідження лінгвістичних особливостей американського варіанту англійської мови. Корпус містив майже 500 тисяч друкованих текстів загальним обсягом біля 1 млн. слів, що належали до найбільш поширених жанрів американської прози – газетні статті, рецензії, художні твори, наукові роботи, релігійна література, фантастика, біографії, гумористичні оповідання і т.д. У подальшому на основі критеріїв відбору текстів Браунівського корпусу були створені інші корпуси, зокрема його британський аналог – Ланкастерсько-Осло-Бергенський корпус, що дозволив проводити порівняльне вивчення американського і британського варіантів англійської мови. В кінці 80-х років був укладений Британський національний корпус обсягом 100 млн. слів, який був доступний через Інтернет і задав нові стандарти, за яким були створені лінгвістичні корпуси багатьох європейських мов. Наразі багатомовний корпус міжнародної організації «European Corpus Initiative (ECI)» складається з 50 корпусів різних мов, кожен обсягом від 12 тисяч до 5 млн. слів [2; 3].

Практичне застосування результатів корпусної лінгвістики можна спостерігати перш за все в галузі лексикографії. В результаті комп'ютерної обробки великих масивів тексту створюються нові словники на кшталт COBUILD (Collins Birmingham University International Language Database), Oxford, Longman, Macmillan, Webster та ін., які на основі матеріалу своїх власних, закритих для загального доступу, корпусів реальних текстів, фіксують найсучасніші мовні тенденції, частотність вживання слів, зміни у лексичному складі мови тощо, що звичайно сприяє їхньому комерційному успіху і попиту серед тих, хто вивчає англійську мову. Наприклад, база даних «Collins Corpus» нараховує 4,5 млрд. слів і містить матеріал з англійських веб-сайтів, газет, журналів, книжок, а також з радіо- і телепрограм та повсякденного мовлення, причому нові дані надходять щомісяця, тому словники постійно оновлюються, щоб надавати найактуальнішу інформацію про вживаність і сполучуваність слів [5]. Зараз вже стало звичною практикою використовувати корпуси для укладання довідкової літератури – граматик, навчальних словників та довідників як для носіїв мови, так і для тих, хто її вивчає [2, с. 107].

Загалом, корпуси як джерела емпіричних даних відіграють вагомую роль у лінгводидактиці. Слід визнати, що штучно створені приклади вживання лінгвістичних одиниць та явищ є малоефективними, оскільки вони можуть суттєво відрізнятись від їх реального функціонування у мові. Приклади, взяті з корпусів, надають більш точний та деталізований опис мовних одиниць, що у свою чергу сприяє

опануванню мови на належному рівні. Крім того, безпосередня робота студентів з «живим» мовленням лінгвістичного корпусу викликає інтерес до вивчення мови і мотивує до самостійного навчання.

Найбільш широко лінгвістичний корпус використовується для розвитку лексико-граматичних навичок в рамках проблемного підходу до навчання, коли студенти мають проаналізувати мовний матеріал, щоб отримати інформацію про правила вживання певної лексичної одиниці або граматичної структури. Такі завдання з використанням лінгвістичного корпусу можуть зокрема полягати у з'ясуванні семантики слова, особливості полісемії слів, сполучуваності, керування дієслів, вживання часових форм тощо. Звичайно, на початковому етапі викладачу потрібно навчити студентів безпосередньо принципам роботи з корпусами, а надалі контролювати їх самостійну дослідницьку діяльність з виявлення граматичних правил або мовних закономірностей і кожен раз заздалегідь перевіряти ефективність того чи іншого корпусу у вирішенні поставленого завдання [3]. Серед найвідоміших корпусів англійської мови у вільному доступі слід зазначити:

- Британський національний корпус – British National Corpus (BNC): <http://www.natcorp.ox.ac.uk/>

- Корпус сучасної американської мови – Corpus of Contemporary American English (COCA): <http://corpus.byu.edu/coca/>

- Мічиганський корпус академічної англійської мови – Michigan Corpus of Academic Spoken English (MiCASE): <http://quod.lib.umich.edu/m/micase/>

Робота з корпусом здійснюється за допомогою спеціальних пошукових систем – корпусних менеджерів – які представляють результати пошуку у вигляді конкорданса, тобто списку, в якому одиниця пошуку представлена у в контекстному оточенні з фрагментів різних текстів (горизонтальні рядки із пошуковим словом посередині називають KWIC – Key Word In Context) і у вигляді статистичних даних, таких як частота словоформ, частота лексем, граматичних категорій, особливість сполучуваності, управління і т. ін.

Процедура корпусного аналізу включає в себе три кроки: ідентифікацію мовних даних за допомогою категоріального аналізу, співвідношення мовних даних за допомогою статистичних методів та інтелектуальну інтерпретацію результатів [2, с. 20]. Лексикологічні дослідження корпусів дають змогу встановити частотність вживання лексичних одиниць мови, відносини між значеннями слів, такі як багатозначність, омонімія, синонімія і антонімія, зміни в словниковому

складі мови, а також відображення в мові соціальних, територіальних, професійних характеристик.

В разі необхідності, дослідник може самостійно укласти прагматично-орієнтовані корпуси невеликого обсягу за допомогою корпусних менеджерів третього покоління Wordsmith, MonoConc, AntConc, Xlara. Ця можливість відкриває нові перспективи у роботі з визначення ознак різних типів дискурсу та виявлення прагматичних особливостей мовної комунікації. Крім того, такі інструменти потрібні, наприклад, у розробці матеріалів для курсу іноземної мови для спеціальних цілей. За допомогою корпусної методики дослідник зможе скласти перелік ключових слів, необхідних для професійного спілкування у певній галузі, визначити найбільш вживані словосполучення, фрази, дискурсивні маркери, граматичні форми, і розробити відповідний комплекс вправ [1].

Отже, корпусна лінгвістика надає методологічний інструментарій, який можна використовувати у сучасних лінгвістичних розвідках, але який також має дидактичний потенціал для розвитку соціокультурної, прагматичної та дискурсивної складової іншомовної комунікативної компетентності.

Літературні джерела

1. Горина О. Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке : автореф. дис. ... канд. пед. наук : 13.00.02 / Горина Ольга Григорьевна; Моск. гос. ун-т им. М. В. Ломоносова. – Москва, 2014. - 24 с.

2. Жуковська В. В. Вступ до корпусної лінгвістики: навчальний посібник / В. В. Жуковська. – Житомир: Вид-во ЖДУ ім. І. Франка, 2013. – 142 с.

3. Сысоев П. В. Лингвистический корпус в методике обучения иностранным языкам / П. В. Сысоев // Язык и культура. 2010. №1 (9). – С. 99-111.

4. Corpus Linguistics. An International Handbook / Ludeling A., Kytö M. (eds.). Volumes 1. – Berlin & New York : Walter de Gruyter, 2008. – 776 p.

5. The Collins Corpus [Електронний ресурс]. – Режим доступу: <http://www.collins.co.uk/page/The+Collins+Corpus>