

ПЕРЕДМОВА

Сьогодні лінгвістика має у своєму розпорядженні важливе джерело формування фактологічних баз досліджень — текстові корпуси [1]. Матеріали текстових корпусів дедалі активніше починають застосовуватися в лінгвістичних дослідженнях. Значний час у лінгвістичній науці тривають дискусії щодо релевантності результатів аналізу, що зумовлено низкою чинників: вибір методів аналізу, фактичні дані, якими оперує дослідник. Теоретично корпуси текстів дають змогу лінгвістові сформулювати такий масив даних, аналіз якого великою мірою унеможливить суб'єктивізм дослідника. Однак лінгвісти в різний спосіб залучають матеріал до аналізу. У “Практичному путівникові по корпусах слов’янських мов” згадано, зокрема, такі “дослідницькі підходи” [2]:

Corpus-assisted analysis (корпуснодопоміжний аналіз) — один із корпусних дослідницьких методів, який ґрунтується на одночасному аналізі корпусних і позакорпусних даних (наприклад, інтерв’ю, анкети тощо) [3].

Corpus-based approach (корпуснобазований підхід) — один із корпусних дослідницьких методів, який полягає у використанні корпусу переважно для представлення, перевіряння чи егземпліфікації теорії або описів, які сформульовано без доступу до великих корпусів [4]. За Є. Тоніні-Бонеллі [5], під час корпуснобазованого дослідження корпус використовують як джерело прикладів, щоб перевірити інтуїцію дослідника або щоб дослідити частотність лінгвальних висловів тощо.

Corpus-driven approach (корпуснокерований підхід) — один із корпусних дослідницьких методів, який полягає на інтерпретації корпусних даних як цілісності. Висновки, які випливають із такого аналізу, мають відображати лише те, що засвідчує корпус. Дослідник має старатися забути про наявні знання щодо аналізованої теми з інших, позакорпусних джерел [6].

Належить зауважити, що Є. Тоніні-Бонеллі розглядає три випадки розбіжностей між теорією і практикою під час використання корпусних даних: а) **ізоляцію** (коли ідеї, отримані з корпусних даних, приймають тільки як допоміжне розширення теоретичних тверджень); б) **стандартизацію** (коли корпус просівають через попередньо наявні теоретичні параметри перед аналізом); в) **конкретизацію** (коли корпус наочно знову забезпечує ймовірніше розширення до вже наявної системи, але не впливає на систему як таку) [5].

У цій колективній монографії здійснено спробу розглянути проблеми створення корпусів текстів різного типу, використати корпуси текстів у лінгвістичних студіях, застосувавши різноманітні методики аналізу корпусного матеріалу, а також звернути увагу на питання, пов’язані зі створенням різного типу корпусів текстів.

У першому розділі розглянуто питання комп’ютерних систем опрацювання мови. В. Широков, І. Шевченко дослідили феноменологічні виміри граматики. У праці наголошено на тому факті, що граматична проблематика привертає увагу не лише теоретиків мовознавства, але дедалі більше прикладників, які створюють системи комп’ютерного опрацювання мови. Поняття систем комп’ютерного опрацювання мови охоплює широке коло завдань, пов’язаних з інформаційним пошуком, перекладом з однієї мови на іншу, інженерією знань, розпізнаванням і синтезом мовних та текстових образів, інтерпретацією змісту текстів, проведенням різних лінгвістичних експертиз, навчанням мови тощо.

І. Кульчицький досліджував технологічні аспекти укладання корпусів текстів. Зокрема, у цьому підрозділі подано аналіз технології укладання корпусів текстів. Автор зупиняє увагу на укладанні електронної бібліотеки текстів, визначенні єдиних правописних правил, а також детально описує процедуру нормалізації, яку мають пройти зібрані в електронній бібліотеці тексти, та подає технологічні вимоги до них.

А. Загнітко у підрозділі “Встановлення функційної характерології і парадигмально-синтагмального вияву часток в експериментальному дослідницькому лінгвістичному корпусі службовості” проаналізував основні різновиди корпусів із розкриттям їхніх кваліфікаційних та класифікаційних ознак, схарактеризував особливості метарозміток у корпусах текстів, визначив статус експериментальних дослідницьких корпусів у системі корпусів. У праці виявлено основні характеристики

конкордансерів та простежено закономірності їхнього використання в корпусах. Розкрито напрями створення Експериментального дослідницького корпусу службовості та схарактеризовано принципи створення бази даних часток української мови на ґрунті такого корпусу, запропоновано кваліфікаційні розмітки створеної бази даних, її полів, шляхів переміщення.

Проблему граматичних лексикографічних систем у дослідженні варіативності та неповноти словозмінної парадигми опрацьовано у підрозділі, авторкою якого є Т. Любченко. Граматичні лексикографічні системи (ГЛС) розробляють в Українському мовно-інформаційному фонді НАН України для різних мов у межах проекту зі створення Національної словникової бази України. У цій частині монографії подано опис концептуальної моделі словозмінної системи мови із флективним компонентом у словозміні, а також структури ГЛС. Головне призначення ГЛС — використання їх у системах автоматичного опрацювання текстів. Окрім того, є змога використовувати ГЛС під час дослідження словозмінної системи відповідної мови, також розглянуто методики дослідження таких властивостей словозмінної парадигми, як її неповнота й варіативність.

У другому розділі монографії звернено увагу на когнітивні дослідження корпусного матеріалу. У підрозділі монографії “Статистичні методи в дослідженні мовних стереотипів”, який написав Є. Бартмінський, висвітлено конкретні питання використання статистичного методу в дослідженні мовних стереотипів на тлі ширшої проблематики відношення між якісною й кількісною сторонами мови. Розглянуто роль стереотипів у міжкультурній комунікації. Представлено комплексний підхід, який використовують у працях з вивчення мовної картини світу і який ураховує як дані, що містяться в мовній системі, у словнику, фразеології, семантиці, використовує різножанрові тексти, від текстів-кліше (прислів'я, приказки, крилаті вислови, гасла, пісні, казки і т. ін.) до текстів публіцистичних, так і звертається до експериментальних методів: даних, зібраних за допомогою різного типу анкет і статистично опрацьованих. Такий комплексний підхід Є. Бартмінський називає аббревіатурою С-А-Т (система, анкета, тести).

У підрозділі “Порівняння як засіб вербалізації концепту “Жінка” (на матеріалі сучасної української художньої прози)” (Олена Левченко) подано приклад створення дослідницького корпусу для потреб когнітивного аналізу. Використання корпусних даних є важливою умовою отримання релевантних результатів, оскільки вивчення порівнянь уможливорює висновки про концептуальні й мовні картини світу сучасних жінок та чоловіків, а також дає емпіричний матеріал для аналізу змін, які відбуваються в мовних картинах світу. За допомогою аналізу вербалізованих порівнянь у праці подано висновки про атрибути, які належать концепту “Жінка” в картинах світу жінок та чоловіків.

Н. Романишин порушила проблему актуалізації концептів “Дім” та “Батьківщина” в ранній ліриці Дж. Байрона. Вказані концепти в обраних поетичних творах розглянуто як основні аспекти національної ідентичності, сутність і природу якої вважають одними з найневизначеніших та найскладніших понять поетичного спадку Байрона. Особливості вербалізації концептів “Дім” та “Батьківщина” досліджено в контексті загальної рецепції поезій Байрона в зарубіжній та вітчизняній лінгвістиці, а також у контексті загальних ідейних, образних та естетичних засад романтизму. Дані для проведення аналізу отримано через застосування програми конкорданс до корпусу повного зібрання творів поета, представленого на дослідницькому інтернет-ресурсі questia.com.

Питання аналізу дискурсу розглянуто у третьому розділі. І. Бехта у підрозділі “Наратив і контекст комунікативної ситуації: стратегії інтерпретації оповідного дискурсу експериментального письма” пропонує використовувати комунікативну стратегічну модель для аналізу дискурсу замість традиційної структурної моделі розуміння й творення тексту, зазначаючи, що проблематика сучасного дискурсного аналізу перебуває під впливом новітніх лінгвістичних парадигм. Висвітлено також у праці центральні положення дискурсного аналізу, які стосуються одиниць дискурсного аналізу, його структури, стратегій інтерпретації та інтерпретативних властивостей наративної моделі експериментального письма. У підрозділі “Дослідницький корпус англійсько-, німецько-, українськомовних текстів релігійно-популярного дискурсу”, автором якого є О. Черхава, визначено критерії ідентифікації релігійно-популярного дискурсу як релігійного типу дискурсу; запропоновано робоче визначення релігійно-популярного виду дискурсу; встановлено співвідношення

релігійно-популярного дискурсу із проповідницьким, пророчим та місіонерським видами, представлено місце релігійно-популярного дискурсу у дискурсивній композиції релігійного дискурсу на мікро- та макрорівнях; скласифіковано тексти релігійно-популярного дискурсу відповідно до їх конситуації (контексту як комунікативної події).

О. Семолюк звернувся до комп'ютерного контент-аналізу політичних карикатур США, ФРН та України, дослідивши суть, функції та типологію політичної карикатури як одного з об'єктів політичної лінгвістики. Показано, що карикатура є не лише інструментом політичної комунікації, а й способом вираження іміджу країни чи національних стереотипів. Проаналізовано чинники, які впливають на формування іміджу країни, визначено роль медіа у формуванні іміджу та стереотипів. Проведено комп'ютерний контент-аналіз американських, німецьких та українських політичних карикатур із використанням програми MAXQDA, проаналізовано подібності та відмінності текстово-візуального оформлення карикатур США, ФРН та України.



Список літератури

1. Корпус текстів української мови (<http://www.mova.info/corpus.aspx?l1=209>); Український національний лінгвістичний корпус (http://corp.ulif.org.ua/virt_unlc/); Das Gralis-Korpus — паралельний корпус сербських, хорватських і боснійських текстів (http://www.gewi.uni-graz.at/gralis/korpusarium/gralis_korpus.html); Dolnoserbški tekstowy korpus (<http://www.dolnoserbški.de/korpus/>); Englesko-črnogorski paralelni korpus (<http://www.eiprevod.gov.me/korpus/>); Hrvatski nacionalni korpus (<http://www.hnk.ffzg.hr/>); Hrvatski jezični korpus (<http://riznica.ihjj.hr/>); The Oslo Corpus of Bosnian Texts (<http://www.tekstlab.uio.no/Bosnian/Corpus.html>); Korpus slovenskega jezika (<http://www.fidaplus.net/>); Nova beseda — словенський корпус (http://bos.zrc-sazu.si/a_beseda.html); GOS (GOvorjene Slovenščine) — усний словенський корпус (<http://www.korpus-gos.net/Support/About>); Macedonian corpus (http://www.tekstlab.uio.no/glossa/html/index_dev.php?corpus=mak); Narodowy Korpus Języka Polskiego (<http://nkjp.pl/>); Polsko-Ukraiński Korpus Równoległy (<http://www.domeczek.pl/~polukr/index.php?option=welcome>); Český národní korpus (<http://ucnk.ff.cuni.cz/>); Slovenský národný korpus (<http://korpus.juls.savba.sk/>); Беларускі N-корпус (<http://bnkorpus.info/>); Corpus Albaruthenicum — корпус наукових білоруських текстів (<http://grid.bntu.by/corpus/>); Национальный корпус русского языка (<http://www.ruscorpora.ru/index.html>); Balanced Corpus of Contemporary Written Japanese (BCCWJ) (<http://www.ninjal.ac.jp/english/products/bccwj/>); Learner's Language Corpus of Japanese (<http://cbllc.tufs.ac.jp/lc/ja/search.php?menulang=en>); Japanese-English Parallel Corpus (<http://www.manythings.org/corpus/>); Base textuelle FRANTEXT (<http://www.frantext.fr/>); Корпуси французької мови (<http://corp.hum.sdu.dk/cqp.fr.html>); BNC — The British National Corpus (<http://www.natcorp.ox.ac.uk/>); BYU-BNC — BRITISH NATIONAL CORPUS (у версії М. Девіса (<http://corpus.byu.edu/bnc/x.asp?w=1024&h=600>)); COCA — CORPUS OF CONTEMPORARY AMERICAN ENGLISH (<http://corpus.byu.edu/coca/>); COHA — THE CORPUS OF HISTORICAL AMERICAN ENGLISH (<http://corpus.byu.edu/coha/x.asp?w=1024&h=600>); CORPUS DEL ESPAÑOL (<http://www.corpusdelespanol.org/>); DeReKo — Das Deutsche Referenzkorpus (<http://www1.ids-mannheim.de/kl/projekte/korpora/>); DGD — Datenbank für Gesprochenes Deutsch (http://dgd.ids-mannheim.de:8080/dgd/pragdb.dgd_extern.welcome); Eastern Armenian National Corpus (<http://www.eanc.net/>); European Parliament Proceedings Parallel Corpus (<http://www.statmt.org/europarl/>); GEKKO — корпус сучасної грузинської мови (<http://iness.uib.no/gekko/overview>); Hellenic National Corpus (<http://hnc.ilsip.gr/en/>); Hindi Corpora (<http://www.cfilt.iitb.ac.in/~corpus/hindi/>); KorpusDK — данський корпус (<http://ordnet.dk/korpusdk>); Līdzsvarots mūsdienu latviešu valodas tekstu korpus (<http://www.korpus.lv/>); Magyar Nemzeti Szövegtár (http://mnsz.nyud.hu/index_hun.html); O CORPUS DO PORTUGUÊS (<http://www.corpusdoportugues.org/>); Корпуси португальської мови (<http://corp.hum.sdu.dk/cqp.pt.html>); ParaSol — A Parallel Corpus of Slavic and other languages (<http://www.slavist.de/>); Romanian corpus (<http://corp.hum.sdu.dk/cqp.ro.html>); Scottish Corpus Of Texts & Speech (<http://www.scottishcorpus.ac.uk/>); Språkbanken — банк шведської мови (<http://spraakbanken.gu.se/>); The Reference Corpus of Estonian (http://www.keeletehnoloogia.ee/projects-1/the-reference-corpus-of-the-estonian-language/comprehensive-corpus-of-estonian?set_language=en); WaCKy — The Web-As-Corpus (<http://wacky.sslmit.unibo.it/doku.php?id=start>); Корпуси італійської мови (<http://corp.hum.sdu.dk/cqp.it.html>); Кыбырымтатар тилининь лингвистик корпусы (<http://korpus.juls.savba.sk/QIRIM/>) тощо.
2. Praktyzny przewodnik po korpusach języków słowiańskich / Red. Milena Hebal-Jezińska. — Warszawa, — 2014.
3. Baker P. Sociolinguistics and Corpus Linguistics. — Edinburgh, 2010. — P. 8.
4. Tognini-Bonelli E. Corpus Linguistics at Work. — Amsterdam/Philadelphia, 2001. — P. 65.
5. Tognini-Bonelli E. Corpus Linguistics at Work, Amsterdam/Philadelphia, 2001. — P. 85.
6. Tognini-Bonelli E. Corpus Linguistics at Work, Amsterdam/Philadelphia, 2001. — P. 84–85.
7. Tognini-Bonelli E. Corpus Linguistics at Work, Amsterdam/Philadelphia, 2001. — P. 182.