КОРПУСНІ ТА ЛЕКСИКОГРАФІЧНІ ТЕХНОЛОГІЇ ОПИСУ МОВНОЇ СИСТЕМИ. ТЕРМІНОЗНАВСТВО

УДК 811. 111:81-13

Ванівська О. І.,
Львівська комерційна академія

ОСНОВНІ ПІДХОДИ ДО АНАЛІЗУ МОВНИХ ДАНИХ У КОРПУСНІЙ ЛІНГВІСТИЦІ

У статті розглянуто різновиди основних підходів до аналізу мовних даних у корпусній лінгвістиці. Зібрано ґрунтовний матеріал щодо наукових розвідок в цьому напрямку. Зосереджено увагу на Британському національному корпусі (BNC). Представлено деякі приклади з BNC. Подано рекомендації щодо здійснення успішного та ефективного корпусного аналізу.

Ключові слова: *Британський національний корпус, конкорданси, корпусна лінгвістика, корпуси мовних даних, мовні (лексичні) одиниці.*

The article deals with a range of main approaches to the language data analysis in Corpus Linguistics. Solid material is collected concerning the study of this subject by various scholars and experts. The attention is turned on British National Corpus (BNC). There are some examples presented from BNC. Recommendations for the successful and effective corpus analysis are given.

Key words: *British National Corpus (BNC), concordance lines, Corpus Linguistics, linguistic corpora, language (lexical) units.*

В статье рассмотрены разновидности основных подходов к анализу языковых данных в корпусной лингвистике. Собрано основательный материал научных исследований этого направления. Сосредоточено внимание на Британском национальном корпусе. Представлены некоторые примеры с BNC. Поданы рекомендации по проведению успешного и эффективного корпусного анализа.

Ключевые слова: *Британский национальный корпус (BNC), конкорданси, корпусная лингвистика, языковые (лингвистические) единицы.*

Сьогодні корпусна лінгвістика та корпуси мовних даних посідають неабияке місце у навчанні та вивченні мов, відкривають нові перспективи для проведення нових лінгвістичних досліджень, допомагають з'ясувати, які зміни відбуваються в мові під впливом різноманітних зовнішніх факторів. О. Демська-Кульчицька пише, що "нові напрями розвитку мовознавчої науки характеризуються високим рівнем технологічності, що забезпечує лінгвістиці особливе місце в сучасному інформаційному світі" [3]. Широкі можливості комп'ютерів і комп'ютерних мереж сприяють і зумовлюють необхідність якісно нових засобів опрацювання інформації, які будуть перспективно розвиватися, особливо у напрямі інтелектуалізації [20]. Як зазначила О. І. Смашнюк, "дослідивши великий обсяг інформації, що міститься в корпусі, можна отримати повне уявлення щодо досліджуваного явища і певної мови в цілому. Велика кількість створених корпусів дає змогу отримати дані аналізу писемного чи спонтанного мовлення, мовлення певної вікової, гендерної, соціальної чи етнічної групи, інформацію про особливості певного діалекту" [17, с. 63].

В. Плугнян називає появу корпусів справжньою корпусною революцією і зазначає, що саме завдяки корпусам мовних даних тепер можна подивитися на мову в реальному використанні [14]. Використання автоматизованих корпусів суттєво змінює уявлення про мовні норми та культуру, представляє надійні критерії для визначення прийнятності та оцінювання тих чи інших явищ вживаності. Укладені будь-які сучасні словники, граматики і довідники, що не ґрунтуються на надійному корпусі, є лише саморобними витворами [33].

А. Б. Кутузов пише, що вивчення корпусів уможливлює отримання точних даних про лексичний склад мов, а також про відносну частотність використання деяких лексичних засобів (слів) [9]. Наприклад, Т. В. Монахова та О. Максимів акцентують увагу на важливості використання даних корпусної лінгвістики для укладання словників чи статистичної обробки лексичних даних. Предметом дослідження можуть бути не лише лексичні значення, а й морфологічні, синтаксичні та фонетичні особливості мови кожного автора [12; 11].

В. В. Риков слушно зазначає, "корпусна лінгвістика дала змогу уточнити результати та висновки, проведені раніше, а також провести нові лінгвістичні дослідження, більш системні й ширші за обсягом емпіричного матеріалу. У центрі уваги корпусної лінгвістики є мовна особистість, тобто її мовленнєва діяльність, масова комунікація, проблема її опису. Дуже важливою властивістю корпусу текстів є його репрезентативність, що визначається фонетичними, морфологічними, синтаксичними, стильовими параметрами" [16].

О. О. Шипнівська пише про те, що "розвиток інформаційного суспільства та суспільства знань спричинив бурхливий прогрес у галузі комп'ютерних технологій опрацювання природної мови, який поставив нові завдання перед лінгвістикою щодо вивчення різних властивостей мовної системи. Завдання, які стоять у цій царині перед усіма ділянками мовознавства, тепер потребують вивчення різнопланових мовних явищ, структур, одиниць, відношень тощо не на окремих, хоча й показових, прикладах, а в їх повному, репрезентативному обсязі. Зрозуміло, що така постановка завдань вимагає застосування спеціального комп'ютерного інструментарію дослідження, зокрема залучення методів корпусної лінгвістики" [19]. Таким чином, чимало науковців вбачають важливим використання лінгвістичних корпусів у методиці та процесі викладання, розробці різноманітних студентських завдань і окремих проектів [1; 13]. Комп'ютерна лексикологія, корпусна лінгвістика та комп'ютерна лінгвісти-

ка допомагають побачити й краще зрозуміти значення, переклад і вживання слів у динаміці [18], простежити за квантитативним навантаженням мовних одиниць [10]. У коло корпусної лінгвістики потрапляє все більше мов. Сьогодні розробляється також українсько-польський паралельний корпус [8]. Корпусна лінгвістика стає все більш поширеною у всьому світі.

Одним із основних підходів до аналізу мовних даних у корпусній лінгвістиці є конкорданси. Загалом конкорданс – це спеціалізована лінгвістична прикладна програма, за допомогою якої здійснюється автоматична вибірка заданих мовних одиниць з електронних текстів, проводиться дослідження корпусу за обраним словом, словосполученням чи фразою [7]. Є. А. Карпіловська описує існування фундаментальних конкордансів, які становлять скарбницю знань про вживання в текстах тієї чи іншої мовної одиниці, і дослідницьких, що підпорядковані розв’язанню конкретного завдання) [6, с. 83]. Залежно від технічних можливостей конкорданс може надавати інформацію про частотність вживання і сполучення тієї або іншої мовної одиниці, а також дає змогу звертатися до конкретного тексту, в якому був знайдений приклад, і демонструє слова, словосполучення чи фрази в центрі комп’ютерного екрану, разом зі словами, що стоять перед і після них, ліворуч і праворуч [7]. Вибране слово, що видається в центрі екрану, відоме як “вузлове”. Лінії конкордансу видають інформацію хаотично, але її можна сортувати, щоб вона надходила за алфавітним порядком, чи групами, які вибрані та організовані для ілюстрації певної особливої поведінки заданого слова чи фрази [30, p. 39].

Така методика і система пошуку даних, з одного боку, суттєво спрощує пошук матеріалу, з іншого боку, вимагає глибокого знання й творчого використання різних підходів, методик лінгвістичних досліджень. Корпусна лінгвістика поступово створює свою метамову, з’являються певні суто “корпусні” принципи досліджень, наприклад, класифікація фактів і мовних явищ як “центральних – типових – прототипових”, що дає змогу побачити різницю в значенні, досліджувати деталі. Така класифікація базується на розбіжностях між системними та функціональними/ комунікативними особливостями мовних одиниць, що були виявлені завдяки дослідженням у корпусній лінгвістиці [30]. Тож із створенням корпусів мовних даних “зросла можливість отримання відомостей щодо функціонування та вживання мови” [2, с. 46].

Термін “типовий” (typical) використовується в корпусній лінгвістиці щодо найбільш характерних випадків як дистрибуції мовних одиниць, так і значень мовних одиниць. Прототиповими вважають такі мовні засоби, частотність яких за інтуїцією носіїв чи користувачів мови мала б бути високою, але, як показують дослідження з корпусної лінгвістики, вони не є настільки частотними як передбачалось. Поняття прототиповості було введене науковцями Дерек Дейві та Девід Крістал. Термін “прототиповий” (prototypical) відображає розбіжності між типовими випадками вживань мовної одиниці і типовими значеннями та даними про частотність. Розбіжності між типовим й прототиповим виявляються на основі різних жанрів, регістрів, соціолінгвістичного варіювання мови. Термін “центральний” (central) більше застосовується до категорій, ніж до окремих слів, дає змогу уточнити ієрархію мовних одиниць, що використовуються для вираження граматичного значення, з урахуванням типовості й частотності [25]. Наприклад, теперішній тривалий час в англійській мові може вказувати на теперішній час (“But the big guy is working hard to change people’s perceptions. She’s cooking supper at the moment”), майбутній час (“Tomorrow we are holding a party in our bungalow, which has room for about 60 people, and I imagine about that number may come. She’s taking an exam tomorrow”) або взагалі не вказувати на жоден з часів (“Rock star Elton John is starting his own Aids charity. She’s always making mistakes”) [30, p. 43]. Таким чином, саме перший з поданих прикладів є центральним/граматичним значенням Present Continuous, – вираження дії, що відбувається в теперішньому часі, в момент мовлення.

Для того, щоб проілюструвати типовість, необхідно розглянути використання певного слова чи словосполучення і визначити, в якому значенні воно вживається найчастіше – це значення буде типовим для досліджуваного слова чи словосполучення. Наприклад, якщо розглянути 100 перших прикладів використання словосполучення “is having” (із 485 можливих прикладів) в BNC, то типовим значенням цього словосполучення є значення “*to possess, own, or able to use or give*” “мати” у 75 прикладах, а в решті 25 прикладах – це словосполучення використовується у модальному значенні. Наприклад: *We are going over to Trame to see what effect this* ***is having****. ’ DENIS Winston Healey* ***is having*** *a certain amount of characteristically mischievous fun with journalists at the moment, on the question of whether he will or will not stand at the next election. Behind me Nathan* ***is having*** *problems, his wooden runners sticking in the snow and causing him to go slowly. The last thing he wants to be bothered with* ***is having to*** *deal with complaints from dissatisfied guests* (модальне значення)*. ’ Less delicious* ***is having to*** *hit the thing apart again* (модальне значення).

У носіїв мови може бути певна інтуїція щодо типовості, але вона не завжди співпадає із даними частотності вживання тих чи інших слів, чи словосполучень. Майкл Барлов і Томас Шортал проводили деякі спостереження і використовували термін “прототиповий”, щоб вказати на використання, яке зазвичай мало б бути типовим, але не завжди чи не обов’язково найбільш часто вживаним. Вони стверджують, що в підручниках для вивчення англійської мови зібраний матеріал сучасного використання, що є саме прототиповим, а не типовим у значенні “найбільш часто вживаних” [21]. Прикладом може бути прикметник *each*, який є саме прототиповим некатегорійним засобом вираження значення часової форми *Present Simple*: *But* ***with each week that passes, Mrs Harris says*** *her concern increases for Alex’s safety and er for the welfare of her two young children*. І, згідно із дослідженням в BNC, із 100 довільно вибраних прикладів використання слова *each*, 36 відображають це значення часової форми теперішнього часу.

За допомогою корпусів мовних даних можна також вирішувати проблеми з використанням і тлумаченням слів, які є однакові або дуже схожі за значенням, але все ж у вжитку їх не можна замінити одне на інше. Тому такий аналіз мовних даних і спостереження за типовим використанням слів, які є “майже синонімами”, може багато чого прояснити в цій ситуації [30, p. 45].

Важливо згадати також і те, що слово є тісно пов'язане з контекстом, тобто з ситуацією, в якій воно вживається. Значення слів розрізняють за паттернами (шаблонами) і фразами, в яких вони типово з'являються. Співвідношення значень і паттернів можна розглядати за допомогою слів, які є багатозначними. З іншого боку, слова зі схожими значеннями вживаються в однакових паттернах. Щоб проаналізувати все вище викладене, необхідно розподілити лінії конкордансу на блоки, кожен з яких би містив приклади вираження одного значення, і тоді чітко було б видно, що кожне значення явно асоціюється з конкретними паттернами [30, p. 46].

Одним з різновидів підходу до аналізу мовних даних у корпусній лінгвістиці є поняття фрейму, – схеми, яка, зазвичай, складається з трьох слів, перше і останнє з яких залишаються незмінними, а слова, що стоять посередині, – змінюються і, таким чином, несуть смислове навантаження. Залежно від виду слотів та їхніх взаємозв'язків С. А. Жаботинська розрізняє п'ять типів фреймових структур [4, с. 16; 5]. Список таких фреймів, що складаються з трьох слів, був запропонований Антуанеттою Ренуф і Джоном Сінклером, які використовували малий корпус (10 млн слів письмової англійської мови і 1 млн слів розмовної англійської мови) [37, p. 128]:

• 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 too, 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.

Ці дослідження демонструють, що наповнення фрейму не просто випадкові, а належать до окремих груп, які є носіями певного значення. Наприклад, слова, що найчастіше використовуються з many … of, класифіковано за такими характеристиками :

– слова, що виражають числа, – thousands, millions, hundreds
– слова, що вказують на тип чи аспект, – kinds, ways, aspects, varieties
– слова, що вказують на тривалість в часі, – years, hours
– слова, що називають групи людей або речей, – members, examples.

Загалом виявлення таких структур особливо корисне для написання комп'ютерних програм, за допомогою яких запрограмовані фрейми будуть видаватися автоматично, без необхідності попередніх знань і уявлень про те, якими вони мають бути. Вони показують частину того, що типове для корпусу, і є фонетичними, морфологічними, синтаксичними, стильовими і більш вживаними, ніж сталі вирази. Фрейми і паттерни можуть стати підґрунтям для узагальнення багатьох тверджень, зроблених лінгвістами дотепер, і підштовхнути їх до нових ідей про сполучення кожного окремого слова в мові та мовленні.

Окрім загального аналізу даних щодо використання слів, їх значень, які асоціюються з певними шаблонами (паттернами), за допомогою конкордансів можна спостерігати за їх функціями та статусом, сполученням з іншими словами, а також за тим, що ці сполучення означають [30, p. 51]. Деніел Кіз вважає, що у людей граматика тісно асоціюється із паттернами (структурою, схемою): закінчення слів, функції слів та їх порядок. Він стверджує, що "паттерни є рушійною силою, яка допомагає нам досліджувати конституанти мови: впізнавання різновиду і значень паттернів дорівнює процесу розуміння людиною граматичної структури мови" [31].

Слід зазначити, що для роботи з корпусами великих розмірів, де кількість даних для обробки є доволі об'ємною, Джон Сінклер запропонував досліджувати щоразу по 30 випадково вибраних ліній конкордансу до тих пір, поки подальша вибірка не перестане видавати чогось нового. Такий аналіз мовних даних є "гіпотетичним тестуванням", в якому мала вибірка ліній стає основою для створення низки гіпотез про паттерни (patterns). Але такий спосіб дослідження застосовується лише для дуже часто вживаного слова [38, p. 157].

Не менш важливим є також те, що з обраних словосполучень можна отримати контури семантичного поля певного досліджуваного слова. Наприклад, Девід Орпін запропонував свій список словосполучень зі словами *bribe* і *bribery*, які разом взяті можна згрупувати за такими семантичними полями [36]:

– слова, що пов'язані з негативними діями (вчинками), – fraud, scandal, corruption, alleged, etc. ;
– слова, що пов'язані з грошима, – dollar, money, tax, etc. ;
– слова, що пов'язані з чиновництвом, – officials, police;
– слова, що пов'язані зі спортом, – players, referee;
– слова, що пов'язані з правовим процесом, – trial, charges, investigation, accused, etc.

Таке групування слів надає інформацію не лише про значення заданих слів, але й про деякі культурні розгалуження поняття, що позначено словом *bribery*.

Так, ми намагалися переглянути також використання дієслова *bribe* в BNC і виявилось, що воно використовується переважно з часткою *to*; це стосується давання хабара у вищезгаданих Д. Орпіном контекстах, і лише у 9-ти з 86 прикладів це дієслово використовується без частки *to*: у п'яти прикладах іде перерахунок дій за допомогою сполучника *and* (наприклад: *I'd have to go to Parliament and bribe them to pass a law specially for my divorce*), у двох – виражена майбутня дія з використанням допоміжного дієслова майбутнього часу *will* (наприклад: *We detest you so much we will bribe you to go away*), і ще у двох прикладах вказані особи, що дають хабаря – *I, they* (наприклад: *They telephone all day; they run after me in the streets; they bribe my barber for locks of my hair; they make my life unbearable*). Форма *bribed* використовується дещо частіше – у 98 прикладах, здебільшого в часовій формі звичайного минулого часу (*I think my guards bribed him to let us pass. They had bribed the executioner to jam a wooden peg in the side of the guillotine to stop the blade from falling all the way down. So you've bribed them along with a promise of sweets*).

Слід зазначити, що під час аналізу словосполучень і сталих виразів головне – це визначити важливість суттєвих, показових словосполучень, а не намагатися їх якимось чином інтерпретувати.

Наступний підхід до аналізу мовних даних стосується розмітки, завдяки якій можна обрати слово і вказати, що воно нас цікавить тільки, наприклад, як дієслово. Також можна порівняти відносну частотність вживання конкретного слова як різних частин мови (у випадках конверсії), або проаналізувати і порівняти, в яких сферах людської діяльності частіше використовується та чи інша частина мови. Зазвичай, точність, з якою розмітка видає інформацію, становить 90 %, тому необхідно пам’ятати, що в деяких випадках слід зважати на судження самої людини-дослідника (користувача), особливо якщо слово вживається у невластивому для нього способі. В такому випадку деякі уточнення вносять вручну [30, p. 80].

Існує також і граматичний розбір речень у корпусі, за допомогою якого ідентифікують загальні межі речення, фрази та звороти, що супроводжуються помітками, такі як прислівниковий зворот (adverbial clause), іменниковий зворот (nominal clause), порівняльний зворот (relative clause), прикметникова фраза (adjective phrase), прийменникова фраза (prepositional phrase). Цей аналіз мовних даних був запропонований такими лінгвістами як Джефрі Ліч і Елізабет Айз [32, p. 34]. Проте, комп’ютерні програми, написані для виконання граматичного розбору, не є на 100 % точними, і тому такі корпуси з граматичним розбором часто редагують вручну для досягнення вищого ступеня точності. Необхідно додати, що на основі таких корпусів виконано і виконуються чимало статистичних праць, пов’язаних з різними реєстрами, зокрема праці Дугласа Байбера [30, p. 84].

Ще один підхід до аналізу мовних даних у корпусній лінгвістиці стосується такої важливої характеристики як зв’язність тексту: аналіз використання слів і фраз у тексті, поєднаних зі словами і виразами, що стоять перед і після них [29]. Деякі слова-зв’язки використовуються для того, щоб підсумовувати, позначати чи виражати певний відрізок дискурсу, і таким чином відіграють роль в організації тексту [26, p. 83; 27]. У схемах, які описують (анотують) зв’язки в текстах, використовується термін анафора [39, p. 6].

Можуть бути використані різні варіанти анотації для того, щоб проаналізувати анафору в тексті, але більшість з них мають щось спільне з наступними [28, p. 66]:

• розпізнає анафору і антецедент (слово чи фраза, до якої належить анафора), або визначає, чи взагалі можна розпізнати антецедент;

• поділяє антецедент на категорії (як номінативний, підрядний та ін.);

• розпізнає напрямок зв’язку (forward or backward);

• розпізнає тип анафори (відношення (reference), заміни (substitution) і ін.);

• визначає відстань між анафорою та її антецедентом.

Мабуть, найцікавішим із вищезазначеного є зв’язок між анафорою і антецедентом. При застосуванні такого підходу кожному антецеденту присвоюється номер і такий самий номер закріплюється за анафорою, що до нього належить [28, p. 72]. Наприклад:

(1 *A man carrying a blue sports bag* 1) … *was arrested when* <REF=1 *he*… Приклад свідчить, що напрямок зв’язку є зворотним (backwards); REF означає “відношення (reference)”. Таким чином, можна простежити за розвитком тексту, показавши, що з чим найчастіше співвідноситься. Але недолік такого виду анотації в тому, що її неможливо робити автоматично, отже, об’єм тексту, який можна закодувати, обмежений [35, p. 261]. Проте, з іншого боку, така форма анотації відкриває нові перспективи й захоплюючі можливості у виявленні цікавих нюансів про типи анафор і антецедентів, що найчастіше трапляються в різних реєстрах, про те, які типові зміни відбуваються з анафорою під час розвитку тексту, і, врешті-решт, дає змогу представити анафоричний аналіз частин текстів у різних реєстрах.

Існує також і семантична анотація даних, яка полягає в тому, щоб розділити слова і фрази в корпусі на категорії за семантичними полями [41, p. 52]. Такий підхід до аналізу мовних даних запропонували Джеррі Томас і Ендрю Вілсон, які досліджували стосунки між лікарями та пацієнтами двох клінік [40, p. 92]. Використовуючи таку систему засобів семантичної класифікації, комп’ютер вираховує найчастіші значення висловлювань, що вживають лікарі, медпрацівники та пацієнти. Результат проведеного аналізу свідчить, що один з лікарів використовував більше особових займенників, підбадьорювальних слів, і, таким чином, його вважали більш комунікабельним і приязним до пацієнтів. Водночас, інший лікар використовував більше медичних термінів, пояснював пацієнтам перебіг хвороби. Як виявили вищезгадані науковці в своєму дослідженні, пацієнти були більше задоволені методами лікування першого лікаря, ніж другого, тобто їм приємніше було розмовляти про їхнє лікування, ніж про перебіг самої хвороби [30, p. 89].

Така автоматична анотація відіграє неабияку роль під час аналізу великої кількості текстових даних, що було б складно і нераціонально робити повністю вручну [40, p. 106]. З метою вивчення отриманих результатів та з урахуванням деяких відмінностей і внесених уточнень, згодом роблять так званий “якісний аналіз дослідження”.

Необхідно згадати ще про таку анотацію, як різновид семантичної анотації, тобто йдеться про часткову анотацію, що стосується певної категорії, наприклад, висловів про “позицію” чи “думку”. Цей підхід до аналізу мовних даних у корпусній лінгвістиці застосували Дуглас Байбер і Едвард Файнеган [22, с. 93], а також Сюзан Конрад і Д. Байбер [24, с. 57]. До цієї категорії увійшли такі мовні дані як, наприклад, прислівники (напр. *probably*), речення (напр. *I think*) і прийменникові фрази (напр. *on the whole*). Таким чином, було проаналізовано і зазначено, що прислівники часто використовують для граматичного вираження думки (позиції) у всіх трьох досліджуваних реєстрах, а саме: в розмовному мовленні, в газетних статтях і в академічній прозі, але найчастіше – в розмовній мові. Такі речення, як *I think* і *I guess* також найчастіше трапляються в розмовній мові. В академічній прозі та газетних статтях, окрім прислівників, ще широко використовують прийменникові фрази.

Корпусна анотація такого плану створює основу для підходу до корпусу з точки зору значення і може поєднуватись зі смисловим (понятійним) підходом до вивчання мови. Як зазначає М. М. Полюжин, морфологічні категорії необхідно вивчати “у зв’язку з когнітивними здібностями людини, що проявляються у проникненні в суттєві структури лексичних і словотворчих категорій” [15, с. 129]. Отже, варто сказати, що такий анотований

корпус дає змогу відповідати на запитання, які слова чи фрази найбільше підходять до ситуації, коли учень (студент) має щось сказати в певному контексті.

Існують три основні методи анотації корпусу: вручну, за допомогою комп'ютера та автоматичний [23, с. 35-37], з яких два останніх методи можуть використовуватись виключно для найменших корпусів. Зрозуміло, що при автоматичній анотації комп'ютер працює самостійно, відповідно до закладених в ньому правил і алгоритмів, і виконує анотацію будь-якого за об'ємом корпусу відносно швидко, але малоймовірно, що результати будуть на 100 % точними порівняно з результатами людини-дослідника. Що ж стосується анотації корпусу за допомогою комп'ютера, то вона дає змогу користувачеві коригувати комп'ютерний вивід даних (як при більшості граматичного розбору) і, таким чином, вручну покращити точність отриманих результатів, хоч ця робота буде виконана повільніше і в невеликому обсязі [30, p. 91].

Підсумовуючи все вище викладене, слід зазначити, що простір електронних текстових корпусів мовних даних дає можливість їх результативного використання, що відкриває перспективи моделювання мовної картини світу. Щоб обрати правильний підхід до аналізу мовних даних у корпусній лінгвістиці, необхідно знати, на яке питання ми хочемо отримати відповідь. Наприклад, якщо ми хочемо знати, як використовується певне слово, то найкраще використовувати звичайний корпус мовних даних; якщо необхідно визначити, яка анафора найчастіше вживається в академічній прозі, то нам потрібен анотований корпус, і т. д.

Щоб провести ефективний корпусний аналіз, треба мати чітко сплановану картину дослідження. Для початку, необхідно визначити мету дослідження. Наступним має бути правильний вибір самого корпусу, який мав би містити необхідний для дослідження матеріал. Далі слід вибрати відповідний програмний пристрій для проведення аналізу та кодування отриманих результатів. Якщо дотримуватись усіх правил, то можна бути спокійним, що отримані результати матимуть неабияку лінгвістичну цінність [34, p. 137].

Література:

1. Бук С. Учнівські корпуси в методиці викладання іноземної мови [Електронний ресурс]. – Режим доступу : http://www.franko.lviv.ua/faculty/Philol/www/teoria_praktyka_ukr_mova/vyp_2/3.%20Buk.pdf, 2007.

2. Гвишиани Н. Б. Корпусная лингвистика и граматика речи / Н. Б. Гвишиани, О. Ю. Герви // Вестн. Моск. ун-та. – М., 2001. – № 2. – С. 46-62.

3. Демська-Кульчицька О. Що нового в науці про мову ? [Електронний ресурс]. – Режим доступу : http://www.kulturamovy.org.ua/KM/pdfs/Magazine61-16.pdf

4. Жаботинская С. А. Концептуальный анализ : типы фреймов // Вісник Черкаського університету. – Черкаси, 1999. – Вип. 11. – С. 12-25.

5. Жаботинская С. А. Когнитивная лингвистика : принципы концептуального моделирования // Лінгвістичні студії. – Черкаси, 1997. – С. 3-11.

6. Карпіловська Є. А. Вступ до прикладної лінгвістики : комп'ютерна лінгвістика / Є. А. Карпіловська. – Донецьк : Юго-Восток, 2006. – 188 с.

7. Корпусна лінгвістика [Електронний ресурс]. – Режим доступу : http://uk.wikipedia.org/wiki/Корпусна_лінгвістика

8. Коциба Н. Морфосинтаксичне тагування польсько-українського паралельного корпусу (PolUKR) [Електронний ресурс]. – Режим доступу : http://www.domeczek. pl/~natko/papers/megaling2008.pdf, 2008.

9. Кутузов А. Б. Корпусная лингвистика. Лекция 2 [Електронний ресурс]. – Режим доступа : http://tc.utmn.ru/files/corpus_2.pdf

10. Левицкий В. В. Квантитативные методы в лингвистике / В. В. Левицкий. – Черновцы : Рута, 2004. – 190 с.

11. Максимів О. Корпус текстів перської мови як джерело матеріалу для навчальних словників-мінімумів [Електронний ресурс]. – Режим доступу : http://www.lnu.edu.ua/faculty/Philol/www/visnyk/45/21. %20Maksymiv.pdf, 2008.

12. Монахова Т. В. Застосування прийомів корпусної лінгвістики в лексикографії [Електронний ресурс]. – Режим доступу : http://www.nbuv.gov.ua/portal/Soc_Gum/Npchdu/Philology.Linguistics/2009_85/85-11.pdf, 2009.

13. Нагель О. В. Корпусная лингвистика и ее использование в компьютеризованном языковом обучении [Електронний ресурс]. – Режим доступа : http://www.lib. tsu.ru/mminfo/000349304/04/image/04-053.pdf

14. Плугнян В. Почему современная лингвистика должна быть лингвистикой корпусов [Електронний ресурс]. – Режим доступа : http://www.polit.ru/lectures/2009/10/23/corpus.html, 2009.

15. Полюжин М. М. Функціональний і когнітивний аспекти англійського словотворення. – Ужгород : Закарпаття, 1999. – 240 с.

16. Рыков В. В. Корпусная лингвистика [Електронний ресурс]. – Режим доступа : http://rykov-cl.narod.ru/c.html, 2002.

17. Смашнюк О. І. Маркери емоційності у спонтанній комунікації (на матеріалі Британського наці. корпусу текстів) : дис.... канд. філол. наук : 10. 02. 04 "Германські мови" / Смашнюк Оксана Іванівна. – К., 2008. – 238 с.

18. Хоменко Ф. В. Комп'ютерна лексикографія при вивченні іноземної мови [Електронний ресурс]. – Режим доступу : http://ev.nuos.edu.ua/content/komp%E2 %80 %99yuterna-leksikograf %D1%96ya-pri-vivchenn %D1 %96-%D1 %96nozemnoi-movi, 2010.

19. Шипнівська О. О. Структурно-семантичні та функціональні характеристики міжчастиномовної морфологічної омонімії сучасної української мови : дис.... канд. філол наук : 10. 02. 01 / Шипнівська Ольга Олександрівна [Електронний ресурс]. – Режим доступу : http://www.lib.ua-ru.net/diss/cont/339734.html, 2007.

20. Широков В. А. Корпусна лінгвістика : [монографія] / В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна, О. М. Костишин, М. Ю. Кригін, Т. П. Любченко, О. Г. Рабулець, О. О. Сидоренко, Н. М. Сидорчук, І. В. Шевченко, О. О. Шипнівська, К. М. Якименко. – К. : Довіра, 2005. – 471 с.

21. Barlow M. Corpora for theory and practice / M. Barlow. – International journal of corpus linguistics. – № 1. – 1996. – P. 1-37.

22. Biber D. and Finegan E. Style of stance in English : lexical and grammatical marking of evidentiality and affect. – Text 9. – 1989. – P. 93-124.
23. Biber D. Longman Grammar of Spoken and Written English / D. Biber, S. Johansson, G. Leech, S. Conrad, E. Finegan. – London : Longman, 1999.
24. Conrad S. and Biber D. Adverbial marking of stance in speech and writing / Eds. Hunston and Thompson. – 2000. – P. 57-73.
25. Crystal D., Davy D. Investigating English style / David Crystal & Derek Davy. – London : Longman, 1969. – 260 p.
26. Francis G. Labelling discourse: an aspect of nominal-group lexical cohesion / Ed. M. Coulthard // Advances in Written Text Analysis. – London : Routledge. – 1994. – P. 83-101.
27. Francis W. N., Kucera H. A. Standard Corpus of Present-Day Edited American English (Brown corpus) / W. N. Francis, H. A. Kucera. – Providence : Brown University, 1979.
28. Garside R., Flidgestone S., and Botley S. Discourse annotation : anaphoric relations in corpora / Eds. Garside et al. – 1997. – P. 66-84.
29. Halliday M. A. K. Cohesion in English / Halliday M. A. K. and Hasan R. – London : Longman, 1976.
30. Hunston S. Corpora in Applied Linguistics / S. Hunston. – Cambridge. – 2002. – 254 p.
31. Kies D. Form and Function of Word Classes in English / D. Kies [Електронний ресурс]. – Режим доступу : http:// papyr.com/hypertextbooks/grammar/word. htm, 2010.
32. Leech G. Syntactic annotation: treebanks / Leech G., Eyes E., Garside et al. – 1997. – P. 34-52.
33. Lifejournal. Maksymus. Корпусна лінгвістика [Електронний ресурс]. – Режим доступу: http://maksymus. livejournal.com/87361.html, 2009.
34. Meyer C. F. English Corpus Linguistics / C. F. Meyer. – 2004. – 168 p.
35. Mitkov R. Towards automatic annotation of anaphoric links in corpora / R. Mitkov // International Journal of Corpus Linguistics 4th ed. – 1999. – P. 261-280.
36. Orpin D. The lexis of corruption in the news: a corpus-based study in ideology / D. Orpin // Unpublished MA dissertation, University of Birmingham. – 1997.
37. Renouf A. Collocational frameworks in English / A. Renouf, J. M. Sinclair; eds. Aijemer and Altenberg. – 1991. – P. 128-144.
38. Sinclair J. M. A way with common words / Eds. H. Hasselgard and Oskefjell // Out of corpora : Studies in honour of Stig Johansson. – Amsterdam : Rodopi. – 1999. – P. 157-179.
39. Sinclair J. M. Written discourse structure / Eds. J. M. Sinclair, M. Hoey and G. Fox // Techniques of description. – London : Routledge, 1994. – P. 6-31.
40. Thomas J. Methodologies for studying a corpus of doctor-patient interaction / Thomas J., A. Wilson; eds. Thomas and Short. – 1996. – P. 92-109.
41. Wilson A. and Thomas J. Semantic annotation / A. Wilson, J. Thomas; eds. Garside et al. – 1997. – P. 53-65.