

Автор статті робить спробу визначити основні особливості перекладу нетипичних запитальних речень українською мовою. Також проводиться аналіз засобів передачі модальності нетипичних запитальних речень в англійському та українському дискурсах. При перекладі речень даного виду перекладач повинен використовувати в цільовому тексті речення, які виражають ідентичну з оригіналом емоційну оцінку ситуації та комунікативну мету, хоча лексико-грамматична структура перекладених запитань може відрізнятися від вихідного речення. Перекладачі, можливо, скористаються результатами дослідження в процесі перекладу текстів художественних творів.

Ключові слова: нетипичні запитальні речення, вторинна функція, модальність, запитання.

Rudenko N.V. Ways of modality rendering of atypical interrogative sentences in English and Ukrainian discourses. The article deals with the ways of modality rendering of atypical interrogative sentences in English and Ukrainian discourses. As a result of transposition the interrogative sentences can be used in the secondary function and are atypical interrogative sentences. In this type of sentences by means of special language means the aspiration of the speaker is expressed, the aspiration to know something, make sure of something, or induce an interlocutor to express opinion, that interests him. The author of the article makes an attempt to determine the basic features of translation of atypical interrogative sentences into Ukrainian language. The analysis of means of rendering of modality of atypical interrogative sentences in English and Ukrainian discourses is also being made. While translating the sentences of such kind the translator should use in the target text the sentences that express a similar emotional estimation of the situation and have the same communicative goal as the original sentence, although the lexicogrammatical structure of translated questions can differ from the original sentence. The translators may use the results of the research in the process of translation of the texts of works of literature.

Key words: atypical interrogative sentences, secondary function, modality, question.

Статтю схвалено
кандидатом філологічних наук, доцентом кафедри
германської філології
Сумського державного університету
С.В. Барановою

Наталія Таценко
(Запоріжжя)

УДК 811.111:81-13

КОРПУСНА ЛІНГВІСТИКА ЯК МЕТОДОЛОГІЯ СУЧАСНИХ ФІЛОЛОГІЧНИХ НАУКОВИХ РОЗВІДОК

У статті проаналізовано базові поняття корпусної лінгвістики в якості методологічної основи сучасних мовознавчих досліджень. Доведено, що корпусні ресурси забезпечують мовознавців емпіричним матеріалом і екстралінгвальною інформацією для підтвердження їхніх гіпотез. Уточнено дефініції понять «корпус» та «конкорданс». Корпус визначається як стандартно подане зібрання писемних або усних репрезентативних машиночитаних текстів фіксованого розміру, які призначені для лінгвістичного опису й аналізу певної мови або діалекту. Конкордансом вважається показник, що пов'язує кожне слововживання з контекстом для отримання статистичних даних про частотність лексем, словоформ, граматичних категорій, динаміку словотворення. Розроблена типологія фактів і мовних явищ як типових та прототипових. Вона базується на виявлених розбіжностях між системними й функціональними особливостями мовних одиниць. Термін «типовий» використовується щодо найбільш характерних випадків дистрибуції й значень мовних одиниць, а «прототиповий» відображає розбіжності між типовими випадками вживань мовної одиниці й типовими значеннями та даними про частотність.

Ключові слова: корпус, конкорданс, метамова, типовий, прототиповий.

Постановка проблеми. Сьогодні позначене розвитком інформаційного суспільства знань, яке спричинило бурхливий прогрес у галузі комп'ютерних технологій опрацювання

природної мови. Перед мовознавством постали нові завдання щодо аналізу властивостей мовної системи, які потребують дослідження різнопланових мовних одиниць, структур, мовленнєвих явищ не на окремих показових прикладах, а в їхньому повному репрезентативному обсязі. Зрозуміло, що це вимагає застосування спеціального комп'ютерного інструментарію, зокрема, залучення методів корпусної лінгвістики. Попит на корпусні дані співпав із появою відповідних технічних можливостей.

Корпусна лінгвістика – це відносно нова методологія аналізу реального функціонування мови, яка дає змогу уточнити висновки й результати, проведені раніше, здійснюючи більш системні та ширші за обсягом емпіричного матеріалу дослідження. У центрі уваги такої методології є мовленнєва діяльність мовної особистості, а також масова комунікація та механізми її опису. При цьому репрезентативність корпусу текстів є дуже важливою властивістю, визначеною фонетичними, морфологічними, синтаксичними та стильовими характеристиками.

Аналіз останніх досліджень і публікацій. Корпусна система даних, з одного боку, суттєво спрощує пошук матеріалу, однак з іншого боку вимагає глибокого знання й творчого використання різних підходів і методик лінгвістичних розвідок. На основі корпусів текстів досліджували фразеологію [11; 12], граматичні моделі [8], лексичні кластери та їхні функції в різних жанрах [4]. Не зважаючи на поширене переконання, що корпус мовних даних повинен оброблятися статистично, доводячи чи спростовуючи гіпотетичні припущення за умови високої частотності випадків, він також є цінним джерелом якісної інформації – прикладів уживання мови в природній комунікації [9, с. 75-77]. Такий вид використання корпусів у когнітивній лінгвістиці називають корпусно-ілюстративним (corpus-illustrated research) [13]. Очевидно, що для когнітивно-дискурсивної практики моделювання пізнавальних метафор корпусна методика є надзвичайно актуальною.

Мета статті. Наведені факти визначили мету нашої статті – розглянути основні принципи й теоретичні засади корпусної лінгвістики, а також проблеми її створення й існування в сучасній теорії мови. Мета дослідження обумовлює такі **завдання**: 1) уточнити дефініції понять «корпус» та «конкорданс»; 2) проаналізувати типологію фактів і мовних явищ як центральних, типових та прототипових; 3) актуалізувати «фрейм» як один з термінів метамови корпусного аналізу.

Виклад основного матеріалу дослідження. Корпусний дослідницький напрям продовжує виформовуватись, а його проблематика пов'язана з розробленням теоретичних засад і практичних прийомів побудови, машинного опрацювання й експлуатації лінгвальних даних, оформлених як корпус текстів. Наразі існує безліч визначень поняття «корпус», тому ми наведемо лише декілька з них. Представники Ланкастерської школи корпусної лінгвістики визначають його як збірник мовних фрагментів, відібраних відповідно до чітких мовних критеріїв для використання в якості моделі мови [9]; крім того, він являє репрезентативну збірку текстів, зазвичай у машиночитаному форматі, яка включає інформацію про ситуацію текстотворення, тобто, інформацію про мовця, автора, адресата й аудиторію [7].

Зауважимо, що в сучасних визначеннях можна простежити основні риси корпусу текстів – мету або логічну ідею, машиночитаний формат, репрезентативність як результат певної відбіркової процедури, а також наявність металінгвістичної інформації. Це уможливорює розуміння **корпусу** як стандартно поданого зібрання писемних або усних репрезентативних машиночитаних текстів фіксованого розміру, які призначені для лінгвістичного опису й аналізу певної мови або діалекту й відібрані та упорядковані відповідно до інтра- й екстралінгвальних критеріїв. Стандартизоване подання словесного матеріалу на машинному носії дозволяє використовувати стандартні програми його оброблення.

Корпуси бувають по-різному організовані в залежності від прагматичної мети їх творців, хоча також використовуються для цілей, не передбачених авторами. Вони поділяються на одномовні та багатомовні, загальні й спеціалізовані, паралельні й порівняльні (перекладацькі корпуси), синхронні та діахронні тощо. Тексти як складові елементи корпусу можуть бути цілісними оригінальними словесними творами або їх частинами. Матеріали, зібрані в текстах – це зразки мови, яка використовується в комунікативній діяльності спільноти, вони охоплюють різні жанри й періоди.

Учені слушно зазначають, що доцільність створення та смисл використання корпусів визначається такими передумовами: достатньо великий (репрезентативний) і збалансований об'єм корпусу гарантує типовість даних і забезпечує повноту подання всього спектру мовних явищ; різнотипні дані перебувають у корпусі у своїй природній контекстуальній формі, що уможливорює їх усестороннє й об'єктивне дослідження; один раз створений і підготовлений масив даних може використовуватись багаторазово, різними дослідниками з різною метою [1, с. 6].

Перший комп'ютеризований корпус текстів – Браунівський корпус (The Brown Corpus) – включає 500 текстів із американських газет, журналів та книг. Мета створення Браунівського корпусу – забезпечити системне дослідження й порівняння окремих жанрів американського варіанта англійської мови – викликала жваві дискусії стосовно принципів відбору текстів і можливих вирішуваних завдань. З одного боку, він будувався на основі статистичних критеріїв, а з іншого – статистика застосовувалась у поєднанні з рішеннями творців корпусу, базованих на професійній інтуїції. Досягнення максимальної об'єктивності вимагало побудови формалізованих, прозорих для перевірки й контролю процедур. Пізніше за такими ж критеріями був створений корпус британського варіанта англійської мови Ланкастер-Осло-Берген (The Lancaster-Oslo-Bergen Corpus or LOB). Наразі обидва корпуси залишаються корисними, на них базуються численні дослідження.

Очевидно, що в наступні роки кількість і різноманітність створюваних комп'ютерних корпусів значно зросла й на сьогодні їх зафіксовано більше ніж 600. Серед сучасних корпусів англійської мови (як американського, так і британського варіантів) найбільш відомими є Британський національний корпус (British National Corpus – BNC), Міжнародний корпус англійської мови (International Corpus of English – ICE), Лінгвістичний банк англійської мови (Bank of English), Корпус сучасної американської англійської мови (Corpus of Contemporary American English – COCA) тощо.

Корпусний пошук дозволяє побудувати *конкорданс* будь-якого слова – список його уживань у контексті з посиланнями на джерело. Загалом конкорданс – це покажчик, що пов'язує кожне слововживання з контекстом; отриманий у автоматичному режимі набір контекстів для заданого явища (слово / словосполучення / граматична форма тощо) [1, с. 138]. Він використовується для отримання різних статистичних даних, зокрема, про частотність лексем, словоформ, граматичних категорій, їх спільне вживання, прослідковування зміни частотності й контекстів у різні періоди часу, динаміку словотворення.

Лінгвісти вказують на існування фундаментальних конкордансів, які становлять скарбницю знань про вживання в текстах тієї чи іншої мовної одиниці, і дослідницьких, що підпорядковані розв'язанню конкретного завдання [2, с. 83]. Вибране слово, що видається в центрі екрана, є «вузловим». Лінії конкордансу видають інформацію хаотично, проте можуть бути сортовані за алфавітним порядком чи групами, які вибрані й організовані для ілюстрації певної особливої поведінки заданого слова чи фрази [8, с. 39].

Науковці, що працюють у цій царині, поступово виформовують свою метамову, з'являються певні суто корпусні принципи досліджень. Наприклад, розроблена типологія фактів і мовних явищ як центральних, типових та прототипових [8]. Вона базується на виявлених розбіжностях між системними й функціональними особливостями мовних одиниць. Термін «типовий» (typical) використовується щодо найбільш характерних випадків дистрибуції й значень мовних одиниць, а «прототиповий» (prototypical) відображає розбіжності між типовими випадками вживань мовної одиниці й типовими значеннями та даними про частотність. Термін «центральний» (central) дає змогу уточнити ієрархію мовних одиниць для репрезентації граматичного значення, з урахуванням частотності й типовості.

Прототиповими вважають такі лексичні засоби, частотність яких згідно інтуїції користувачів або носіїв мови мала б бути високою, але за результатами корпусних досліджень відрізняється від тієї, що передбачалась. Розбіжності між типовими й прототиповими одиницями виявляються на основі різних реєстрів, жанрів, соціолінгвального варіювання мови. Для ілюстрації типовості слід розглянути використання певного слова чи словосполучення й визначити, у якому значенні воно вживається найчастіше – це значення й буде типовим.

Наприклад, дієслово *sympathize* має два значення: 1) *to feel or express sympathy for*; 2) *to agree with or support* [6, с. 896]. Ми розглянули 50 перших із 80 випадків ужитку дієслова

sympathize у BNC й виявили значення «*to feel or express sympathy for*» в 46 контекстах. Наведемо декілька прикладів:

I sympathize with the Duke of Wellington who, towards the end of his life it is said, grieved continually over the brave men killed at the Battle of Waterloo (BNC, *sympathize*, 4).

Madge Garland had a terrible experience, which I can sympathize with (BNC, *sympathize*, 21).

I sympathize because in my little way I grieve too, but over the people in my life I have allowed to die without ever having shown them any particular affection (BNC, *sympathize*, 5).

Лише в чотирьох контекстах із 50-ти це слово вживається в значенні «*to agree with or support*»:

They often sympathize with feminist critiques of traditional theory (BNC, *sympathize*, 27).

It is very possible to sympathize in good part, if not entirely, with the psychologists and philosophers who were sceptical or uncertain of the worth of introspection as a source of knowledge, reluctant to attempt to deal with the unquantifiable, keen to be in accord with certain principles of scientific methodology, resistant to such free speculation as the Freudian kind, and who thus took the step of analysing ascriptions of consciousness into claims about no more than behaviour (BNC, *sympathize*, 33).

It is possible to sympathize with these arguments, but there are more fundamental criticisms of key settlement policies (BNC, *sympathize*, 37).

On the Monday, after making one mistake by phoning the Evening Standard – and I can sympathize with your reasons for that mistake; after all, it was a heaven-sent opportunity to assert your father's continued existence – anyway, after that you get a train back to London ... (BNC, *sympathize*, 50).

Таким чином, корпусні дані дають змогу стверджувати, що типовим значенням лексеми *sympathize* є «*to feel or express sympathy for*». Це значення співпадає з інтуїтивно визначеною нами типовістю. Однак, інтуїтивна типовість не завжди збігається з даними частотності вживання лексем. Досить часто типове значення лексем виявляється прототиповим. Так, за результатами корпусних досліджень, деякі науковці схиляються до думки, що матеріал сучасного використання, зібраний у підручниках для вивчення англійської мови, є не типовим, а прототиповим [3].

Корпуси уможливають вирішення проблем із тлумаченням і використанням синонімів та «майже синонімів», тобто слів, які не є взаємозамінними [8, с. 45]. Оскільки лексеми тісно пов'язані з контекстом, їхні значення розрізняють за шаблонами або патернами (patterns) і фразами, в яких вони типово з'являються. Співвідношення значень і патернів можна розглядати за допомогою полісемантичних слів. Водночас, слова зі схожими значеннями вживаються в однакових патернах. Для аналізу цих співвідношень необхідно розподілити лінії конкордансу на блоки, кожен з яких буде містити приклади репрезентації одного значення, що дасть змогу чітко прослідкувати прив'язку патерну до конкретного значення.

Окрім загального аналізу даних щодо використання лексем та їхніх значень, які асоціюються з певними патернами, за допомогою конкордансів можна спостерігати за їх статусом і функціями, сполученням з іншими словами й за тим, що ці сполучення означають [8, с. 51]. Маємо зауважити, що патерни є імпульсом дослідження конститuentів мови, оскільки виявлення типу й значень патернів можна прирівняти до процесу розуміння людиною граматичної структури мови.

Одним із понять метамови корпусного аналізу мовних даних є фрейм – схема, яка зазвичай складається з трьох слів, перше й останнє з яких залишаються незмінними, а ті, що посередині, змінюються й несуть смислове навантаження. Список таких фреймів був запропонований на матеріалі малого корпусу (10 млн. слів письмової й 1 млн. слів розмовної англійської мови):

- **a ... of** – *a lot of, a matter of, a number of, a sort of, a couple of, etc.*;
- **an ... of** – *an example of, an element of, an act of, an average of, etc.*;
- **too ... to** – *too late to, too much to, too easy to, too late to, too young to, too close to, etc.* ;
- **for ... of** – *for most of, for all of, for fear of, for both of, etc.*;
- **many ... of** – *many years of, many kinds of, many parts of, many millions of, many thousands of, etc.*;

- **had ... of** – *had enough of, had plenty of, had thought of, had heard of, had one of, had died of, etc.*;

- **be ... to** – *be able to, be allowed to, be expected to, be said to, etc.* [10, с. 128].

Лексика, що наповнює наведені фрейми, належать до окремих груп, які несуть певні значення. До прикладу, слова, що найчастіше використовуються з *many ... of*, виражають числа (*thousands, millions, hundreds*), вказують на тип чи аспект (*kinds, ways, aspects, varieties*), позначають тривалість у часі (*years, hours*), і називають групи людей або речей (*members, examples*).

Виявлення й аналіз таких фреймових структур є надзвичайно корисним для створення комп'ютерних програм, за допомогою яких запрограмовані фрейми видаватимуться автоматично, безвідносно до попередніх знань і уявлень про те, якими вони мають бути. Фонетичні, морфологічні, синтаксичні, стильові фрейми й патерни репрезентують типове для корпусу та є більш уживаними, ніж сталі вирази. Вони можуть стати підґрунтям для узагальнення багатьох лінгвістичних теорій, зроблених мовознавцями раніше, й надихати їх на нові ідеї про сполучуваність кожного окремого слова в мовленні.

Завдяки розмітці в корпусі можна порівняти відносну частотність уживання певних лексем як різних частин мови (у випадках конверсії), або проаналізувати й порівняти, в яких сферах людської діяльності частіше використовується та чи інша частина мови [8, с. 80]. Як правило, точність розмітки становить дев'яносто відсотків, тому в деяких випадках слід зважати на судження дослідника, особливо якщо слово вживається невласливим для нього способом. У такому випадку деякі уточнення вносять вручну.

Висновки дослідження та перспективи подальших наукових розвідок у визначеному напрямі. Корпуси допомагають проаналізувати спонтанне мовлення, мовлення різних гендерних, соціальних, етнічних та вікових груп, інформацію про особливості певного діалекту; суттєво змінюють уявлення про мовленнєву культуру й мовні норми, являючи собою надійні критерії для оцінки й визначення прийнятності певних узусних явищ. Наявний масив мовленнєвих даних за певний період уможливорює здійснення аналізу лексико-граматичних характеристик мовлення різних авторів та різних жанрів, перевірки автентичності текстів, компаративних розвідок, інтертекстуальності, перекладу, прихованих (потенційних) моделей уживання лексики, її квантитативного навантаження, розвитку й динаміки концептів у часі, й навіть використання в розслідуванні злочинів (*forensic linguistic analysis*).

Слід додати, що корпуси слугують джерелом багатовекторних лексикографічних робіт зі створення сучасних та історичних словників. Двоїстий характер корпусної лінгвістики (налаштованість як на створення, так і на використання корпусів) обумовлюється подвійним характером її об'єкта – текстів, які являють собою вихідний мовленнєвий матеріал для мовознавців і водночас є продуктом корпусних досліджень. Окремий аспект складає теорія й практика програмного оброблення корпусних ресурсів.

Таким чином, корпусні ресурси забезпечують мовознавців емпіричним матеріалом і екстралінгвальною інформацією для підтвердження їхніх гіпотез. Проте арсенал корпусного методу не обмежується спеціально створюваними лінгвістичними базами електронних даних. Науковці досить часто звертаються до Інтернету як надзвичайно багатого джерела фактичного матеріалу. На наш погляд, Інтернет-ресурси програють корпусам, створюваним у наукових і освітніх цілях, за своїми пошуковими можливостями (надаючи неопрацьовані дані), проте відображають рідкісні мовні явища й окаціоналізми, які ще не встигли потрапити в спеціальний корпус. Детальний розгляд цього питання є перспективою подальших наукових досліджень.

ЛІТЕРАТУРА

1. Захаров В. П. Корпусная лингвистика : учебник для студентов направления “Лингвистика” / В. П. Захаров, С. Ю. Богданова [2-е изд., перераб. и дополн.]. — СПб : СПбГУ. РИО. Филологический факультет, 2013. — 148 с.
2. Карпіловська Є. А. Вступ до прикладної лінгвістики : комп'ютерна лінгвістика / Є. А. Карпіловська. — Донецьк : Юго-Восток, 2006. — 188 с.
3. Barlow M. Corpora for theory and practice / M. Barlow // *International journal of corpus linguistics*. — 1996. — № 1. — P. 1—37.

4. Biber D. Corpus linguistics : investigating language structure and use / D. Biber, S. Conrad, R. Reppen. — Cambridge University Press, 1998. — 300 p.
5. BNC : British national corpus [Electronic resource] // BNCweb (CQP-Edition). — Available at : <http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php>.
6. Collins English Dictionary / editors : J. Crozier, A. Grandison, C. McKeown, E. Summers, P. Weber. — Glasgow : HarperCollins Publishers, 2008. — 1040 p.
7. Finegan E. Language : its structure and use / E. Finegan. — N.Y. : Harcourt Brace College Publishers, 2004. — 575 p.
8. Hunston S. Corpora in applied linguistics / Susan Hunston. — Cambridge : Cambridge University Press, 2002. — 254 p.
9. McEnery T. Corpus linguistics : an introduction / T. McEnery, A. Wilson. — 2nd ed. — Edinburgh : Edinburgh University Press, 2001. — 235 p.
10. Renouf A. Collocational frameworks in English / A. Renouf, J. M. Sinclair. — Aijemer and Altenberg. — 1991. — P. 128—144.
11. Sinclair J. M. Corpora for lexicography / John Sinclair // A Practical guide to lexicography. — Amsterdam : Benjamins, 2003. — P. 167—178.
12. Stubbs M. Language corpora / M. Stubbs // In a Davies & C Elder eds. handbook of applied linguistics. — Oxford : Blackwell, 2004. — P. 106—132.
13. Tummers J. Usage-based approaches in cognitive linguistics : a technical state of the art / J. Tummers, K. Heylen, D. Geeraerts // Corpus Linguistics and Linguistic Theory. — 2005. — Vol. 1, № 2. — P. 225—261.

Таценко Н. Корпусная лингвистика как методология современных филологических научных исследований. В статье анализируются базовые понятия корпусной лингвистики в качестве методологической основы современных языковедческих исследований. Доказано, что корпусные ресурсы обеспечивают языковедов эмпирическим материалом и экстралингвистической информацией для подтверждения их гипотез. Уточняются дефиниции понятий «корпус» и «конкорданс». Корпус определяется как стандартно представленное собрание письменных или устных репрезентативных машиночитаемых текстов фиксированного размера, которые предназначены для лингвистического описания и анализа определенного языка или диалекта. Конкорданс является указателем, связывающим каждое словоупотребление с контекстом для получения статистических данных о частотности лексем, словоформ, грамматических категорий, динамике словообразования. Разрабатывается типология фактов и языковых явлений как типичных и прототипичных. Она базируется на выявленных расхождениях между системными и функциональными особенностями языковых единиц. Термин «типичный» используется в отношении наиболее характерных случаев дистрибуции и значений языковых единиц, а «прототипичный» отражает разногласия между типичными случаями употреблений языковой единицы и типичными значениями и данными о частотности.

Ключевые слова: корпус, конкорданс, метаязык, типичный, прототипичный.

Tacenko N. Corpus Linguistics as a Methodology of Modern Philological Scientific Studies. This article analyzes basic concepts of corpus linguistics as a methodological basis of modern linguistic research. It is proved that corpus resources provide linguists with empirical data and extralinguistic information to confirm their hypotheses. The definitions of «corpus» and «concordance» concepts are specified. The corpus is defined as a conventionally submitted collection of written or oral representational machine readable texts of fixed size, which is intended for linguistic description and analysis of specific language or dialect. Concordance is an index linking each linguistic usage with its context for getting statistics on the frequency of lexical units, word forms, grammatical categories, and word formation dynamics. The typology of facts and linguistic phenomena as typical and prototypical is developed. It is based on the differences between system-based and functional characteristics of language units. The term «typical» is used for the most typical cases of distribution and meanings of linguistic units, and «prototypical» reflects the differences between typical occurrences of linguistic units and typical meanings and data on their frequency.

Key words: corpus, concordance, metalanguage, typical, prototypical.