

## КОРПУСНІ ДОСЛІДЖЕННЯ В СУЧАСНІЙ ЛІНГВІСТИЧНІЙ ПАРАДИГМІ

Вінницький національний технічний університет

### *Анотація.*

*Робота має на меті визначення статусу комп'ютерних методів та створення мовних корпусів у сучасній лінгвістичній парадигмі. Ці методики, що з'явилися завдяки Інтернету, довели свою плідність і перспективність у всіх галузях сучасної науки. У роботі також описано можливі напрямки вивчення текстового простору Мережі.*

**Ключові слова:** мовний корпус, Інтернет дискурс, корпусна лінгвістика, методика навчання іноземних мов, лінгвістична парадигма.

### **Abstract.**

*The suggested paper is aimed at defining the status of computer methods and corpus creating in modern linguistic paradigm. These techniques brought to life by the Internet and up-to-date information technologies have proved to be diverse and seminal. At present they go viral in all sciences and humanities. Some prospects of the Net text expanse studies are described. .*

**Key words:** language corpus, Internet discourse, corpus linguistics, foreign language teaching methodology, linguistic paradigm.

Сучасні лінгвістичні дослідження Інтернету включають вивчення таких аспектів як гендерний, ідентифікація та презентація віртуальної особистості, комунікативний простір Мережі як середовище функціонування комп'ютерного (або електронного) дискурсу. В цьому напрямку виконані дослідження Л.Ф.Компанцевої, О.І.Горошко, О.Є.Войскунського та інших. Із закордонних авторів варто згадати Г.Лернер, Е.Шерман, Х.Кларк, М.Костельс. Проблемам мовної синтагматики в Інтернеті присвячені дослідження Л.Ю.Іванова, С.А.Недобуха, Г.М.Трохимової, О.М.Галічкіної, А.Е.Жичкіної, а також Дж.Барбатіса, Дж.Нільсона, Р.Дуокінс, Д.Крістал, І.Дісон та інших. Інтерес дослідників до функціонально-стилістичних характеристик Інтернет-комунікації, виділення специфіки жанрів відображено в роботах В.П.Захарова, О.В.Булдакова, Л.А.Капанадзе, П.Лінч, С.Хортон, Дж.Челенджер.

Використанню Мережі як лінгвістичного корпусу присвячено роботи М.Фолька, Г.Греффенстета, Ф.Ресніка, А.Кілгаріфа, Е.Агіра, Д.Мартінеса, а також представників сучасної корпусної русистики – М.В.Копотева, А.Мустайоки, В.П.Захарова, В.І.Белікова, Т.І.Резнікової. В роботах цих авторів вивчаються як ідіолекти, зокрема, авторські, так і узагальнюється методологія, яка формує зміст корпусної лінгвістики, теорію та методику, створення корпусів, а також власне корпусні дослідження, тобто дослідження різних аспектів мови за допомогою корпусних методів. Створення лінгвістичних корпусів української мови на сучасному етапі не носить системного характеру.

Українські лінгвістичні дослідження Інтернету сьогодні, в основному, ведуться у когнітивному, лінгвопрагматичному, лінгвокультурному ракурсах (Компанцева Л.Ф., Дмитрук О.В., Вінарева О.В., Столярова М.О. та інші) переважно на англомовному матеріалі.

Метою нашої статті є аналіз взаємного впливу методів традиційної та комп'ютерної лінгвістики, визначення ролі Інтернету у розвитку корпусної лінгвістики, її перспективних напрямків, зокрема методологічному та дидактичному.

Незважаючи на твердження Ф. де Соссюра про дуалізм природи мови, і те, що виділена ним дихотомія *langue-parole* є загально прийнятою в науці, не можна однозначно стверджувати, що сучасна лінгвістика має одну авторитетну теорію, що пояснює асиметричний характер мови, втілений у тезі С.Карцевського. Що в решті решт є об'єктом лінгвістичного дослідження: *langue* чи *parole*, мова чи мовлення, парадигма чи синтагма. Не існує єдиної думки з цього приводу, але очевидно, що саме поєднання інтуїтивного і текстологічного підходів здатне забезпечити верифікованість наукових результатів. Поєднання традиційних лінгвістичних дослідницьких підходів із сучасними інформаційними технологіями зробило можливість створення надзвичайно великих за розмірами і

різноманітними за характером сукупностей мовного і мовленнєвого матеріалу. Ці сукупності називаються корпусами, а напрямок мовознавства, що використовує такі інструменти – корпусною лінгвістикою.

Звичайно, перші корпуси виникли задовго до появи електронно-обчислювальних методів обробки інформації. Діахронічні дослідження завжди будувались на інтроспективі якогось обмеженого текстового матеріалу. Діалектологічні польові екскурси також узагальнюють певну сукупність природного мовленнєвого матеріалу. Перехід на електронні форми зберігання і обробки лінгвістичних корпусів не змінює суті методології дослідження, проте безмежно розширює його можливості і перспективи. Отож, корпусна лінгвістика є достатньо новим напрямком мовознавства, тісно пов'язаним з комп'ютерною і когнітивною лінгвістикою. З першою її об'єднує технологія та інструменти обробки мовного (текстового) матеріалу, а з другою вона єдина у базовому постулаті: об'єктом її інтересу є мовленнєва діяльність, представлена безмежною і невичерпною кількістю текстів. Корпусна лінгвістика в певному сенсі змінює пріоритети філології як науки. «Об'єктом изучения становится речь, не сводимая к языковой абстракции, нормам литературного языка, суждениям о правильности и неправильности в языке, основанным исключительно на интуиции образованного исследователя. Вторым важным теоретическим следствием корпусных исследований можно считать то, что сосюрсовская дихотомия заменяется представлением о первичности речевой деятельности с плавной шкалой генерализаций от речевого штампа до грамматического правила» [1, с.12].

Комп'ютерна лінгвістика, безперечно, забезпечує методологічну основу корпусних досліджень. Її значення різко і драматично виросло разом із зростанням і популяризацією Інтернету. Багато дослідників вважають WWW найбільшим і найпотужнішим лінгвістичним корпусом, що містить більш як мільярд документів. Два десятиліття тому корпус, що містив мільйон слів, вважався великим, сього дні досліджуються корпуси, що нараховують понад сто мільйонів [2, с.10].

Очевидно, що використання WWW як мовного корпусу є новітнім напрямком. Кількість релевантних підходів досить обмежена, але успішно вирішуються завдання на різних рівнях мовної системи: синтаксису, семантики, а також у практично-перекладному аспекті; в лексикографії та перекладі. Наявність у Мережі найрізноманітніших структурно і стилістично текстів усіма унормованими мовами світу, а також штучними та вигаданими мовами, надають дослідникам практично безмежні можливості. Надзвичайно важливо і те, що у WWW містяться паралельні тексти. Виникають нові можливості для діахронічного вивчення мовних змін. І хоча застарілі документи часто вилучаються з сайтів, уже існують успішні результати вивчення семантичних змін.

Фахівці в галузі теорії та практики перекладу вважають, що WWW є найкориснішим інструментом для спостереження за особливостями вживання слова або сполучення. Оскільки запит може бути обмежений мовою, а через URL і країною, то стає можливим отримання інформації як про власне мовленнєву реалізацію, так і про частотність. Лексикографія користується як словниковими ресурсами Інтернету, так і виявленням неологізмів, класифікацією їх у різних мовах, визначенням валентності тощо.

Із самого свого виникнення корпусна лінгвістика була тісно пов'язана з викладанням мови як іноземної. Відомо, наприклад, що 60% англійського розмовного мовлення у США припадає на 50 найчастотніших лексем. Безперечно, цей факт, отриманий експериментально [3, р.XVII] повинен бути врахованим під час добирання вокабуляру на відповідному етапі навчання. На результатах корпусних досліджень будуються навчальні словники і посібники лексичного і граматичного спрямування. Нажаль, такі дидактичні матеріали не створені поки що для української, україномовної аудиторії. Підручник А.К.Головіної «Частотный курс ускоренного обучения английскому языку по профилю радиоэлектроники», що був виданий у Ленінграді ще у 1978 році і активно використовувався у технічних вишах, на сьогодні застарів, а послідовників не знайшов.

Перспективним є і створення корпусів студентських тестів, що дозволило б класифікувати помилки, визначати типові і враховувати їх у процесі викладання. Такі відомості містяться у деяких словниках (Collins Cobuilt Student's Dictionary, Oxford Learner's English Dictionary), але вони звичайно не враховують мовну інтерференцію. Створення посібника, де враховувалися б помилки саме україномовних студентів під час вивчення англійської мови, було б кроком уперед у вітчизняній лінгводидактиці.

М.Вольф відзначає, що проблемою сучасного підходу до використання WWW як лінгвістичного корпусу є відсутність спеціалізованої пошукової машини. «We ... have to live with the operators and options they offer. But these search engines are not tuned to the needs of linguists» [2, с.7]. Автор навіть

сформулював вимоги до ідеального пошукового двигуна для лінгвістичних корпусних потреб, серед яких можливість обмеження пошуку за синтаксичними, семантичними, текстологічними (стилістичними), жанровими ознаками, певними характеристиками сфери вжитку мовної одиниці (від літери, слова до складного синтаксичного цілого) тощо.

Корпусна лінгвістика набуває все більшої популярності останнім часом, хоча і була жорстко розкритикована патріархом сучасної наукової лінгвістики Н.Хомскі [4, р.10]. Очевидно, що вона становить повноцінну альтернативу традиційній філології.

Іншою зоною взаємовпливу Інтернету і наукової лінгвістики є вивчення Інтернет-комунікації. Тут поєднуються конгнітивно-прагматичний і лінгвокультурний аспекти вивчення. Інтернет-комунікація трактується сьогодні як особливе середовище актуалізації мовних одиниць, де переважає вербальна складова. Фахівці говорять про формування специфічної мови Інтернету, вивчають її стилістичні характеристики, утворення жанрів (сайту, блогу, чату).

Інтернет-лінгвістику інколи розуміють як «наївно-наукові» дослідження користувачів Інтернету, покликанні до життя текстовим простором WWW. Це і створення так званих вигаданих мов (конлангів), і «популярні» переклади, створення аматорських словників, спроби упорядкування термінології тощо. Інтернет, на відміну від інших ЗМІ, зокрема телебачення, передбачає співучасть, а не є культурою споживання, тому і формується Інтернет-мова (Netspeak або e-talk), яка є багатоаспектним явищем для лінгвістичного аналізу.

На сьогодні Інтернет можна вважати символом нашого часу, оскільки його вплив на розвиток світової цивілізації є безпрецедентним. Це складна соціотехнічна система, що безперервно працює і змінюється, популярність і значення якої постійно зростають. Вивчення численних наслідків цього зростання, як і внутрішнього простору Мережі потребує міждисциплінарного підходу, який отримав назву Інтернет-стадіз (Інтернет-студії). Антропоцентризм сучасної науки висуває на передній план дослідження Web-простору такі гуманітарні науки, як соціологія, психологія, інформатика і, звичайно, лінгвістика.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. М.В.Копотев, А.Мустайоки.Современная корпусная русистика// Инструментарий русистики: корпусные подходы. – Хельсинки, 2008.
2. Martin Volk. Using the Web as Corpus for Linguistic Research// Publications of General Linguistic 3. University of Tartu. – 2002.
3. Nation I.S.P. Teaching and Learning Vocabulary. – New York, 1990/
4. Chomsky N. The Master and His Performance// Intercultural Pragmatics. - 1993. – P.111.

*Степанова Ірина Сергіївна* – канд. філол. наук, доцент, завідувачка кафедри іноземних мов Вінницького національного технічного університету

*Stepanova Iryna S.* – Candidate of Philology, Associate Professor, Head of the Foreign Language Department of Vinnytsia National Technical University