

Дем'янчук Ю. І.,

кандидат економічних наук,

викладач кафедри технічного перекладу

Львівського державного університету безпеки життєдіяльності

КАТЕГОРІЯ БАГАТОМОВНОГО КОРПУСУ ТЕКСТУ У СУЧАСНІЙ ЛІНГВІСТИЦІ: ІСТОРІОГРАФІЧНИЙ АНАЛІЗ

Анотація. У статті розглядаються наукові дослідження багатомовних корпусів тексту у працях відомих лінгвістів. Здійснюється історіографічний аналіз вітчизняних та іноземних праць, які обґрунтовують проблеми реалізації паралельного корпусу тексту. Характеризуються історичні передумови становлення корпусної лінгвістики та їх вплив на розробку багатомовних корпусів. Вказується на проблеми та перспективи розвитку наукових розробок щодо можливості застосування багатомовних корпусів та їх конотаційних особливостей.

Ключові слова: контрактивна лінгвістика, конотація, лексикографічний аналіз, багатомовні корпуси, паралельний корпус тексту, національний корпус російської мови, історіографія дослідження, порівняльний аналіз.

Постановка проблеми. У сучасній лінгвістиці багатомовний корпус тексту залишається недослідженою категорією, яка потребує не лише теоретико-методологічного, а й глибинного порівняльно-історичного аналізу. Передусім це зумовлено активною розробкою двомовних та багатомовних корпусів тексту, які б позначали лексикографічні та стилістичні особливості досліджуваного перекладу. Саме зіставна, чи контрактивна, лінгвістика, яка вивчає фонетичні, морфологічні, стилістичні, лексико-граматичні розбіжності між корпусами тексту, може вирішити проблему ефективного вивчення однієї (або декількох) мов в процесі перекладу. Тому багато науковців чітко співвідносять корпусну та контрактивну лінгвістику для глибшого розуміння тексту та якісних розробок багатомовних паралельних корпусів.

Метою статті є історіографічний та теоретичний аналіз багатомовного корпусу в галузі корпусної та контрактивної лінгвістики. Згідно з цією метою і гіпотезою дослідження поставлені такі завдання: розглянути дослідження українських науковців в галузі корпусної лінгвістики щодо застосування багатомовних корпусів; проаналізувати розробки іноземних лінгвістів; здійснити порівняльно-історичний аналіз створення та застосування багатомовних корпусів тексту (від XX ст. до сучасності) з перспективою їх розвитку.

Виклад основного матеріалу. Оскільки багатомовні паралельні корпуси тексту є малодослідженим лінгвістичним явищем, то їх варто співвідносити з історією становлення корпусної та контрактивної лінгвістики.

Щодо важливості досліджень корпусної лінгвістики та багатомовних корпусів Т. Бобкова робить висновки, що корпусна лінгвістика є [2, с. 4]: філософською галуззю; характеризується власним предметом, об'єктом, метою дослідження, термінологічним апаратом, теоретичним, практичним підґрунтям і власною методологією; визначається застосуванням статистичних методів, комп'ютерних технологій і програмного забезпечення (що, відповідно, дає можливість глибоко розглядати різні види

корпусів тексту з їх стилістичними особливостями); базується на визнанні тексту як головного об'єкта лінгвістичного дослідження й науковому характері власної методології дослідження (враховується лексикографічний, морфологічний та стилістичний рівні); корпусні дослідження ґрунтуються на певній дослідницькій філософії.

У статті «Історичні та концептуальні передумови корпусної лінгвістики» [1, с. 15] Т. Бобкова розглядає історичні періоди становлення корпусного перекладу, різновидів електронних корпусів тексту:

1) рання корпусна лінгвістика (кінець XIX ст. – 1960 рр.), що характеризується створенням доелектронних паперових корпусів текстів і формуванням теоретичного підґрунтя корпусного аналізу. Головна корпусна фонологічна розробка вказаного періоду – корпус Г. Дьюї для вивчення частотності звуків (1923 р.);

2) лексикологічний (і лексикографічний): 1) словники-конкорданси для перекладу Біблії; італійсько-англійський словник Дж. Флоріо (1598 р.); словник англійської мови С. Джонсона (1775 р.); Oxford English Dictionary (1928 р.);

3) морфологічний і синтаксичний: переклад античних джерел граматики Паніні (IV ст. до н.е.), наукових праць Л. Блумфілда (Language, 1933 р.), З. Харпика (Methods in Structural Linguistics, 1951 р.), Ч. Фріза (The structure of English, 1952); розробка корпусу англійського вживання (SEU, 1959 р.) Р. Кварка;

4) семасіологічний, у цей період досліджується лексична грамика французької мови Дж. Гадженгайма та Севаджета (1956 р.), здійснюється компаративний реєстр частоти значень слів в англійській, французькій, німецькій та іспанській мовах;

5) психолінгвістичний, у цей період досліджуються примітивні щоденники; мови дітей в онтогенезі Р. Брауна (1964 р.);

6) лінгводидактичний, у цей період був винайдений частотний словник англійської мови (1921 р.);

7) прикладний рівень, у цей період розробляється частотний словник німецької мови Ф.В. Кедінга (1897–1898 рр.), укладений для вдосконалення системи стенографії.

Серед найвідоміших методологічних розробок [8, с. 302] слід назвати такі: автоматизація лінгвістичного аналізу – дослідження лабораторії лексикографічного аналізу (Безансон, 1957 р.) і лабораторії Центру з автоматизації філологічного аналізу (Галараті, 1953 р.), розробка інструментарія машиночитаних корпусів – комп'ютерний конкордансер (1949–1967 рр.) Р. Буза при підтримці ІВМ; механолінгвістика й методологія формування вибірки (1956–1970) А. Джіланда.

В. Жуковська у дослідженні «Корпусна лінгвістика: історична перспектива та сучасний стан» [6, с. 23] особливу увагу приділяє системі розвитку двомовних та багатомовних кор-

писів. Особливу увагу автор приділяє електронному корпусу The Survey of English Usage, укладений Рендольфом Квіркком у 1959 р. Цей проект став перехідним етапом у становленні корпусної лінгвістики, оскільки він зосереджувався на лінгвістичних особливостях щоденного спілкування (як письмового, так і усного) звичайних громадян, і від початку збереження даних в електронному форматі не передбачалося. У свою чергу О. Демська-Кульчицька, здійснюючи огляд сучасного стану корпусної лінгвістики у своїй монографії «Основи національного корпусу української мови» [4, с. 140], зазначає, що це напрям досить розгалужений і передбачає студії, по-перше, загальної теорії корпусної лінгвістики, над якою працюють Д. Байбер, Дж. Синклер, В. Тойберт; по-друге, кореляції корпусної лінгвістики та інших лінгвістичних дисциплін; по-третє типології корпусів та методи інтерпретації корпусних даних; по-четверте, розроблення загальних засад створення природних мов тощо (праці Б. Алтенберга, М. Баньки, У. Френсиса, Г. Кеннеді, Г. Ліча, А. Баранова, М. Михайлова, Р. Рикова, Л. Ричкової, С. Шарова та ін.).

Г. Лук'янець наголошував на тому, що основою корпусної лінгвістики є розроблення теоретичних засад і практичних прийомів побудови, машинного опрацювання, експлуатації та аналізу мовних даних, оформлених як корпус текстів. В дослідженні «Основні напрями сучасних корпусних досліджень мови та перспективи їх подальшого розвитку» [7, с. 128] автор наголошує на тому, що багатомовні корпуси тексту – це свого роду інформаційно-довідкова система, яка слугує базою для дослідження одиниць та явищ різних мовних рівнів (фонетичного, морфологічного, лексико-семантичного та синтаксичного) для вивчення особливостей використання природної людської мови у формах усного та писемного мовлення та для визначення специфіки функціонування мови в різних стилях (художньому, офіційно-діловому, публіцистичному, науковому, розмовному).

Загалом ідея створення корпусу тексту зародилася в 60-х рр. ХХ ст. під значним впливом здійснених масштабних емпіричних досліджень, про які ми вже зазначали. До кінця 1960-х рр. існувало декілька невеликих корпусів, укладених на різних принципах. У 1980-х рр. – на початку 1990-х рр. відбулися вагомі досягнення в сфері комп'ютеризації, сканування текстів та зберігання інформації, і лінгвісти швидко зрозуміли, що ці нові інструменти пропонують величезні можливості для опису і аналізу мовних явищ.

У сучасній корпусній лінгвістиці визначення корпусу ґрунтується на чотирьох ознаках, які запропонували дослідники Т. Макенері і А. Вілсон. Цими ознаками, що визначають специфіку цієї словесної єдності, є «розміщення корпусу на машинному носії; стандартизоване відображення словесного матеріалу на цьому машинному носії, що дає змогу застосувати стандартні програми його обробки; кінцевий розмір і репрезентативність як результат особливої процедури відбору» [12, с. 116].

Під корпусом текстів В. Шевчук розуміє «величезний масив природних текстів (як письмових, так і усних) природної мови, представлених в комп'ютерному вигляді, тобто на машинному носії і належним чином упорядкованих, з метою їх використання в наукових і практичних дослідженнях» [11, с. 124]. Корпус текстів в ширшому сенсі можна трактувати як сукупність всіх текстів однієї мови, доступних для перекладача. Загалом під корпусом текстів розуміємо загальноприйняте поняття інтерпретації. Найвідомішим багатомовним національним корпусом тексту можна вважати НКРМ (національний корпус російської

мови). В. Шевчук підкреслює, що «національний лінгвістичний корпус – це багатомовна колекція усних і письмових текстів різних жанрів, стилів, регіональних і соціальних варіантів, що становлять великий пласт для вивчення мови. Національний лінгвістичний корпус застосовують для вивчення лексики і граматики мови, а також для відстеження найменших змін в мові. На основі саме національного корпусу створюються граматичні довідники та академічні словники» [11, с. 137].

Паралельні корпуси НКРМ, за ствердженням В. Січінаєвої [10, с. 208], вміщують такі паралельні двомовні корпуси, як англійську, вірменську, білоруську, болгарську, іспанську, італійську, латиську, німецьку, польську, українську, французьку, естонську. Практично для всіх з цих мов представлені симетричні мовні пари. У багатомовний корпус входять поліваріантні тексти. В паралельні корпуси також входять як художні, так і офіційні тексти з тієї ж класифікацією (публіцистика, виробничо-технічні, навчально-наукові, церковні, юридичні тексти). Тексти на більшості мов отримують морфологічну розмітку. Отже, у сучасній лінгвістиці мовні корпуси з легкістю можуть вміщувати мільйони слів реальних текстів, а кількість складних контрактивних аналітичних інструментів для обробки цих текстів швидко зростає.

Сама ж контрастивна лінгвістика почала розвиватися в кінці 1940-х рр. як окремий напрям прикладної лінгвістики, який мав вирішити багато практичних проблем, що виникають під час навчання іноземної мови. Для позначення цієї дисципліни використовуються й інші терміни: «контрастивний аналіз» і «міжмовні дослідження» (концепція Р. Булгарські) тощо. Ключовим положенням контрастивної лінгвістики є систематичне порівняння двох і більше мов на різних рівнях: від граматики до лексики, опису, зіставлення і порівняння. Особливо велика увага в контрастивній лінгвістиці приділяється розбіжностям між мовами: лінгвістичний опис двох мов може проводитися на основі будь-якої лінгвістичної моделі, проте та ж сама модель повинна використовуватися в обох випадках. Найчастіше як зіставний критерій використовується еквівалентність, особливо еквівалентність перекладу. В кінці 1970-х рр. стало очевидно, що отримані зіставні результати. Інтерес до контрастивної лінгвістики зменшився і знову почав розвиватися лише в 1990-х рр., коли з'явилися нові електронні інструменти, зокрема багатомовні комп'ютерні текстові корпуси. У сучасних умовах зіставлення мов і мовних явищ фактично отожднюється з лінгвістичною теорією перекладу. Цей погляд на співвідношення теорії перекладознавства і контрастивної лінгвістики досі знаходить свій вияв у деяких теоретичних побудовах, в яких вся проблематика перекладу зводиться до зіставлення окремих лексичних одиниць, морфологічних категорій і синтаксичних конструкцій у початковій мові і мові перекладу. Зокрема, у відомій праці Ж.-П. Віне і Ж. Дарбельнез порівняльної стилістики французької та англійської мов поєднуються поняття порівняльної стилістики та теорії перекладу. Загалом метою контрастивної лінгвістики є синхронне зіставлення мов, виявлення їх загальних і розпізнавальних рис на основі єдиної мови, що виступає як *tertium comparationis*, або однією з досліджуваних мов. Можна стверджувати, що контрактивна лінгвістика стала основою для глибинного розуміння багатомовних паралельних корпусів. Як матеріал для зіставлень використовуються тексти, у яких часто відсутній семантичний інваріант, але які дають змогу виявити диференціальні ознаки текстів конкретного жанру в мовах, які зіставляються.

У сучасний період ведеться детальна робота зі складання, дослідженню й опису електронних багатомовних корпусів різної тематики. Незважаючи на велику кількість робіт західних дослідників, в науковій літературі з корпусної лінгвістики цей напрям не отримав широкого поширення. Для дослідження процесу перекладу використовується кілька видів корпусів: а) паралельний корпус текстів, б) порівняльний (зіставний) корпус текстів. У дослідженні «Корпус паралельных текстов как инструмент сопоставительного описания языков» Д. Добровольский наголошує на тому [5, с. 290], що використання багатомовних паралельних корпусів дає можливість знайти перекладні еквіваленти для кожного слова і вирази їх в конкретних контекстах, на автентичному текстовому матеріалі. До подібних факторів належать синтаксичне оточення конкретного слова, його функція в структурі дискурсу, а також комунікативно-прагматичні умови його вживання. На думку дослідника, аналіз лексичних паралелей на значному за обсягом автентичному текстовому матеріалі дає можливість засумніватися у традиційних системах контрастивної лінгвістики і двомовної лексикографії.

Водночас паралельний корпус текстів деякими дослідниками інтерпретується також як перекладацький корпус (parallel / translational corpus). Особливою рисою паралельного корпусу є розмітка, яка при великих обсягах корпусів виконується автоматичним способом. При підготовці такого корпусу вихідній мові має відповідати пропозиція мови перекладу, проте в процесі перекладу таке співвідношення часто порушується, тому потрібно вдаватися до ручного редагування, щоб не допустити розбіжність пропозицій мови оригіналу з пропозиціями на мові перекладу при розмітці. Такий корпус дає можливість встановити відповідності перекладу між пошуковими елементами корпусу, виявити їх частотність і список контекстів, в яких живаються мовні одиниці.

Різновидом паралельного корпусу є зворотний перекладацький (back-translational corpus) та порівняльний (comparable corpus) корпуси. Другий тип вміщує в себе два різних набори текстів однією мовою: перший складається з оригінальних текстів, другий – з перекладених текстів. Такий корпус дає можливість розкрити внутрішню структуру мови: він сприяє якісному аналізу і опису мовних одиниць.

А. Віланденберг, Н. Беляєва, В. Захаров, С. Коваль, Ю. Марчук, В. Рубашкін, В. Риков, Л. Ричкова, А. Шайкевич особливу увагу приділяли спеціальному корпусу текстів. Такий корпус є багаторівневою системою багаторазового використання, що дає змогу здійснювати різні операції в ході вирішення конкретних дослідницьких завдань.

За твердженням А. Віланденберг [3, с. 72], корпус спеціальних паралельних текстів є адекватною базою для гармонізації термінології відповідної предметної сфери, зокрема офіційних документів. Дослідник робить висновки відносно стилістичних особливостей перекладу юридичних документів: для забезпечення семантичної однозначності правового документа і точності його юридичних формулювань в спеціальному двомовному та багатомовному словнику підлягають фіксації стійкі багатоконпонентні термінологічні словосполучення; функціональна класифікація термінів права вимагає виділення номінацій, уточнюючих і регулюючих термінів; на відміну від наукових і науково-технічних текстів, в правовому тексті особливу роль відіграють терміни – модальні дієслівні терміносистеми,

що є засобом вираження прагматичної спрямованості тексту, а також уточнюючі терміни – приєднані терміносполучення як засіб необхідної конкретизації правової норми.

Висновки. Таким чином, розглянувши еволюцію становлення контрастивної лінгвістики та її вплив на розробку багатомовних корпусів тексту, вдалося встановити ряд закономірностей, які протягом історії суттєво впливали на вироблення перекладацької тактики з офіційними документами. Оскільки більшість корпусів була створена в 90-х рр. ХХ ст., їх стилістична якість доволі низька. Що стосується перекладу сучасних книг та документів, то потрібно враховувати доповнюючі морфологічні, лексикографічні та стилістичні елементи, які додаються до сучасних багатомовних корпусів. Якісним можемо вважати НКРМ, на основі якого перекладацька робота з юридичними та офіційними документами дає можливість виявити позитивні мовні показники та алгоритми.

Література:

1. Бобкова Т. Історичні та концептуальні передумови корпусної лінгвістики / Т. Бобкова // Наукові записки Ніжинського державного університету ім. Миколи Гоголя. Сер.: Філологічні науки. – 2014. – Кн. 2. – С. 13–17.
2. Бобкова Т. До визначення корпусної лінгвістики у сучасному мовознавстві / Т. Бобкова // Наукові записки Національного університету «Острозька академія». Серія «Філологічна»: збірник наукових праць. – Вип. 45. – Острог: Видавництво Національного університету «Острозька академія», 2014. – С. 3–5.
3. Виландеберг А. Принципы и методы гармонизации терминологии на основе корпуса специальных параллельных текстов: на материале документов ООН: дисс.... канд. филол. наук: спец. 10.02.21 «Прикладная и математическая лингвистика» / А. Виландеберг. – СПб., 2005. – 266 с.
4. Демська-Кульчицька О. Основи національного корпусу української мови: [монографія] / О. Демська-Кульчицька. – К., 2005. – 218 с.
5. Добровольский Д. Корпус параллельных текстов как инструмент сопоставительного описания языков / Д. Добровольский // Русская и сопоставительная филология: состояние и перспективы: Международная научная конференция, посвященная 200-летию Казанского университета (Казань, 4–6 октября 2004 г.) / под. общ. ред. К. Галиуллина. – Казань: Изд-во Казан. ун-та, 2004. – С. 289–290.
6. Жуковська В. Корпусна лінгвістика: історична перспектива та сучасний стан / В. Жуковська // Матеріали за 8-а міжнародна научна практична конференція, «Ключові впроєкти в сьвременно та наука». – Софія, 2012. – С. 20–24.
7. Лук'янець Г. Основні напрями сучасних корпусних досліджень мови та перспективи їх подальшого розвитку / Г. Лук'янець // Наукові праці Національного університету харчових технологій. – 2012. – № 44. – С. 127–132.
8. Лингвистический энциклопедический словарь / гл. ред. В. Ярцева. – М.: Советская энциклопедия, 1990. – 688 с.
9. Кружков М. Информационные ресурсы контрастивных лингвистических исследований: электронные корпуса текстов / М. Кружков // Системы и средства информации. – 2015. – Т. 25. – Вып. 2. – С. 140–159.
10. Сичинава Д. Параллельные тексты в составе НКРЯ: новые направления развития и результаты / Д. Сичинава // Труды Института русского языка им. В. Виноградова. – 2015. – № 6. – С. 194–234.
11. Шевчук В. Электронные ресурсы переводчика: Справочные материалы для начинающего переводчика / В. Шевчук. – М.: Либрайт, 2010. – 159 с.
12. McEnery T. Corpus Linguistics / T. McEnery, A. Wilson. – Edinburgh: Edinburgh University Press, 2001. – 420 p.

Демьянчук Ю. И. Категория многоязычного корпуса текстов в современной лингвистике: историографический анализ

Аннотация. В статье рассматриваются научные исследования многоязычных корпусов текста в трудах известных лингвистов. Осуществляется историографический анализ отечественных и зарубежных работ, обосновывающих проблемы реализации параллельного корпуса текста. Характеризуются исторические предпосылки становления корпусной лингвистики и их влияние на разработку многоязычных корпусов. Указывается на проблемы и перспективы развития научных разработок по поводу возможности применения многоязычных корпусов и их коннотационных особенностей. Особое внимание уделяется контрактивной лингвистике, ее связям с общей лингвистикой и системой создания многоязычных корпусов текста.

Ключевые слова: контрактивная лингвистика, коннотация, лексикографический анализ, многоязычные корпуса, параллельный корпус текста, национальный корпус русского языка, историография исследования, сравнительный анализ.

Demyanchuk Yu. Body category bilingual text in modern linguistics: historiographical analysis

Summary. The article deals with research of multilingual text buildings in the works of famous linguists. Implemented historiographical analysis of domestic and foreign works, justifying the implementation of the problem of parallel corpus. They are characterized by the historical background of the formation of corpus linguistics, and their impact on the development of multi-language packages. It points out the problems and prospects of development of scientific research, the possibility of using multi-language corps and their connotations features. Particular attention is paid to contractive linguistics, its relations with the general linguistics and system for creating multi-lingual text hulls. The author made a conclusion on the appropriateness of NKRM (the national body of the Russian language), as a modern, updated translation tools to work with legal and official documents (in the body has all the necessary tools, the program is constantly updated).

Key words: contractive linguistics, connotation, lexicographical analysis, multi-lingual body parallel to the body of the text, national body of the Russian language, historiography of the study, comparative analysis.